Cellular and Metabolic Engineering **Biotechnology and Bioengineering**
DOI 10.1002/bit.25629

CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining†

Ravichandra Bachu[1], Iñigo Bergareche[1,2], Lawrence A. Chasin[1*]

[1] Department of Biological Sciences, Columbia University, New York, NY 10027, USA

[2] Present address: Synpulse USA Inc, New York, NY 10036, USA

*Correspondence to: Lawrence A. Chasin

912 Fairchild Center, Mail Code 2433, Columbia University, New York, NY 10027, United States.

Tel.: +1 212 854 4645.

E-mail address: lac2@columbia.edu

Running title: Targeted plasmid insertion in mammalian cells

†This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/bit.25629]

This article is protected by copyright. All rights reserved
Received April 10, 2015; Accepted April 27, 2015

Accepted Preprint

Accepted Preprint

Abstract

Mammalian cells are widely used for the production of therapeutic recombinant proteins, as these cells facilitate accurate folding and posttranslational modifications often essential for optimum activity. Targeted insertion of a plasmid harboring a gene of interest into the genome of mammalian cells for the expression of a desired protein is a key step in production of such biologics. Here we show that a site specific double strand break (DSB) generated both in the genome and the donor plasmid using the CRISPR-Cas9 system can be efficiently used to target ~5 kb plasmids into mammalian genomes via nonhomologous end joining (NHEJ). We were able to achieve efficiencies of up to 0.17% in HEK293 cells and 0.45% in CHO cells. This technique holds promise for quick and efficient insertion of a large foreign DNA sequence into a predetermined genomic site in mammalian cells. This article is protected by copyright. All rights reserved

This article is protected by copyright. All rights reserved

Introduction

The most commonly used hosts for the production of recombinant proteins for pharmaceutical use are cultured mammalian cells, as these permit folding and post translational modifications close to that of the natural protein. In many cases genetic engineering has been used to provide further specific advantages over wild type host cells. Examples include mutant Chinese hamster ovary cells (CHO) with altered glycosylation patterns (Yamane-Ohnuki et al. 2004), CHO cells deficient in dihydrofolate reductase (*DHFR*) to facilitate the isolation of transfectants with amplified transgenes (Urlaub and Chasin 1980; Urlaub et al. 1983), and mouse NS0 myeloma cells that no longer produce a background of endogenous immunoglobulin (Galfre and Milstein 1982). Recently, the genetic engineering of mammalian cell lines has been greatly facilitated by the use of three major classes of targetable nucleases that can produce site-specific double strand breaks (DSB): zinc finger nucleases (ZFNs), transcription activator–like effector nucleases (TALENs), and more recently CRISPR-Cas RNA-guided nucleases. Both ZFNs and TALENs consist of a DNA binding domain that binds to the targeted sequence in the genome joined to a nuclease domain, which produces the DSB. The endonuclease for both these classes of targetable nucleases is derived from the restriction enzyme *FokI* (Joung and Sander 2013; Kim et al. 1996). The DNA binding domains in ZFNs and TALENs are usually modeled after transcription factors; (Gaj et al. 2013).On the other hand, the recently discovered CRISPR-Cas system uses an RNA-guided nuclease, which is a fusion of a guide RNA that selectively binds to the DNA and the Cas9 endonuclease protein derived from *Streptococcus pyogenes* (Jinek et al. 2012; Mali et al. 2013a; Ran et al. 2013b).

Once the DSB is introduced into the genome, it can be repaired via two major pathways: high fidelity homology-dependent repair (HDR) or error-prone non-homologous end joining (NHEJ) (Huertas 2010; Lieber 1999; Pardo et al. 2009; Rouet et al. 1994) HDR is much less frequent than NHEJ in

