Published online 5 June 2017

Nucleic Acids Research, 2017, Vol. 45, No. 13 ***8105–8115***
doi: 10.1093/nar/gkx497

The contribution of homology arms to nuclease-assisted genome engineering

Oliver Baker[1,2], Sarah Tsurkan[2], Jun Fu[2,3], Barbara Klink[4], Andreas Rump[4], Mandy Obst[1,2], Andrea Kranz[2], Evelin Schröck[4], Konstantinos Anastassiadis[1,*] and A. Francis Stewart[2,*]

[1]Stem Cell Engineering, Biotechnology Center, Technische Universität Dresden, BioInnovationsZentrum, Tatzberg 47, Dresden 01307, Germany, [2]Genomics, Biotechnology Center, Technische Universität Dresden, BioInnovationsZentrum, Tatzberg 47, Dresden 01307, Germany, [3]Shandong University–Helmholtz Joint Institute of Biotechnology, State Key Laboratory of Microbial Technology, Shandong University, Shanda Nanlu 27, 250100 Jinan, People's Republic of China and [4]Institute for Clinical Genetics, Faculty of Medicine, Carl Gustav Carus, Technische Universität Dresden, Fetscherstrasse 74, Dresden 01307, Germany

Received February 02, 2017; Revised May 17, 2017; Editorial Decision May 24, 2017; Accepted May 30, 2017

ABSTRACT

Designer nucleases like CRISPR/Cas9 enable fluent site-directed damage or small mutations in many genomes. Strategies for their use to achieve more complex tasks like regional exchanges for gene humanization or the establishment of conditional alleles are still emerging. To optimize Cas9-assisted targeting, we measured the relationship between targeting frequency and homology length in targeting constructs using a hypoxanthine-guanine phosphoribosyl-transferase assay in mouse embryonic stem cells. Targeting frequency with supercoiled plasmids improved steeply up to 2 kb total homology and continued to increase with even longer homology arms, thereby implying that Cas9-assisted targeting efficiencies can be improved using homology arms of 1 kb or greater. To humanize the Kmt2d gene, we built a hybrid mouse/human targeting construct in a bacterial artificial chromosome by recombineering. To simplify the possible outcomes, we employed a single Cas9 cleavage strategy and best achieved the intended 42 kb regional exchange with a targeting construct including a very long homology arm to recombine ∼42 kb away from the cleavage site. We recommend the use of long homology arm targeting constructs for accurate and efficient complex genome engineering, particularly when combined with the simplifying advantages of using just one Cas9 cleavage at the genome target site.

INTRODUCTION

Site-directed cleavage of target genomes by designer nucleases is fundamentally changing genome engineering (1). Pioneering work with rare cutting endonucleases and zinc finger nucleases indicated the substantial benefits of targeted nuclease cleavage for genome editing and engineering (2,3). The advent of TALENs (4,5), CRISPR/Cas9 (6,7) and Cpf1 (8) has further simplified the technology and many genomes are now accessible for targeted damage (9,10). At lower but still workable frequencies, site-directed mutations can be inserted by co-introduction of an oligonucleotide to repair the DNA damage by homologous recombination (HR) and introduce an intended small mutation (11,12). However beyond site-directed damage or point mutagenesis, more complex genome engineering tasks remain challenging with many unresolved issues related to experimental design.

Nuclease-assisted targeting clearly has the potential to expand the ambit of genome engineering. To exploit this potential for challenging exercises, such as the establishment of alleles for conditional mutagenesis, simultaneous biallelic targeting or large regional replacements, the targeting strategy must be optimized. For example, before the recent development of nuclease-assisted targeting, knowledge of the relationship between the length of homology in the targeting construct and targeting frequency was critical for the development of efficient conventional strategies to target the mouse genome (13–15). Because nuclease-assisted targeting is promoted by the DNA break in the genome, it differs mechanistically from conventional targeting, which is promoted by the DNA ends of the linear targeting construct. Consequently, the relationship between the length of homology in the targeting construct and targeting frequency

*To whom correspondence should be addressed. Tel: +49 351 4634 0129; Fax: +49 351 4634 0143; Email: stewart@biotec.tu-dresden.de
Correspondence may also be addressed to Konstantinos Anastassiadis. Tel: +49 351 4634 0127; Fax: +49 351 4634 0143; Email: anastass@biotec.tu-dresden.de
Present address: Oliver Baker, Horizon Discovery Group plc, 8100 Cambridge Research Park, Waterbeach, Cambridge CB25 9TL, UK.

needs to be independently calibrated for nuclease-assisted targeting.

Previously, we presented an integrated recombineering platform for rapid Cas9-assisted targeting to establish protein tagged and conditional alleles in mouse embryonic stem (ES) cells (16). Whereas we achieved very high monoallelic targeting frequencies with targeting constructs built on 1 kb homology arms (i.e. 2 kb total homology), no simultaneous biallelic targeting events were found unless we increased the lengths of the homology arms. Therefore we decided to evaluate systematically the relationship between the length of homology and targeting efficiency in order to design better strategies for challenging applications such as our goal to humanize a mouse gene by a large regional replacement. We aimed to introduce the human version of the important cancer gene, KMT2D (also called Mll2 or Mll4) (17,18) in place of the mouse counterpart. This exercise involved replacing the mouse gene from initiating methionine codon to stop codon with the equivalent human region including introns, which is an ∼42 kb exchange. Because this is an order of magnitude greater than any exchange yet reported (19), we were faced with a variety of uncertainties regarding strategy, including the length of homology arms for the targeting construct.

MATERIALS AND METHODS

Feeder-free R1 mouse ES cells (5×10^6) were co-electroporated with 15 μg gRNA expression plasmid that recognizes exon 6 of hypoxanthine-guanine phosphoribosyl-transferase (HPRT) and 5 μg Cas9 expression plasmid (16). Seven days after plating at low density, 6-thioguanine (6-TG; 10 μM) was added. Isolated colonies were expanded and analyzed for HPRT exon 6 mutations by polymerase chain reaction (PCR) and sequencing. Clone 18 carried a 3 bp deletion and 1 bp substitution at the Cas9 cleavage site and was taken for the experiments of Figure 1. HPRT targeting constructs were made by recombineering to subclone pieces (400 bp, 1, 2, 10 and 20 kb) from mouse bacterial artificial chromosome (BAC) RP23–13N1 centered on the exon 6 mutation site and flanked by AgeI sites. Equimolar amounts of these subclones (corresponding to 2.5×10^{-13} moles), digested with AgeI or not, were electroporated with 15 μg gRNA expression plasmid that recognizes the mutant exon 6 and 5 μg of Cas9 expression plasmid into clone 18 (5×10^6 cells). Twenty four hours after plating at low density in standard Dulbecco's modified eagles medium (DMEM) supplemented with 10% heat-inactivated fetal calf serum (FCS) and self-made recombinant leukemia inhibitory factor (LIF), HAT (hypoxanthine, aminopterin, thymidine) selection was applied for 9 days. The plates were subsequently fixed and stained with Coomassie blue for counting. The HPRT rescue assay was also used with a 100 mer oligonucleotide centered on the exon 6 mutation at increasing inputs. The lowest input (Figure 1C; 0.5 μg) is 6 times more than the molar inputs used in the plasmid targeting series.

