Genome editing is a powerful tool to study gene function. The clustered regularly interspaced short palindromic repeat (CRISPR) system is an adaptive immune system found in bacteria. It can destroy naturally occurring and engineered phages and plasmids (1). A CRISPR genome editing tool was developed based on this system and has been used to edit the genomes of many species (2-4). CRISPR cleavage causes DNA double-strand breaks (DSBs), which are usually repaired by the nonhomologous end-joining (NHEJ) repair pathway (5,6). NHEJ is an error-prone process that often causes insertion/deletion (Indel) mutations, a portion of which result in frameshift mutations. In one study, a 51%–79% CRISPR-targeting efficiency was obtained for different genes in human embryonic stem (ES) cells. Similarly, CRISPR-induced mutation rates of up to 78% were obtained in mice (4,8–10).

For precise genome editing by inserting or deleting a designed fragment, gene targeting based on homologous recombination (HR) is often the preferred methodology. However, the efficiency of traditional gene targeting is generally low, ranging from undetectable to 0.1% (11). The DSBs induced by CRISPR can also be repaired through the homology-directed-repair (HDR) pathway. Here, a DNA fragment flanked by two sequences that are homologous to the sequences flanking the cleaved site can be inserted into the cleaved site by HR with efficiencies 5000 times higher than traditional HR (12).

Gene knock-in techniques mediated by the CRISPR/HDR pathway have been less well studied, but CRISPR/HDR

**METHOD SUMMARY**

Here we show that large DNA fragments (up to 7.4 kb) can be inserted into the genome of mouse embryonic stem (ES) cells at high efficiency using the CRISPR/homology-directed-repair (HDR) technique without the need for non-homologous end-joining (NHEJ) inhibitors.
targeting efficiencies have been reported to be only in the range of 0.5–20%, much lower than the efficiency of CRISPR mediated by NHEJ (4,8). Since HDR and NHEJ compete with each other for the same DSB site, the efficiency of HDR can be significantly increased by inhibiting NHEJ (13,14). However, because NHEJ is essential for DNA repair and embryonic viability, its deficiency may cause deleterious effects, such as mutagenesis and toxicity. Furthermore, it is unknown whether or not inhibiting NHEJ is effective for knocking in a DNA fragment at any locus, especially large fragments (>4 kb) containing multiple genes or a single large gene. So far, only small inserts and a few loci have been tested with this method (13,14). More importantly, the targeting efficiency of CRISPR in vivo is only ~3% (15). Therefore, additional investigation is required to enhance gene knock-in efficiencies for larger DNA fragments.

Here, large DNA fragments, 7.4 and 5.8 kb, were knocked in to the genomes of mouse ES cells and zygotes, respectively, using the CRISPR/HDR technique without any NHEJ inhibitors, at high efficiencies equivalent to that of CRISPR/NHEJ knockout in human ES cells (7) and to that of CRISPR/HDR using NHEJ inhibitors (13,14). Furthermore, a high survival rate of zygotes was obtained by injecting only one fourth of the regular concentrations of CRISPR reagents. This study also demonstrates that NHEJ inhibitors and homologous sequences longer than 2.2 kb are not prerequisites for knocking in a large (7.4 kb) fragment into the genomic locus of the gene under study at high efficiency.

Materials and methods

Mouse ES cell culture

The mouse ES cell line (JM8A3-N1, passage 7) was purchased from the Mouse Biology Program (MBP) at the University of California Davis. ES cells were grown on gelatin-treated cell culture plates with mitomycin-treated feeder cells (PMEF; neo resistant) at 2.5 × 10^4 cells per 10 cm plate using Dulbecco’s modified Eagle’s medium (DMEM), which contains 20% ES cell qualified fetal bovine serum (FBS), 0.1 mM 2-mercaptoethanol, MEM nonessential amino acids, and 1000 U recombinant murine leukemia inhibitory factor/mL (Millipore, Billerica, MA).