This article is protected by copyright. All rights reserved

mammalian cells (Chu et al. 2015)and requires hundreds of base pairs of homologous DNA sequences that span the DSB. Exogenous DNA can provide this homology to serve as a template for faithfully repairing the break or for introducing a mutation. Thus homologous recombination-based gene targeting strategies are not so efficient but can generate precise gene modifications. In contrast, NHEJ occurs at a much higher frequency than HDR (Iyama and Wilson 2013; Sedivy and Sharp 1989) and requires no DNA repair template. It involves processing of the ends of the break and joining them, which is frequently associated with indels, small base pair insertions and deletions, (Carbery et al. 2010; Iyama and Wilson 2013; Pan et al. 2013). Due to resulting frame shift mutations and premature stop codons, NHEJ has been extensively used for generating gene knock-outs (Carbery et al. 2010; Pan et al. 2013; Shalem et al. 2014). Gene knock-ins via NHEJ have been successfully performed as well (Auer et al. 2014; Maresca et al. 2013; Pan et al. 2013). Thus NHEJ, though error-prone, can be useful due to its greater simplicity and efficiency compared to HDR.

The potential advantages of the CRISPR-Cas technology over ZFNs and TALENs are that it is relatively low cost, less time consuming and does not require complicated protein engineering (Carroll 2014). Our goal is to use this highly efficient CRISPR-Cas technology to introduce a DNA fragment into a specific location in mammalian genomes via NHEJ. The type II CRISPR-Cas system of *S. pyogenes* has been successfully employed for editing the genomes in different species such as nematodes, fruitflies, zebrafish, plants, mice, human cells (Auer et al. 2014; Bassett et al. 2013; Chang et al. 2013; Cho et al. 2013; Cong et al. 2013; Gratz et al. 2013; Hwang et al. 2013; Jiang et al. 2013; Mali et al. 2013b; Ronda et al. 2014; Waaijers et al. 2013; Wang et al. 2013; Yang et al. 2013) and most recently CHO cells (Ronda et al. 2014). All these studies have focused on gene disruption or on the HDR-mediated rescue of disrupted genes. Gene knock-in via the NHEJ mechanism using CRISPR-Cas technology has been successfully demonstrated in zebrafish (Auer et al. 2014). Here we

This article is protected by copyright. All rights reserved

demonstrate for the first time, CRISPR-Cas mediated insertion of ~5 kb long plasmid DNA sequences into the genome of HEK293 cells and CHO-K1 cells via the NHEJ pathway.

Materials and Methods:

Cell Culture and Transfection:

CHO-K1 and HEK293 cells were grown in Alpha Minimal Essential Medium with nucleosides and deoxynucleosides (Hyclone) supplemented with penicillin/streptomycin at 100 U/100 µg per ml and 10% fetal bovine serum (Atlanta Biologicals). We transfected ~ 2 x 10^5 cells grown to 40 – 50 % confluence in 1 ml of medium in a 12-well plate with 0.12 ml of Opti-MEM (Life Technologies) containing 250 ng each of two CRISPR-Cas plasmids and 500 ng of donor plasmid and 3.2 µl of Lipofectamine 2000 reagent (life Technologies).

Antibiotic selection: Kill curves were performed in order to determine the appropriate antibiotic concentration for cell selection. Sensitive (i.e., non-transfected) cells (~1.5 x10^5) were cultured in 6-well plates with increasing concentrations of Geneticin (G418), 6-thioguanine (6-TG) or puromycin and the medium was changed every 3 days. After 2 weeks, the lowest antibiotic concentration that killed all cells was chosen for transfection experiments. Kill curves were carried out for both cell lines used.

Puromycin: One day after transfection the cells were transferred to 10 cm dishes in non-selective medium. Two days later, with the cells at ~ 10% confluence, fresh medium containing 4 µg/ml and 7 µg/ml of puromycin was provided for HEK293 and CHO-K1 cells respectively. The medium was replaced every 3 days.

This article is protected by copyright. All rights reserved

Accepted Preprint

Geneticin: Five days after transfection CHO cells were replated at 25% confluence in 12-well plates. Selection was performed for 10 days in medium containing 700 µg/ml of Geneticin with medium replacement every day.

6-Thioguanine: The cells that survived the Geneticin selection were split and diluted to a density of 5,000 cells per well in 12-well plates. 6-TG selection was performed in medium containing 4 µg/ml of 6-TG (Sigma-Aldrich); the medium was replaced every 2 days. After 9 days colonies were pooled and used for genomic DNA extraction.