Kmt2d BACs (mouse: RP23–367C9 and human: RP11–437A15) were obtained from CHORI and modified as outlined in Figure 2. Oligonucleotides were purchased from Sigma (Steinheim, Germany) and are listed in Supplementary Table S2. All recombineering steps were performed as previously described (16,20). The PacI digested (2 μg) modified human KMT2D BAC was gently extracted with phenol/chloroform, ethanol precipitated, washed with 70% ethanol, dried, resuspended in 5 μl water and then electroporated as a standard recombineering sample. The coding region of the human KMT2D sequence in the final chimeric BAC was sequence verified using Illumina's TruSight-One NGS-Panel for sequence library construction and BACs were purified using Nucleobond BAC 100 kit (Macherey-Nagel). E14Tg2a cells were cultured in standard ES cell medium (DMEM with 4.5 g/l glucose supplemented with 15% heat-inactivated FCS, 2 mM L-glutamine, 1 mM sodium pyruvate, 1% P/S and 100 μM non-essential amino acids, 100 μM β-mercaptoethanol) containing self-made recombinant LIF and 2i inhibitors (PD-0325901 [1 μM] and CT-99021 [3 μM]) (21). E14Tg2a (2×10^5 cells/well) were lipofected (Lipofectamine 2000; ThermoFisher) in a 6-well plate with chimeric BACs containing different lengths of homology arms (HA 4/7, HA 4/50 and HA 50/7) at various concentrations (500, 1000 and 2000 ng) in the presence or absence of both gRNA 1 and 2 (500 ng each) expression constructs and CAG-FLAG-NLS-linker-Cas9-IRES-puro (500 ng) (16). After 24 h cells were selected by G418 (0.2 mg/ml) for 7 days, resistant clones picked into 96-well plates, expanded and split into hygromycin B (0.2 mg/ml) containing ES media. Clones resistant to G418 and hygromycin B were screened by quantitative PCR (qPCR) for loss of allele (LOA) events.

Genomic DNA (gDNA) was prepared from 96-well plates and diluted to 10 ng/μl. qPCR reactions were set-up in 20 μl total volume containing 0.5 ng/μl gDNA and 0.4 μM primers using 2× GoTaq® qPCR Master Mix (Promega) and analyzed on the Mx3005P qPCR system (Stratagene). Each clone was run in triplicates and cycle threshold (Ct) values were normalized against average Ct of an internal control, Nxt2 (copy number = 2N). Expression was calculated by fold changes relative to wild-type control using $^{\Delta\Delta}$Ct method (22). Mouse LOA primer pair (mLOA a/b) was tested on human H9 DNA for specificity (not shown).

For fluorescent *in situ* hybridization (FISH) analysis of the integration site, the human KMT2D-spec-amp-ccdB-hyg BAC was labeled via Nick-translation using tetramethyl-rhodamin-5-dUTP (Roche) and hybridized on metaphase spreads together with control probes targeting mouse chromosome 15 (RP23–35D4—15qA1; RP23–278C10—mouse Kmt2d on 15qF1) and human chromosome 12 (RP11–240H4—12q14), respectively.

RESULTS

The relationship between homology and targeting frequency

To evaluate the relationship between the length of homology in the targeting construct and targeting frequency, we employed selection for expression of the HPRT gene, which is a well-known model system in the history of gene targeting (23–25). HPRT is on the X chromosome, so is a single copy gene in male ES cells. It can be both selected for, using HAT medium and against using 6-thioguanine (26). We