Figure 1. Schematic of CRISPR/HDR targeting strategy. (A) Strategy to knock-in a transcription control and labeling cassette (TCLC) into the Lrba locus in mouse embryonic stem (ES) cells using the CRISPR/homology-directed-repair (HDR) technique. The single guide (sgRNA) sequence is in red, and the protospacer-adjacent motif (PAM) sequence (NGG) is in green. The underlined sequences are the SpeI cleavage site and the translation start codon of Lrba. The cleavage site of Cas9 nuclelease is indicated by a red arrowhead. (B) The sgRNA will guide the Cas9 nuclease to the target site and cleave the DNA. (C) HDR by the left and right arms (blue) of the target donor will then insert the TCLC into the Lrba locus between the promoter and the Lrba translation start codon (ATG). E: EcoRV site. Red boxes: Southern blotting probes. Triangles: foxP, locus of cross P1 site. STOP: three different polyadenylation signal sites for transcription termination and polyadenylation of mRNA. rtTA: reverse tetracycline controlled transcriptional activator. TRE: tetracycline responsive element.

PCR screen of knock-in-positive ES cell clones

To screen knock-in-positive ES cell clones by PCR, the primers mLALF and mLALR (Supplementary Table S1) were designed from the mouse genomic sequence immediately outside of the left homologous arm and the CAG promoter in the target vector, respectively (Figure 1C). The fragment was amplified by PCR in a 20 μl reaction mixture composed of 4 μl 5x PrimeSTAR (Clontech Laboratories, Mountain View, CA) GXL buffer, 1.6 μl 2.5 mM dNTP mix, 0.4 μl 10 μM forward and reverse primers, 1 μl ES cell DNA from the 96-well preparation, 0.4 μl DMSO, and 0.4 μl PrimeSTAR GXL DNA polymerase (5 U/μl). Sterile distilled water was added to bring the total reaction volume to 20 μl. The PCR cycling parameters were: 98°C
for 5 min; 5 cycles of 94°C for 10 s and 75°C for 3 min 30 s; 5 cycles of 98°C for 10 s and 72°C for 3 min 30 s; 26 to 30 cycles of 98°C for 10 s and 69.4°C for 3 min 30 s; and 1 cycle of 72°C for 10 min. To detect the presence of the wild-type (WT) allele, the primers LAFWD1 and LAREV1 (Supplementary Table S1) that are located upstream and downstream of the targeting site, respectively, were designed from the mouse genomic sequence. The PCR conditions were the same as the above except for the PCR cycling parameters, which were: 98°C for 5 min; 5 cycles of 98°C for 10 s and 72°C for 1 min; 5 cycles of 98°C for 10 s and 69°C for 1 min; 30 cycles of 98°C for 10 s and 66°C for 1 min; and 1 cycle of 72°C for 10 min. PCR products (15 μl) were digested with 5 units of SpeI (NEB, Ipswich, MA) and 3 μl of CutSmart buffer in a total volume of 45 μl at 37°C for 4 h.

We also examined the 3’ end integrity of the knock-in by detecting the presence of the third loxP sequence using Taqman real-time PCR with primers loxP3F and loxP3R and dual-labeled Taqman probe loxPProbe (Supplementary Table S1). One microliter genomic DNA was used for real-time PCR. All reactions were run on the CFX96 Real-Time System (Bio-Rad Laboratories) in 20 μl reactions using Premix Ex Taq (Probe qPCR) (Clontech Laboratories) with 10 μM primers and probe. The PCR cycling parameters were: 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. The baselines and quantification cycle (Cq) were set automatically.

**Southern blotting confirmation of knock-in-positive ES cell clones**

Southern blotting was performed using standard methods. The DNA probe was prepared with the PCR DIG Probe Synthesis Kit (Roche Diagnostics, Risch, Switzerland) by using the donor vector as the template with the primers LrbaFp and LrbaRP (Table S1). Ten micrograms of genomic DNA were digested with EcoRV overnight and run on a 0.7% agarose gel. The DNA was transferred to a nylon membrane and hybridized with a digoxigenin-labeled probe. After stringent washing, the membrane was incubated with alkaline phosphatase-conjugated anti-digoxigenin antibodies (#11093274910; Roche Diagnostics). The chemiluminescent detection of positive signal on the membrane was achieved by incubation with CSPD (a chemiluminescent substrate for alkaline phosphatase) and developed on an X-ray film.