Cell staining with crystal violet: Cells were washed twice with phosphate buffered saline (PBS), fixed for 10 minutes in 3.75% formaldehyde in PBS and stained with 0.1 % crystal violet solution in water for 20 min. Finally, plates were rinsed with tap water and air-dried.

Genomic DNA extraction: Genomic DNA was extracted with the GenElute mammalian genomic DNA miniprep kit (Sigma-Aldrich) following the manufacturer's instructions.

Southern blot: Southern blot analysis was performed according to (Southern 2006). The 600 base pair *Puro* coding sequence was labeled with digoxigenin (DIG) by PCR and used as a probe. Approximately 10 µg of genomic DNA was digested with NsiI (NEB) for 12 hours; the reaction mixture was precipitated with 2.5 M ammonium acetate and 2.5 volumes of ethanol, washed, dried and resolubilized in 20 µl of water. The resulting DNA fragments were electrophoresed on a 0.8 % Tris-borate agarose gel at 15 volts for 16 hours. The fragments were transferred to NYTRAN nylon membranes (Schleicher & Schuell). The membranes were hybridized with the DIG labeled probe (i.e., promoterless *Puro* CDS). The hybridization signals were developed using CDP-Star (Roche

This article is protected by copyright. All rights reserved

Accepted Preprint

Diagnostics) as a substrate for alkaline phosphatase, which is conjugated to the anti-DIG-antibody (Roche Diagnostics).

Oligonucleotides and sequencing: Primers used for PCR and sequencing were ordered from Fisher Scientific and designed so that their melting temperature was close to 60º C and the last nucleotide was A or T. The purchased salt-free oligonucleotides were diluted in DNase- and RNase-free water to a stock concentration of 100 µM. Table I lists all the primers used. Plasmid DNA was sent to Eton Bioscience for sequencing. Results were analyzed using FinchTV and SerialCloner.

Inverse PCR: Genomic DNA isolated from CHO-K1 cells was independently digested with three restriction enzymes: AluI, NlaIII and HpyCH4V (NEB). After heat inactivation of the restriction enzymes, the DNA fragments were circularized using T4 DNA ligase at a diluted concentration of 0.6 ng/µl. Two rounds of PCR were performed using the primers 2 and 3 followed by primers 1 and 4. The PCR products were gel purified, cloned into the pGEMT-EZ vector (Promega), amplified in *E. coli* and sequenced.

Results:

CRISPR-Cas mediated gene insertion in HEK293 cells via NHEJ

In order to introduce the DNA fragment into a defined site in the genome, we developed the following puromycin resistance assay. The strategy was to target and cleave the DNA sequence upstream of a promoter-less *Puro* transgene integrated into the HEK293 genome as well as the sequence downstream of the CMV promoter in a transfecting donor plasmid. As a host, we used a transgenic cell line, HEK293-FW, developed by Arias et al. (2015) that carries a 5300 bp plasmid (pFW) harboring a promoterless puromycin gene (*Puro*) inserted into the PLEKHG1 gene of HEK293 cells (chr.6, 150,823,513). As a donor we used a 4300 bp plasmid (pIC3) containing a CMV promoter (Figure 1).

This article is protected by copyright. All rights reserved

Accepted Preprint

The insertion of the *in vivo* linearized pIC3 plasmid into the DSB upstream of the promoter-less *Puro* gene, with the proper orientation of CMV promoter, will drive the expression of the *Puro* gene thus rendering the cells resistant to puromycin. The use of such a promoter complementation test ensured that random insertion in the genome or transient transcription of the donor plasmid would not yield resistant clones.

In order to introduce the DSB via CRISPR-Cas, we used the 8.5 kb long pX330 expression vector developed in the Zhang laboratory (Ran et al. 2013b). It carries two expression cassettes: a human codon-optimized Cas9 endonuclease derived from *Streptococcus pyogenes* with an SV40 nuclear localization signal and a customizable chimeric guide RNA (gRNA) driven by an RNA Polymerase III promoter. To direct the Cas9 to induce a DSB at a defined site in the pIC3 plasmid as well as in the HEK293 genome, corresponding gRNA sequences were introduced into the BbsI restriction site of the pX330 expression vector using standard cloning protocols. Since gRNAs can have variable activities, two gRNAs were designed to induce each DSB: P1and P2 targeting the plasmid and G1 and G2 (Table I) targeting the HEK293 genome, respectively (Ran et al. 2013a; Ran et al. 2013b).