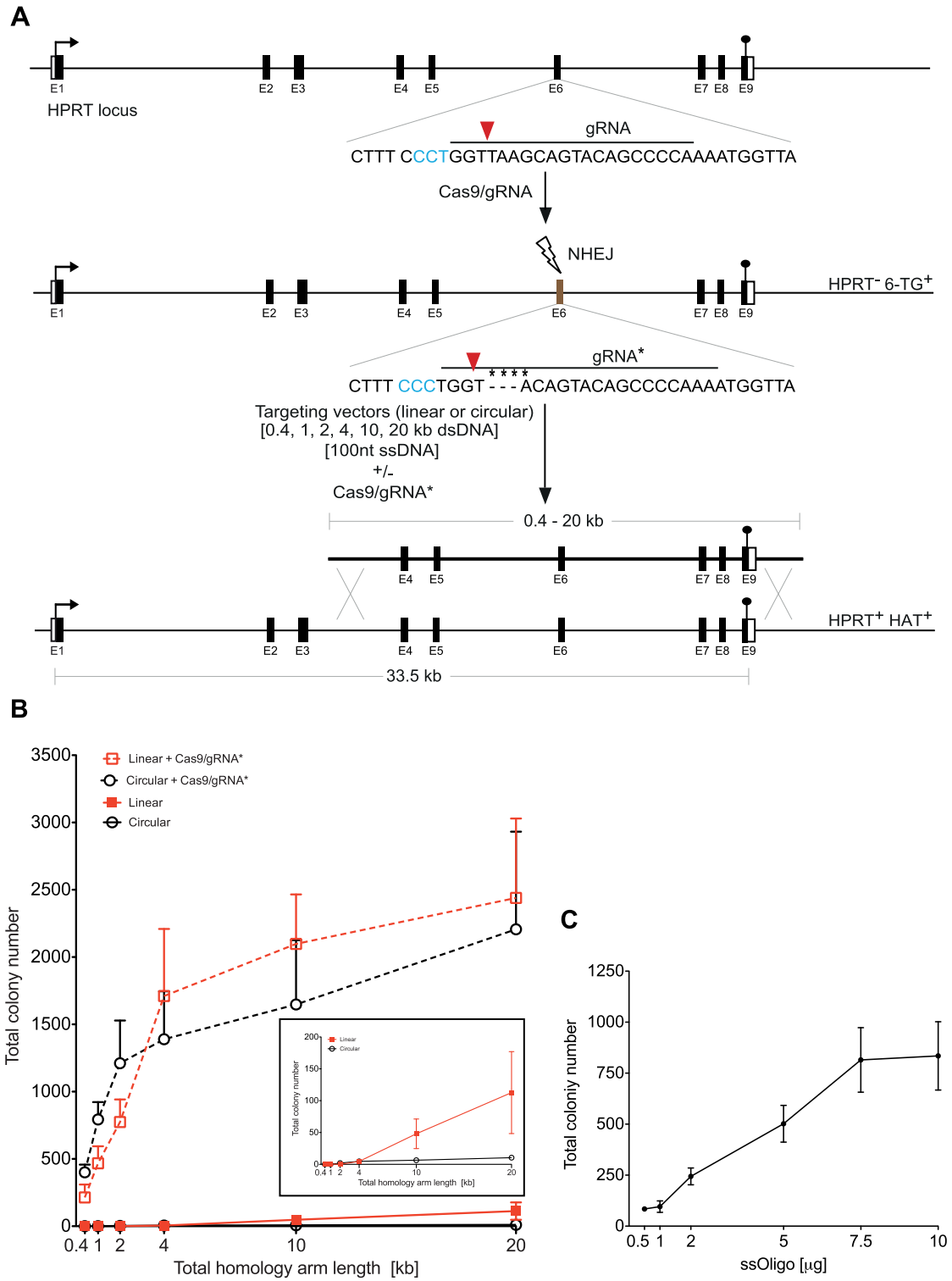

Figure 1. The relationship between homology arm length and nuclease-assisted targeting efficiency. (**A**) Diagram of the hypoxanthine-guanine phosphoribosyl-transferase (HPRT) assay used in R1 embryonic stem cells. The HPRT gene was mutated using the gRNA indicated and 6-thioguanine (6-TG) resistant cells characterized for damage by non-homologous end joining (NHEJ) at the Cas9 cleavage site. A clone with a 4 bp mutation shown below was taken for repair using the illustrated gRNA and selection for HAT (hypoxanthine, aminopterin, thymidine) resistance. (**B**) Plot of targeting efficiency expressed as total HAT resistant colonies against the total length of homology in the targeting construct. Four conditions are plotted based on the use of linearized (red lines) or uncut (circular; black lines) targeting construct plasmids with added Cas9 and gRNA expression plasmids (dotted lines) or with empty plasmids (solid lines). The inset shows an expansion of the data without Cas9. Results show the average of three independent experiments (error bars; SEM). (**C**) A 100 mer oligonucleotide centered on the exon 6 mutation was used in the same HPRT rescue assay at increasing inputs. The lowest input (0.5 µg) corresponds to a 6-fold molar excess over the molar inputs of the targeting constructs. Two independent experiments were performed in duplicates and standard errors calculated after subtraction of the background value achieved by omitting the oligonucleotide.

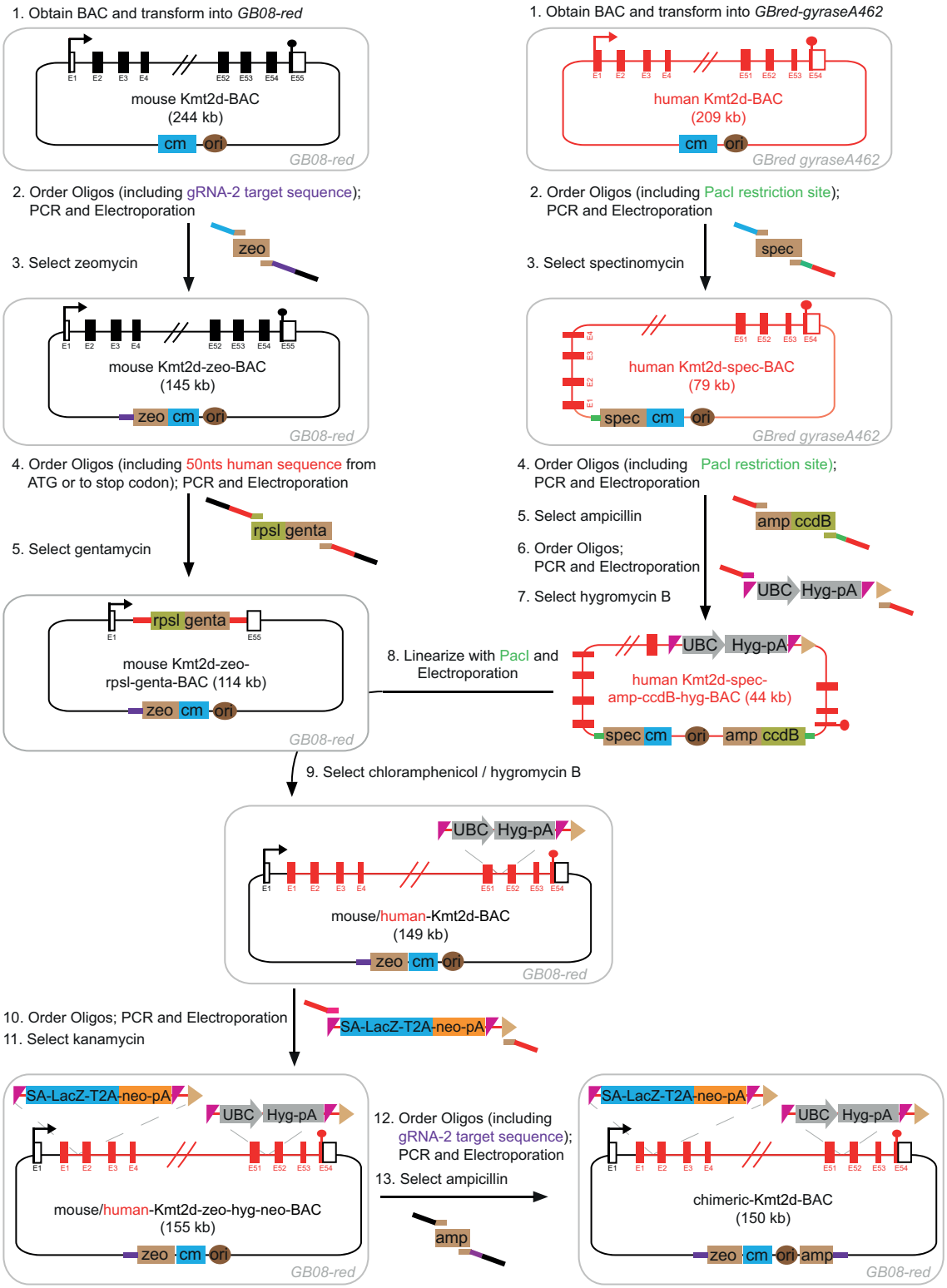

Figure 2. Recombineering to generate the hybrid mouse/human targeting bacterial artificial chromosome (BAC). The starting mouse Kmt2d BAC (black) is shown at the left and the starting human KMT2D BAC (red) at the right. Step 1 on both the left and right sides involved transforming the BACs into

established an HPRT rescue assay by first damaging exon 6 of the wild-type gene in R1 ES Cells with Cas9 to select a 6-TG resistant ES cell clone that can be rescued by HAT selection to restore the wild-type allele (Figure 1A). We rescued HPRT using segments of the wild-type HPRT gene subcloned from a BAC by recombineering (27,28) carrying homology regions either side of the exon 6 mutation from 200 bp to 10 kb (that is, total homology from 400 bp to 20 kb). None of these segments included the HPRT promoter so random integration cannot convey HAT resistance. Targeting was performed with either supercoiled (circular) or restriction digested (linear) plasmids with or without plasmids expressing Cas9 and a gRNA that is specific to the mutant HPRT allele. As expected without Cas9, very few HAT resistant colonies were obtained with circular plasmids and the targeting frequency with linearized targeting constructs increased with increasing homology length (Figure 1B) (13,14). The addition of Cas9 cleavage greatly promoted targeting frequency with both linear and circular targeting constructs. Apart from decreased efficiencies with linearized constructs at short homology arm lengths, which is likely due to exonuclease action, both circular and linear targeting constructs delivered the same outcome. This observation adds to the utility of Cas9-assisted targeting because the extra work to linearize the targeting construct, which was previously essential for conventional targeting (as again illustrated in Figure 1B insert) can be omitted. The degree to which Cas9 promoted targeting compared to conventional targeting varied with the amount of homology but a maximal promotion of ∼1000-fold was achieved at 2 kb total homology.