**Microinjection of mouse embryonic stem (ES) cell clones and blastocysts targeted by CRISPR/homology-directed-repair (HDR).** (A) PCR screening for Lrba knock-in-positive ES cell clones. The predicted size of the PCR product is 3.5 kb. Ctrl: control. PCR template was from untargeted ES cells. loxP3: the third loxP site. (B) Southern blotting. EcoRV-digested genomic DNA was hybridized with a 5’ external probe with expected fragment sizes of 5 kb (WT) and 13 kb (knock-in, KI). Lane 1, linearized donor vector (22 kb, 100 pg) control. Other lanes: G418-positive clones. #4 and #9 were incorrectly targeted, and #6 is WT. The other clones (#2, 3, 5, 7, 8, 10, and 11) were correctly targeted heterozygotes. (C) The presence of the third loxP site in these ES cell clones was detected by real time PCR using the loxP sequence as a Taqman probe. Some results are shown at the bottom of (A). RFU: relative fluorescence units. Positives: with the presence of the third loxP site; negatives: without the presence of the third loxP site. (D) Nested PCR detection of correctly targeted blastocysts derived from zygotes injected with CRISPR reagents. One quarter of the DNA from a blastocyst was used for the first PCR, as in (A). Ctrl: control. PCR template was from an uninjected zygote. (E) Targeting efficiency of CRISPR/HDR in ES cells and zygotes.
into 10 μl water and stored at -80°C. To extract blastocyst genomic DNA, 10 μl 2× digestion buffer (100 mM KCl, 20 mM Tris-HCl pH 9.0, 0.2% Triton X-100, and 0.8 mg/mL proteinase K) were added into each tube containing 1 blastocyst and incubated at 55°C overnight, followed by incubation at 94°C for 10 min to inactivate the proteinase K. Five microliters of the blastocyst DNA preparation were used as PCR template for the first PCR reaction using the same conditions and parameters as the PCR screening of knock-in-positive ES cell clones, with the exception that the number of cycles in the third phase of the touchdown PCR program was changed to 20. The second (nested) PCR reactions were conducted as follows: the nested PCR primers were LAFWD1 and FL30R (Supplementary Table S1), and the PCR conditions were the same as the above except for the cycling parameters, which were: 98°C for 5 min; 5 cycles of 98°C for 10 s and 69.5°C for 1 min; 5 cycles of 98°C for 10 s and 66.5°C for 1 min; 25 cycles of 98°C for 10 s and 63.5°C for 1 min; and 1 cycle of 72°C for 10 min.

### Results and discussion

We sought to integrate a 7.4 kb transcription control and labeling cassette (TCLC) at the Lrba genomic locus in mice (Figure 1). A donor Lrba vector was constructed from a BAC subclone, had homologous arms of 3.4 kb and 9 kb flanking the TCLC, and was characterized by restriction digestion and then confirmed by sequencing (Figure 1A, Supplementary Figure S1). We found that 67% of G418-positive ES cell clones were correctly targeted for the 5’ HR (Figure 2A), while more than 82% of G418-positive clones also had the third loxP sequence (loxP3, Figure 2C and bottom of Figure 2A), the presence of which may be an indication of the integrity of the 3’ end of the insert. The clones that were positive for both 5’-recombination and the third loxP sequence have likely been correctly targeted. Since PCR approaches cannot detect concurrent random integration, Southern blotting was used to confirm the presence of the TCLC. We found that 70% of clones were correctly targeted based on the Southern blot (Figure 2B).