The constructed pX330 vectors along with the pIC3 donor plasmid were co-transfected into ~50% confluent HEK293 cells. The selection was applied for 8 days, after which the colonies formed were fixed and stained for colony counting. The results of targeting different combinations of genomic DNA and plasmid DNA sequences are shown in Table II.

The P1 and B2 targeting sequences proved to be the most effective. Indeed, P2 was ineffective, suggesting considerable sequence specificity and suggesting that a wider screen of targeting sequences might improve integration frequencies further. Linearizing the donor plasmid with a restriction enzyme before transfection yielded few insertions compared to targeted DSB cleavage, consistent with

This article is protected by copyright. All rights reserved

earlier studies in zebrafish and CHO cells (Auer et al. 2014; Cristea et al. 2013). No colonies appeared in negative controls without genomic DNA cleavage.

In parallel to colony counting, duplicate treated cultures from transfections #1 to #5 were expanded for 3 more days and their DNA extracted for PCR and sequencing. In order to verify the integration of the pIC3 plasmid into the HEK293-FW genome, we amplified the integration junctions by PCR using primers designed to bind ~150 bp upstream and downstream of the putative DSB 3 nucleotides upstream of the PAM sequence (Jinek et al. 2012). PCR products were ligated into the pGEMT-EZ vector via TA cloning and amplified in *E. coli* to obtain cloned integration junctions from the HEK293-FW cells. Almost all (31/32) the junction sequences exhibited small deletions (dashes in Table III) of 1 to 43 nucleotides, with only a single insertion of 15 nt (italicized in Table III). Surprisingly, all four 3' junction sequences from transfection 1 ended in the sequence CCCGGG (underlined in Table IIIB). These hexamer represents the last 6 nt of the 23 nucleotide plasmid target sequence of which the nucleotides GGG are the PAM sequence (Fig. 1). It is not clear from our results whether the cleavage happened immediately after the PAM sequence or that cleavage occurred as expected 3 nucleotides upstream of the PAM (Jinek et al. 2012), followed by a 6 nucleotide insertion of a CCCGGG sequence. We sequenced at least 3 different clones derived from the PCR product of each pool of cells. Surprisingly for each of the individual junctions most of the sequences yielded identical indels. The cause of this uniformity is not clear (see Discussion).

In order to verify that the complete vector was integrated into the HEK293-FW genome in these puromycin resistant trasnfectants, we performed Southern blotting on the genomic DNA extracted from transfections 1 and 3 using the *Puro* CDS as a hybridization probe. As a control we used the untreated host HEK293-FW cells carrying the original pFW plasmid with no insertion. Cleavage of the host DNA with NsiI should yield a 1782 bp fragment from the integrated pFW plasmid sequence, while cleavage after insertion of the donor plasmid should yield a fragment of 6075 bp (Figure 2a). As

This article is protected by copyright. All rights reserved

can be seen in Figure 2b, the Southern blot showed a single band in the wells with transfection #1 and #3 corresponding to the length of ~6kb confirming the integration of the complete plasmid.

CRISPR-Cas mediated plasmid insertion in CHO cells via NHEJ

To test the generality of NHEJ integration we turned to CHO cells. CHO cells are widely used for the commercial production of therapeutic biologics, a process that involves integration of plasmids bearing minigenes specifying recombinant proteins (Cacciatore et al. 2010). CRISPR-Cas was used to target plasmid integration into the hypoxanthine phosphoribosyltransferase locus (*Hprt*) of CHO K1 cells. The idea was to introduce the pFW plasmid coding for the neo gene into the disrupted *Hprt* gene, thus conferring resistance to both G-418 and 6-TG. The X-linked *Hprt* gene is present in only a single copy in CHO cells (Grant and Worton 1989; Xu et al. 2011) and its disruption confers resistance to 6-thioguanine (6-TG), allowing the ready selection of knockouts. The 5.3 kb pFW plasmid, which carries a *neo* gene, was used as the donor. As we did for the HEK293 cells, we transfected the CHO cells with four separate combinations of two CRISPR-Cas plasmids with guide RNAs targeting the CHO *Hprt* gene (pHprt1 and pHprt2) and 2 CRISPR-Cas plasmids targeting the pFW plasmid (pFW1 and pFW2, Table IV).