The Cas9-assisted targeting frequency with circular and linear substrates revealed the potentially important observation that both curves are bimodal showing steep increases up to 2 kb total homology with further moderate increases with longer homology arms. Why are these curves bimodal? We speculate that full stability of the synapse requires ∼1 kb of homology on either side and thereafter the moderate further gains reflects the increased probability that longer targeting constructs will find their matching sequences in the genome.

Also using the HPRT assay, we used a 100 mer oligonucleotide centered on the mutation site to restore HAT resistance. The lowest input used (0.5 μg; Figure 1C) corresponds to a 6-fold molar excess over the molar input of the plasmid targeting constructs. Even at a 120 molar excess (i.e. 10 μg), which appears to be saturating, the efficiency achieved does not match the targeting construct with 1 kb homology arms. At these high concentrations, ectopic oligonucleotide integration into the genome also becomes a concern.

Therefore challenging Cas9-assisted targeting exercises clearly benefit from long homology arms of at least 1 kb each and further benefit can be acquired with even longer arms.

Humanization of the Kmt2d gene in mouse ES cells

Very few genome engineering exercises to insert more than ∼10 kb into a specific site have been accomplished. Most large insertions have relied on illegitimate, or more recently transpositional, recombination (29,30), however these methods do not provide site-specific guidance. After a first round of HR to position site-specific recombinase target sites in the genome, site-specific recombination has been employed for large insertions. However this strategy for directed insertion of large regions has only been employed a few times (31,32). Multiple rounds of HR with BAC-based targeting constructs were used to achieve the spectacular feat of humanizing the mouse immunoglobulin loci (33). Apart from these examples, very few other large exercises have been reported primarily because of the difficulties and inefficiencies involved. To evaluate whether CRISPR/Cas9 could facilitate large, precise, engineering exercises, we aimed to exchange 42 kb of human KMT2D for its mouse counterpart. Our strategy involved (i) exchanging the region from start (initiating methionine) to stop codons including the introns; (ii) building the exchange in a BAC using recombineering; (iii) including gRNA cleavage sites (termed here gRNA-2) flanking the BAC targeting construct to employ the benefit of *in vivo* cleavage (34); (iv) evaluating different homology arm lengths; and (v) using Cas9-assisted targeting with only one Cas9 cleavage site (termed here gRNA-1), which cleaves at the 5′ end of mouse Kmt2d in the second intron. Although a strategy using two Cas9 gRNAs to cleave at each end of the mouse gene may appear intuitively more logical, we opted for a single cleav-

Escherichia coli hosts for recombineering; either *GB08-red* (a recombineering host with the arabinose-inducible red operon integrated into the chromosome) or *GBred-gyrA462* (*GB08-red* with a mutation to permit introduction of the *ccdB* gene) as indicated. Steps 2 and 3, on both sides, involved recombineering with polymerase chain reaction (PCR) products generated using oligonucleotides that include the homology boxes as well as the 5′ gRNA-2 target sequence or 5′ PacI restriction site as illustrated. Both recombineering steps deleted unwanted 5′ sequences: for mouse, deletion of ∼100 kb to leave 50 kb 5′ of the start codon: for human, deletion of all sequences 5′ of the start codon. On the left side, Steps 4 and 5 involved recombineering to delete the entire coding region of the mouse gene and replacement with 37 nt homology arms for the human gene (red) and the *rpsl* gene for later streptomycin counter-selection. On the right side, Steps 4 and 5 involved recombineering to delete the unwanted human sequences 3′ of the gene and simultaneously insert (i) the 3′ PacI site and (ii) the *ccdB* gene for later counter-selection (by the absence of the *gyr462* mutation). Steps 6 and 7 (on the right side) involved inserting a dual, *E.coli* and ES cell, hygromycin (hyg) selectable gene into a human 3′ intron to provide selection for the next recombineering step and later use during gene targeting in mouse ES cells. This hygromycin cassette is flanked by F3 sites (purple triangles) and includes the Ubiquitin C promoter (UBC). Step 8 involved cutting the human BAC with PacI to separate the human DNA from the BAC vector followed by electroporation of the PacI digest into GB08-red containing the product of Step 5, followed by selection for chloramphenicol (cm) and hygromycin (hyg) in Step 9. Candidate double resistant colonies were then checked for resistance to streptomycin and lack of resistance to spectinomycin (spec). Steps 10 and 11 involved insertion of an FRT flanked lacZ gene trap cassette, which conveys kanamycin (kan) resistance in *E.coli* and G418/neomycin (neo) resistance in ES cells (SA: splice acceptor; pA: polyadenylation signal), into the first human intron. Steps 12 and 13 involved insertion of the gRNA-2 target sequence at the 3′ end to either delete ∼43 kb or not. Subsequent optional deletion of ∼46 kb at the 5′ end is not illustrated. NB: according to ENSEMBL, the human KMT2D gene has one less exon than the mouse because the mouse has a 5′ non-coding exon whereas the human does not. We suspect that the human gene annotation has yet to include the corresponding 5′ non-coding exon because, as currently annotated, human exon 1 allows for almost no 5′ non-coding region.

Table 1. Summary of Kmt2d genotyping

Homology arms 5′/3′	LOA$^+$/tested	Southern/tested	FISH/tested	Correct
4/7 kb	8/66	5′ 6/8 3′ 3/8	3/3	3
4/50 kb	10/56	5′ 8/8	8/8	8
50/7 kb	6/62	3′ 1/5	1/5	1

For the 4/7, 4/50 and 50/7 kb homology arm constructs, 66, 56 and 62 colonies respectively were tested using the LOA assay. Of the successful candidates, 8, 8 and 5 respectively, were tested by Southerns at the indicated ends and the successful clones were tested by FISH. Further negative LOA clones were also tested by Southerns and FISH and are not summarized here.

age site to reduce the number and variety of unwanted mutagenic events that can arise from a double cleavage strategy (19,35). A single cleavage strategy also delivers easier genotyping because it bypasses the complications that can arise from the various possibilities provoked by two cleavages.

A multistep recombineering strategy was used to modify human and mouse BACs carrying the Kmt2d gene and build the compound targeting BAC (Figure 2). For the mouse Kmt2d BAC, 100 kb of unwanted sequence was removed to leave 50 kb 5′ of the start codon and simultaneously introduce a gRNA-2 cleavage site. Then the mouse coding region was replaced with the counter-selectable *rpsL* gene flanked by short stretches of human sequence immediately 3′ of the start and 5′ of the stop codons respectively. For the human KMT2D BAC, all sequences 5′ and 3′ to the start and stop codons were deleted and PacI restriction sites inserted along with the *ccdB* counter-selectable gene (20), then a hygromycin selection cassette including both *Escherichia coli* and mammalian promoters was inserted in intron 51.

All of the above steps are routine recombineering exercises. The next step was unusual. The human BAC was cut with PacI to release the 42 kb KMT2D coding region from the BAC vector, electroporated into cells carrying the modified mouse BAC followed by selection for hygromycin, chloramphenicol and streptomycin resistance. The counter-selectable genes, *rpsL* and *ccdB* served to eliminate unwanted products. The human/mouse hybrid BAC was further modified by the introduction of our standard *lacZ-neo* gene trap cassette (36) in the first intron and then by shaving at both ends to generate three 5′ and 3′ combinations of mouse homology arm lengths; namely 50/7, 4/50 and 4/7 kb. The 3′ end shaving included the introduction of a second gRNA-2 cleavage site. All recombineering junctions were sequenced to confirm accurate recombination and exclude any mutations introduced by faulty oligonucleotide synthesis.