This high CRISPR/HDR-mediated knock-in efficiency obtained in ES cells prompted us to test whether similar efficiencies could be obtained in mouse zygotes. The CRISPR reagents injected at the regular concentrations (8) interfered with the normal development of zygotes. Only 9% of zygotes injected with the CRISPR reagents developed into blastocysts (Table 1). To reduce this toxicity, the reagents were diluted 4-fold. The diluted reagents produced the best results when compared with the positive controls with or without the DNA ligase IV inhibitor Scr7, which inhibits NHEJ (Table 1). Nested-PCR genotyping revealed that the targeting efficiency for mouse zygotes was 32% (Figure 1D). This frequency would have been higher if the regular concentrations of CRISPR reagents had been used. High efficiency of HDR-mediated knock-in in mice is expected as the CRISPR knockout efficiencies are almost the same in ES cells and mice (4). The advantages of using lower concentrations of CRISPR reagents may include lower toxicity, lower off-target rates, and lower cost. Evidence of genome editing in blastocysts indicated that genome-edited mice could be produced. We then injected 311 embryos with the diluted CRISPR reagents and obtained a survival rate of 82% for the blastocysts, which were transferred to pseudopregnant recipients. We obtained a 34% birth rate (86 pups). The high survival rate of blastocysts and the high birth rate support the use of low concentrations of CRISPR reagents.

The high efficiency of CRISPR/HDR (Figure 1E) without NHEJ inhibitors suggests that long homologous arms (3.2 and 7.6 kb) (Supplementary Table S2), which are much longer than the recommended 800 bp (8), are required to obtain high knock-in efficiency (17). The frequency of traditional gene targeting is roughly proportional to the extent of homology shared by the transgene and its target locus (18). For example, a 40-fold increase in the rate of targeting has been observed with an increase in homology from 4 kb to 9.1 kb (19). A 25-fold increase was seen over the range of 2.5 kb to 9.5 kb of homology (20). We thus conducted the parallel comparison of targeting efficiency of two targeting vectors with different lengths of homologous arms flanking the same insert (Supplementary Table S2). Although there is total of ~7 kb difference of the homologous sequences between the 2 targeting vectors, the results show that there is no obvious difference in targeting efficiency between the 2 targeting vectors (Figure 3A). This indicates that the targeting efficiency of CRISPR/HDR is not dependent on the length of homologous arm in the range of 1.7 kb to 7.6 kb, and that increasing the length of each homologous arm in this range may not increase the CRISPR/HDR targeting efficiency as it does for traditional gene targeting. Our results also show that a ~2 kb homologous arm at each end should be sufficient for CRISPR/HDR, and it can be easily obtained by PCR amplification, thus making it easier to construct a targeting vector. Although extremely short (50 bp) homologous arms of double-stranded DNA can mediate HDR at 5%–10% efficiency, they may not efficiently mediate HDR of a larger insert such as the 7.4 kb fragment used in this study (21). Indeed, increasing the insert length from 99 bp to 720 bp considerably reduced targeting efficiency (9-fold), but that can be compensated for by increasing the homology arm size (22). Increasing the lengths of homologous arms in the range of 50 bp to 2 kb likely increases the targeting efficiency. While high CRISPR/HDR gene targeting efficiency was obtained for both ES cells and zygotes, no biallelic knock-ins were detected by Southern blotting and PCR screening in ES cells. It is possible that the mutated alleles may have NHEJ-mediated small insertion/deletion (indel) mutations that cannot be distinguished from the WT.
alleles by the two methods used above. These mutations would destroy the SpeI site (Figure 1B) in the Cas9/sgRNA recognition site, since the Cas9/sgRNA complex cleaves 3–8 nucleotides upstream of the protospacer adjacent motif (PAM) (23–25), and this SpeI site is located just within the cleavage range. Although the feeder cells contribute to the PCR products, their numbers in each culture well were ~20 times less than that of ES cells. Therefore, the PCR products should mainly originate from the ES cells. SpeI digestion can thus be used to detect NHEJ-induced mutations. In a CRISPR/NHEJ experiment, SpeI can digest the PCR product (713 bp) from the WT allele, producing 527 bp and 186 bp fragments, but it cannot cut the CRISPR-mutated allele as the SpeI site would be destroyed by the CRISPR mutation (Figure 3B). However, in the CRISPR/HDR experiments, all PCR products were digested by SpeI (Figure 3C); there was no uncleaved band for any clone among the 41 targeted clones and the 19 non-targeted clones. These results indicate that all of these clones have at least one WT allele, but there are no mutated alleles resulting from CRISPR/NHEJ, while this Cas9/sgRNA system can completely cleave its substrate in vitro (16) and successfully mutated the target site in ES cells at the absence of a donor vector (Figure 3B).

This is a surprising result given that the frequency of CRISPR/NHEJ mutations is usually higher than that of CRISPR/HDR (4,8). NHEJ occurs throughout the cell cycle, while HDR occurs only during the S and G2 phases (13). These two repair pathways seem to compete for DSBs, and inhibiting NHEJ significantly increases the efficiency of HDR (13,14). On the other hand, the CRISPR/Cas9 cleavage site is restored and can be cleaved again after complete repair of a DSB by NHEJ. This cycle of cleavage/NHEJ-repair may be repeated multiple times until the DSB is repaired with a mutation. However, once a DSB is repaired by HDR and a donor DNA fragment is inserted, the CRISPR/Cas9 cleavage site is destroyed and cannot be cleaved again by CRISPR/Cas9 (26). In this sense, CRISPR/HDR is more efficient than CRISPR/NHEJ, which may require multiple rounds of cleavage/repair. In
this study, it is possible that the NHEJ repairs almost all of the DSB sites at this genomic site without mutation, or that the NHEJ-mediated mutations are inhibited by some unknown mechanisms. As for no biallelic knock-ins being detected, it is likely that the double knock-in is lethal to the cells because knockdown of Lrba induces apoptosis (27). Since whether cells choose NHEJ or HDR is fundamental to CRISPR/Cas9 gene targeting, further investigation of these surprising results is warranted.

Another intriguing observation is that most of the G418-positive clones resulted from HR. This result is in sharp contrast to the traditional targeting, where most of the G418-positive clones result from random integration. The frequency of random integration is typically 1000 times more than that of targeted integration (18).

These results indicate that by increasing DNA DSBs at the correct site, CRISPR favors on-target integration as opposed to random integration.

We have demonstrated that (i) the targeting efficiency of CRISPR HDR without using any NHEJ inhibitors is equivalent to that of CRISPR HDR with NHEJ inhibitors; (ii) CRISPR/NHEJ is not detectable; (iii) biallelic knock-in is not detectable; (iv) left and right homologous arms longer than 2.1 and 1.7 kb, respectively, do not increase HDR efficiency; (v) one fourth of the regular concentrations of CRISPR reagents can be used to knock in a large DNA fragment into the genome of mouse zygotes at high efficiency with low toxicity; and (vi) large DNA fragments, 7.4 and 5.8 kb, can be knocked in to the genomes of ESC cells and zygotes, respectively. These results should inspire interest in further study of the mechanisms of high CRISPR targeting efficiency, as well as provide an example of how to obtain high efficiency gene knock-in results with large DNA fragments using CRISPR HDR without the need for NHEJ inhibitors.

Acknowledgments

This work was supported by the Joy McCann Culverhouse Endowment to the Division of Allergy, Department of Internal Medicine, and the University of Michigan Transgenic Animal Model Core.

Competing interests

The authors declare no competing interests.

References


Received 30 March 2015; accepted 27 July 2015.

Address correspondence to Jia-Wang Wang, Division of Allergy and Immunology, Department of Internal Medicine, University of South Florida Morsani College of Medicine, Tampa, FL. E-mail: jwang@health.usf.edu

To purchase reprints of this article, contact: biotechniques@fosterprinting.com