The Kmt2d gene is expressed in mouse ES cells. Hence the acquisition of neomycin resistance was used to select for targeting in ES cells (Figure 3A). Hygromycin selection, either alone or combined with neomycin, was found to be unnecessary. However it provided further support for the genotyping (Supplementary Table S1).

Candidate colonies were initially genotyped using qPCR to count the changed number of mouse alleles from two to one (37). About 1/6th of all candidates from each of the three homology arm variations passed this LOA assay (Table 1). To verify Kmt2d targeting by Southern analysis, we had first validated the probes and restriction strategy before finalizing the recombineering decisions to make the two shorter homology arms (4 kb 5′ and 7 kb 3′; Figure 3A and data not shown). For each of the three homology arm combinations (4/7; 4/50; 50/7), eight candidates were taken for Southern analysis including three, eight and five clones respectively that were positive by the LOA assay (Figure 3B). Notably, all eight 4/50 LOA candidates were confirmed by Southern analysis as targeted at the 5′ end. For the 4/7 candidates, all three LOA candidates were confirmed homologous recombinants with both 5′ and 3′ probes. Unexpectedly, three of the five 4/7 LOA negative clones also displayed targeting at the 5′ end. However none of these three were targeted at the 3′ end. In contrast to the accuracy of the LOA assay for 4/7 and 4/50 clones, only one of the five LOA positive 50/7 candidates was confirmed by the 3′ end Southern analysis.

Because we observed half sided targeting with the 4/7 candidates and could not use Southern analysis to validate targeting with the 50 kb homology arms, we employed FISH (38) as a third genotyping criterion. The probes for the mouse and human Kmt2d genes were validated (Supplementary Figure S1A) and then applied to all clones that were positive in both LOA and Southern assays as well as selected negative clones. All clones that were double positive by the LOA and Southern assays showed introduction of the human KMT2D BAC at the correct locus by FISH and no other positive clones were found (Figure 3C and Supplementary Figure S1B). Furthermore no Cas9 induced damage was found on the other Kmt2d allele (Supplementary Figure S2), partially vindicating our choice to employ a single cleavage strategy.

From these three genotyping assays, we found that (i) the 4/50 targeting construct was better than the 4/7 or 50/7 configurations for Cas9-assisted targeting at the mouse Kmt2d locus; and (ii) both the LOA and one-sided Southern assays did not deliver conclusive genotyping. Each alone insufficiently defined the recombination event and additional genotyping was required to identify the correctly targeted 42 kb replacement with confidence.

These observations, particularly the finding that 6 of the 4/7 colonies were homologous recombinants at the 5′ end but only three of these were also correct at the 3′ end, reaffirm the need for careful and accurate genotyping. For further analysis, we proceeded with two correctly genotyped clones, F11 and G4.

Expression of the hybrid locus

Expression of the hybrid mouse/human mRNA is the ultimate proof for correct engineering in this exercise. Before we could confirm expression, the two selection cas-

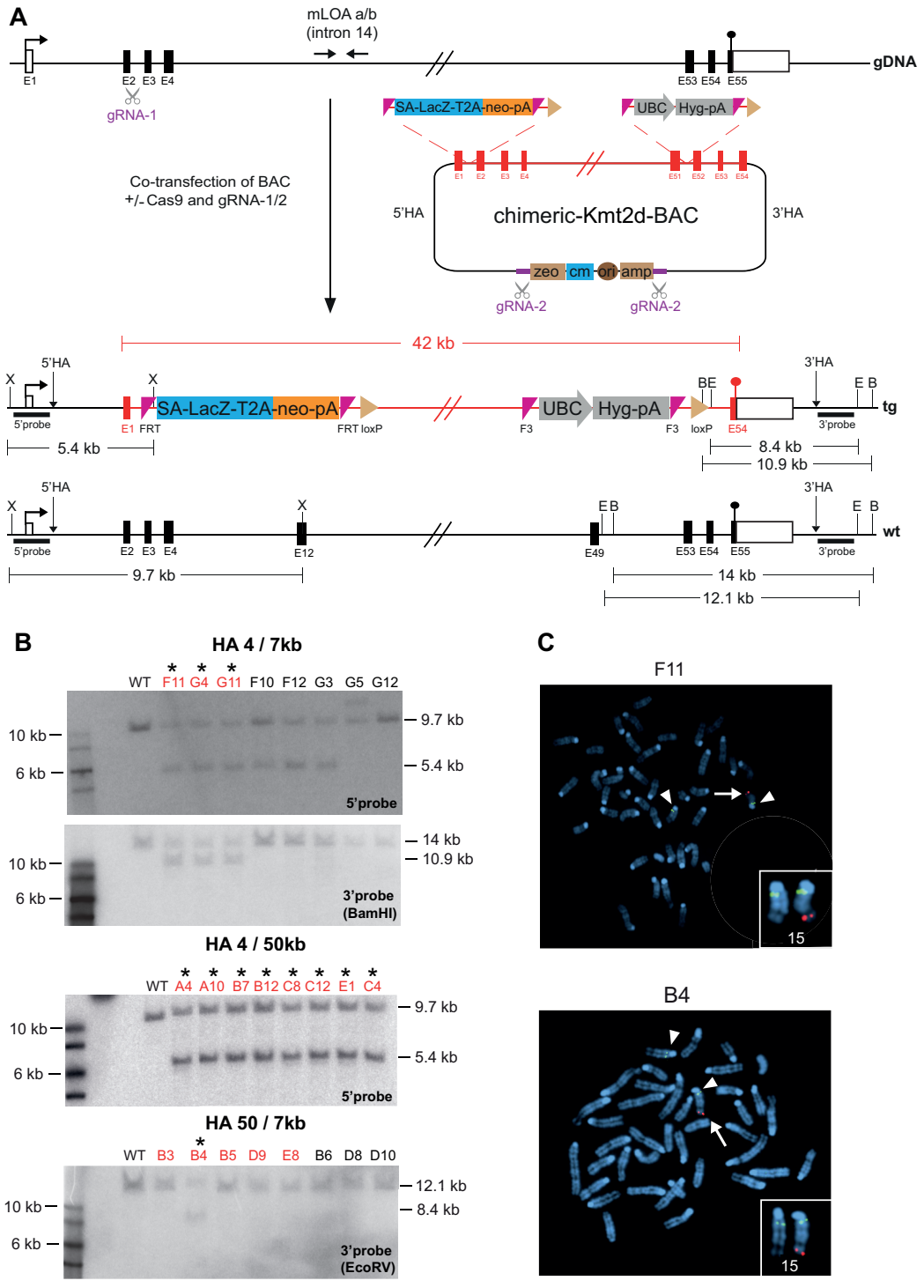

Figure 3. Replacement of the mouse Kmt2d gene with its human counterpart. (**A**) The endogenous mouse Kmt2d locus is depicted at the top along with the gRNA cleavage site and loss of allele (LOA) PCR primer pair. After co-transfection of the Cas9 and gRNA expression plasmids with one of the three (4/7, 4/50 and 50/7) circular BAC targeting constructs, candidate targeted events were selected using G418 and screened by LOA, followed by Southern blotting using the XbaI (X), EcoRV (E) and BamHI (B) digests depicted below. HA: homology arm. (**B**) Southern blots with LOA positive candidates denoted in red font and LOA negatives in black font. The DNA was digested with XbaI (X) and hybridized with the 5′ probe or digested with either EcoRV (E) or BamHI (B) and hybridized with the 3′ probe. Clones confirmed by fluorescent *in situ* hybridization (FISH) are marked with an asterisk. (**C**) FISH confirmation of correctly targeted ES cell clones. The human KMT2D BAC was labeled in red and hybridized together with a control probe for mouse chromosome 15 (chromosomal band 15qA1) labeled in green. Representative metaphase for clone F11 and B4 with a green signal (arrowhead) on both chromosome 15 and an additional red signal on one chromosome 15 at chromosomal band 15qF1 (where the endogenous Kmt2d gene is located) that corresponds to the integration of the human KMT2D BAC (arrow) is shown.

settes had to be removed using Flp recombinase (39). The FRT flanked *lacZ-neo* and UBC-*hyg* cassettes were flanked by FRT and F3 sites respectively to ensure no unwanted Flp recombination between the two cassettes (Supplementary Figure S3A). Transient expression of Flpo (40) resulted in about 90 and 36% excision of the *lacZ-neo* and UBC-*hyg* cassettes respectively with almost all of UBC-*hyg* excised clones also showing *lacZ-neo* excision (Supplementary Figure S3B). Both clones F11 and G4 showed the same ∼ 90:36% ratio of Flp recombination.

To evaluate mRNA expression, reverse transcriptase-PCR (RT-PCR) was performed using a primer specific to the mouse 5′ UTR combined with mouse and human primers specific to early exons (Figure 4A). Three Flp recombined subclones from each of F11 and G4 were examined and both the endogenous mouse and chimeric mouse/human transcripts were detected in all six subclones (Figure 4B).

DISCUSSION

Designer nucleases, strongly boosted by the technical simplicities of CRISPR/Cas9, present new opportunities for sophisticated genome engineering, especially in combination with other ways to engineer genomes including HR, endogenous ligation or site-specific recombination. Here we focused on Cas9-assisted HR by first examining the relationship between targeting efficiency and the length of homology in the targeting construct. Using an HPRT assay in mouse ES Cells, we found that efficiency increased steeply up to 2 kb total homology with further gains rendered by even longer homology arms. Although our findings agree with previous indications from nuclease-assisted targeting (3,41,42), they stand in apparent contrast to the widespread exploration of Cas9-assisted targeting with single stranded oligonucleotides that have much less total homology (11,12). Consequently we included a 100 mer in the HPRT assay to find that targeting efficiency was much lower although increasing molar input ∼100-fold (compared to the molar input of the targeting constructs used here) promoted HPRT rescue to about half of that achieved with the targeting construct with 1 kb homology arms (total 2 kb homology).

Although oligonucleotides only permit small changes, such as site-directed mutations or the insertion of a short tag, their use with Cas9-assisted targeting has become popular because the convenience of ordering oligonucleotide synthesis bypasses the recombinant DNA work needed to make a targeting construct. We suggest that this convenience is often a false economy because the labor saved by avoiding recombinant DNA work is less than consequential additional work elsewhere such as with the genotyping search for the intended mutation. Genotyping can be problematic if either the number of candidates that have to be screened becomes onerous or the intended recombination event has to be identified among various possibilities. An example of the first case involves Cas9-assisted targeting in human induced pluripotent stem cells, where often large numbers of candidates need to be screened to identify a suitable, biallelically modified, clone. An example of the second case involves Cas9-assisted targeting in zygote injections where mosaicism and non-homologous end joining can complicate the outcome. Herein lies a choice between the convenience of ordering an oligonucleotide followed by extensive genotyping or some recombinant DNA work to make a targeting construct followed by less genotyping. Recombineering has greatly simplified the generation of targeting constructs (15,43,44) and recently we described an even faster and simpler method for making targeting constructs with 1 kb homology arms (16). Even in the cases where the intended mutation is small enough to be included in an oligonucleotide, we suggest that it will often be less work to make and use a targeting construct than to use an oligonucleotide and have to screen more candidates. Furthermore, in the case of zygote microinjection, optimal ways to generate the intended mutation reduce the number of both zygotes needed and candidate founders to be screened, thereby reducing the number of animals used.

The complications of genotyping were also a consideration in our choice to utilize only one cleavage site for humanization of the Kmt2d gene. To avoid the potential complications arising from a Cas9 cleavage near each end of the replacement region, we decided to first examine the merits of a single cleavage strategy using variations of the targeting construct with different homology arm lengths. All three variations worked suggesting that only a single Cas9 cleavage is needed when employing a targeting construct with long homology arms. However the combination that delivered the best result positioned a 4 kb homology arm near the Cas9 cleavage site at the 5′ end and a very long 3′ homology arm (50 kb) to recombine with the distant 3′ homology region 42 kb from the Cas9 cleavage site. Indeed this configuration delivered a remarkable 14% targeting efficiency amongst G418 resistant colonies. Although more cases need to be documented before the design principles of nuclease-assisted targeting are finalized, we suggest that the use of long homology arms may circumvent the complications provoked by more than one Cas9 cleavage in the target region. For example, whether with selection in cells or without selection in zygote injections, we suggest that the introduction of loxP sites to establish conditional alleles will be best accomplished using a targeting construct with >1 kb homology arms and a single Cas9 cleavage site, ideally at the position of one of the loxP sites so that the targeting construct will not be cleaved. Based on the 4/50 example presented here, it may also be useful to make one homology arm extend well past the other loxP site. For example, a one Cas9 cleavage strategy to achieve a loxP allele could employ a short arm of ∼1 kb to the loxP at the Cas9 cleavage site and 4–5 kb past the other loxP site (with the smallest practical distance between the two loxP sites). However further work is required to compare the relative merits of one cleavage strategies using long homology arms and two cleavage strategies.

Recently other ways to introduce DNA into designer nuclease cleavage sites have been described employing ligation (41,45–47) or bridging oligonucleotides (48). Whether these methods are more efficient and reliable than nuclease-assisted HR using long homology arms remains to be determined but the best strategy will likely also be affected by the context, for example with or without selection, and the cell type, for example zygote or ES cell. However accu-

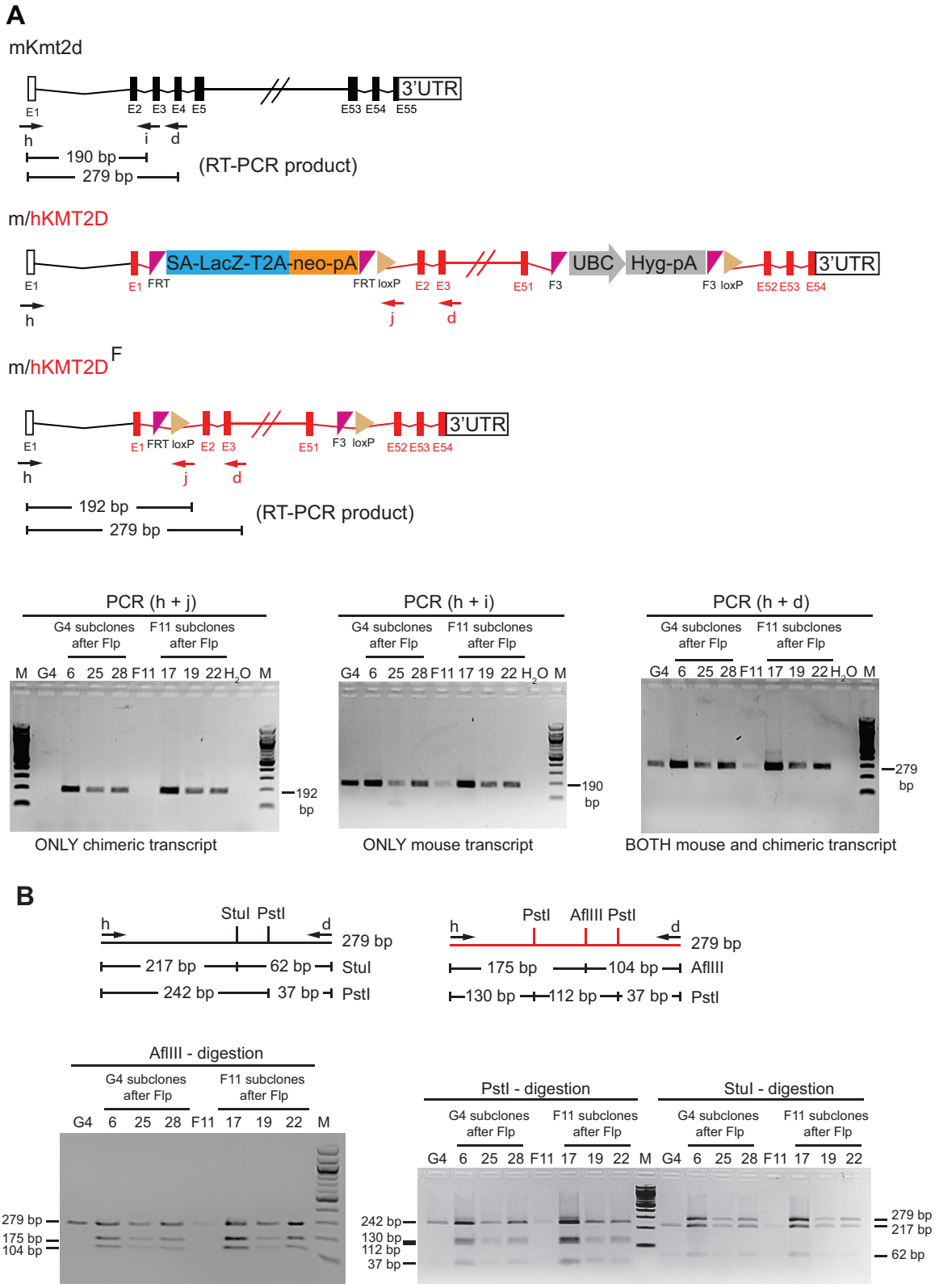

Figure 4. The hybrid mouse/human KMT2D gene is expressed in the targeted embryonic stem (ES) cells. (**A**) The primers (h, i, j and d) for Reverse transcriptase-PCR (RT-PCR) of wild-type (mKmt2d), targeted (m/h Kmt2d) or Flp recombined (m/h Kmt2d^{F}) Kmt2d alleles are illustrated on the gene diagrams and the results from the primer combinations (h + j; h + i and h + d) are shown below. (**B**) The RT-PCR products were digested with the indicated restriction enzymes to distinguish mouse and human Kmt2d transcripts from three independent Flp recombined subclones derived from either G4 or F1.

rately constructing the intended outcome in *E.coli* is a likely way to improve the yield of correct recombinants, especially across nuclease cleavage sites where unwanted events are common. Determining the most efficient strategies will be particularly important for applications that involve the generation of genetically modified animals or future cell type-specific genetic modifications of patients.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Stefanie Weidlich for helping with qPCR and Susan Schneider for subcloning segments of HPRT.
Author contributions: O.B., K.A., J.F. and A.F.S. designed the experiments. O.B., M.O. and K.A. performed experiments and generated the reagents except for the construction of the chimeric BAC which was done by J.F. S.T. performed HPRT assay. B.K., A.R. and E.S. performed and analyzed FISH experiments. O.B. and A.F.S. wrote the manuscript.

FUNDING

EU 7th Framework project, PluriMes (602423 to K.A.); Deutches Krebshilfe (110560 to A,K and A.F.S.); Else Kröner–Fresenius Stiftung (2012 A300 to A.K and A.F.S.). Source of Open Access funding: DFG Support the Best program of TUD.
Conflict of interest statement. A.F.S. is a shareholder in Gene Bridges GmbH, which holds exclusive commercial rights to recombineering.

REFERENCES

1. Carroll,D. (2014) Genome engineering with targetable nucleases. *Annu. Rev. Biochem.*, **83**, 409–439.
2. Rouet,P., Smih,F. and Jasin,M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.*, **14**, 8096–8106.
3. Moehle,E.A., Rock,J.M., Lee,Y.L., Jouvenot,Y., Dekelver,R.C., Gregory,P.D., Urnov,F.D. and Holmes,M.C. et. al. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3055–3060.
4. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.
5. Mussolino,C., Morbitzer,R., Lütge,F., Dannemann,N., Lahaye,T. and Cathomen,T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.*, **39**, 9283–9293.
6. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
7. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2579–E2586.
8. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E., Joung,J., van der Oost,J., Regev,A. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.
9. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
10. Komor,A.C., Badran,A.H. and Liu,D.R. (2016) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, **168**, 20–36
11. Chen,F., Pruett-Miller,S.M., Huang,Y., Gjoka,M., Duda,K., Taunton,J., Collingwood,T.N., Frodin,M. and Davis,G.D. (2011) High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods*, **8**, 753–755.
12. Wefers,B., Meyer,M., Ortiz,O., Hrabé de Angelis,M., Hansen,J., Wurst,W. and Kühn,R. (2013) Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 3782–3787.
13. Hasty,P., Rivera-Pérez,J., Chang,C. and Bradley,A. (1991) Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Mol. Cell. Biol.*, **11**, 4509–4517.
14. Deng,C. and Capecchi,M.R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.*, **12**, 3365–3371.
15. Skarnes,W.C., Rosen,B., West,A.P., Koutsourakis,M., Bushell,W., Iyer,V., Mujica,A.O., Thomas,M., Harrow,J., Cox,T. *et al.* (2011) A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*, **474**, 337–342.
16. Baker,O., Gupta,A., Obst,M., Zhang,Y., Anastassiadis,K., Fu,J. and Stewart,A.F. (2016) RAC-tagging: recombineering and Cas9-assisted targeting for protein tagging and conditional analyses. *Sci. Rep.*, **6**, 25529.
17. Rao,R.C. and Dou,Y. (2015) Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat. Rev. Cancer*, **15**, 334–346.
18. Kantidakis,T., Saponaro,M., Mitter,R., Horswell,S., Kranz,A., Boeing,S., Aygün,O., Kelly,G.P., Matthews,N., Stewart,A. *et al.* (2016) Mutation of cancer driver MLL2 results in transcription stress and genome instability. *Genes Dev.*, **30**, 408–420.
19. Byrne,S.M., Ortiz,L., Mali,P., Aach,J. and Church,G.M. (2015) Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.*, **43**, e21.
20. Wang,H., Bian,X., Xia,L., Ding,X., Muller,R., Zhang,Y., Fu,J. and Stewart,A.F. (2014) Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res.*, **42**, e37.
21. Ying,Q.-L., Wray,J., Nichols,J., Batlle-Morera,L., Doble,B., Woodgett,J., Cohen,P. and Smith,A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.
22. Schmittgen,T.D. and Livak,K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.*, **3**, 1101–1108.
23. Doetschman,T., Gregg,R.G., Maeda,N., Hooper,M.L., Melton,D.W., Thompson,S. and Smithies,O. (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, **330**, 576–578.
24. Thomas,K.R. and Capecchi,M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, **51**, 503–512.
25. Gravells,P., Ahrabi,S., Vangala,R.K., Tomita,K., Brash,J.T., Brustle,L.A., Chung,C., Hong,J.M., Kaloudi,A., Humphrey,T.C. *et al.* (2015) Use of the HPRT gene to study nuclease-induced DNA double-strand break repair. *Hum. Mol. Genet.*, **24**, 7097–7110.
26. Stout,J.T. and Caskey,C.T. (1985) HPRT: gene structure, expression, and mutation. *Annu. Rev. Genet.*, **19**, 127–148.
27. Zhang,Y., Buchholz,F., Muyrers,J.P. and Stewart,A.F. (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.*, **20**, 123–128.
28. Maresca,M., Erler,A., Fu,J., Friedrich,A., Zhang,Y. and Stewart,A.F. (2010) Single-stranded heteroduplex intermediates in λ red homologous recombination. *BMC Mol. Biol.*, **11**, 54.
29. Rostovskaya,M., Fu,J., Obst,M., Baer,I., Weidlich,S., Wang,H., Smith,A.J.H., Anastassiadis,K. and Stewart,A.F. (2012) Transposon-mediated BAC transgenesis in human ES cells. *Nucleic Acids Res.*, **40**, e150.
30. Rostovskaya,M., Naumann,R., Fu,J., Obst,M., Mueller,D., Stewart,A.F. and Anastassiadis,K. (2013) Transposon mediated BAC transgenesis via pronuclear injection of mouse zygotes. *Genesis*, **51**, 135–141.

31. Wallace,H.A.C., Marques-Kranc,F., Richardson,M., Luna-Crespo,F., Sharpe,J.A., Hughes,J., Wood,W.G., Higgs,D.R. and Smith,A.J.H. (2007) Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell*, **128**, 197–209.
32. Lee,E.-C., Liang,Q., Ali,H., Bayliss,L., Beasley,A., Bloomfield-Gerdes,T., Bonoli,L., Brown,R., Campbell,J., Carpenter,A. *et al.* (2014) Complete humanization of the mouse immunoglobulin loci enables efficient therapeutic antibody discovery. *Nat. Biotechnol.*, **32**, 356–363.
33. Murphy,A.J., Macdonald,L.E., Stevens,S., Karow,M., Dore,A.T., Pobursky,K., Huang,T.T., Poueymirou,W.T., Esau,L., Meola,M. *et al.* (2014) Mice with megabase humanization of their immunoglobulin genes generate antibodies as efficiently as normal mice. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 5153–5158.
34. Cristea,S., Freyvert,Y., Santiago,Y., Holmes,M.C., Urnov,F.D., Gregory,P.D. and Cost,G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.*, **110**, 871–880.
35. Lee,H.J., Kweon,J., Kim,E., Kim,S. and Kim,J.-S. (2012) Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.*, **22**, 539–548.
36. Testa,G., Schaft,J., van der Hoeven,F., Glaser,S., Anastassiadis,K., Zhang,Y., Hermann,T., Stremmel,W. and Stewart,A.F. (2004) A reliable lacZ expression reporter cassette for multipurpose, knockout-first alleles. *Genesis*, **38**, 151–158.
37. Frendewey,D., Chernomorsky,R., Esau,L., Om,J., Xue,Y., Murphy,A.J., Yancopoulos,G.D. and Valenzuela,D.M. (2010) The loss-of-allele assay for ES cell screening and mouse genotyping. *Meth. Enzymol.*, **476**, 295–307.
38. Liyanage,M., Coleman,A., du Manoir,S., Veldman,T., McCormack,S., Dickson,R.B., Barlow,C., Wynshaw-Boris,A., Janz,S., Wienberg,J. *et al.* (1996) Multicolour spectral karyotyping of mouse chromosomes. *Nat. Genet.*, **14**, 312–315.
39. Buchholz,F., Angrand,P.O. and Stewart,A.F. (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.*, **16**, 657–662.
40. Raymond,C.S. and Soriano,P. (2007) High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS One*, **2**, e162.
41. Orlando,S.J., Santiago,Y., DeKelver,R.C., Freyvert,Y., Boydston,E.A., Moehle,E.A., Choi,V.M., Gopalan,S.M., Lou,J.F., Li,J. *et al.* (2010) Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res.*, **38**, e152.
42. Beaton,B.P., Kwon,D.-N., Choi,Y.-J., Kim,J.-H., Samuel,M.S., Benne,J.A., Wells,K.D., Lee,K., Kim,J.-H. and Prather,R.S. (2015) Inclusion of homologous DNA in nuclease-mediated gene targeting facilitates a higher incidence of bi-allelically modified cells. *Xenotransplantation*, **22**, 379–390.
43. Angrand,P.O., Daigle,N., van der Hoeven,F., Schöler,H.R. and Stewart,A.F. (1999) Simplified generation of targeting constructs using ET recombination. *Nucleic Acids Res.*, **27**, e16.
44. Sarov,M., Schneider,S., Pozniakovski,A., Roguev,A., Ernst,S., Zhang,Y., Hyman,A.A. and Stewart,A.F. (2006) A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nat. Methods*, **3**, 839–844.
45. Maresca,M., Lin,V.G., Guo,N. and Yang,Y. (2013) Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.*, **23**, 539–546.
46. Lackner,D.H., Carré,A., Guzzardo,P.M., Banning,C., Mangena,R., Henley,T., Oberndorfer,S., Gapp,B.V., Nijman,S.M.B., Brummelkamp,T.R. *et al.* (2015) A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat. Commun.*, **6**, 10237.
47. Suzuki,K., Tsunekawa,Y., Hernandez-Benitez,R., Wu,J., Zhu,J., Kim,E.J., Hatanaka,F., Yamamoto,M., Araoka,T., Li,Z. *et al.* (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144–149.
48. Yoshimi,K., Kunihiro,Y., Kaneko,T., Nagahora,H., Voigt,B. and Mashimo,T. (2016) ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat. Commun.*, **7**, 10431.