Transfected cells were first selected for resistance to G418 and this population was subsequently selected for resistance to 6-TG. These cells survived both G418 and 6-TG selection requiring that they had both a *neo* gene and a disrupted *Hprt* gene. The pHprt1 and pFW1 targeting sequences proved to be the most effective (Table IV, rows 1 to 4). Omission of the gRNAs targeting the donor plasmid also yielded some colonies (Table IV, rows 5 and 6), as expected whenever disruption of the *Hprt* gene and the random insertion of the pFW plasmid occurred in the same cell. No colonies appeared in the negative controls without targeting the *Hprt* gene.

This article is protected by copyright. All rights reserved

Accepted Preprint

We used inverse PCR to determine the genomic location of the inserted pFW plasmids in the CHO genome. Genomic DNA samples from transfections 2 and 4 were digested with a 4-cutter restriction enzyme to generate linear fragments (see Materials and Methods). The linear fragments were circularized under conditions that favor intramolecular rather than intermolecular ligations. We used primers specific to the pFW sequence to amplify a fragment that includes the expected plasmid junction and the flanking chromosomal region (Figure 3). PCR products were ligated into pGEMT-EZ (Promega) via TA cloning and amplified in *E. coli* for sequencing.

Transfection 4 yielded 4 sequences: all 4 showed perfect end joining between the targeted plasmid site and the targeted *Hprt* site (CHOgenome.org scaffold NW_003613932.1 at nucleotide 1137989); both orientations of the plasmid were represented. The sequences from transfection 2 mapped onto the genomic scaffold NW_003613932.1 at nucleotide 1129841, which is ~2 kb downstream of the DSB targeted by pHprt1 gRNA. This integration site bears no similarity to the intended target. It is not clear if this integration represents a nearby off target event or if the cleavage happened at the targeted DSB, with a subsequent ~2kb deletion event. In addition to these sequences each of the two transfections yielded an off target integration (Fu et al. 2013; Hsu et al. 2013) each of which unexpectedly included plasmid sequences from untargeted regions. Since these cells are resistant to 6-TG, it is likely that they all contain disruptions in the *Hprt* gene; the off target integrations would then be occurring in addition to the targeted integrations.

In order to compare the efficiency (transfectant colonies per treated cell) of the NHEJ insertion in CHO cells to that in HEK293 cells, we used the same *Puro* gene reconstruction strategy employed in the HEK293-FW cells. As a target we used the population of G418 and 6-TG resistant CHO-K1 cells from transfection 4 in the experiment described above. We picked the guide RNAs that gave the highest efficiency in HEK293-FW cells, i.e., G1 and P2, to target the upstream region of the *Puro*

This article is protected by copyright. All rights reserved

Accepted Preprint

coding sequence present in the integrated pFW plasmid sequence (i.e., now genomic) and the downstream region of the CMV promoter present in pIC3. The CRISPR-Cas vectors along with the pIC3 donor plasmid were co-transfected into ~70% confluent CHO-K1 cells. After 3 days of expression in non-selective medium, selection for puromycin resistance was applied for 5 days, after which the colonies formed were fixed and stained for counting. The results of different combinations of genomic DNA and plasmid DNA targets are shown in Table V.

We observed 2700 colonies in the sample treated with the guide RNAs targeting both the genomic DNA and donor plasmids, representing an efficiency of 0.45 %, about 3 times higher than the 0.17% achieved with HEK293-FW cells and identical guide RNAs. It is interesting to note that we found ~500 colonies (an efficiency of 0.08%) by targeting the genomic *Puro* gene but not the donor plasmid, suggesting that the CMV-bearing plasmid may undergo cleavage inside the cell and that many of these linearized plasmids were able to provide the necessary promoter activity to the genomic *Puro* CDS. It should be noted, however, that no such untargeted promoter provision was seen using HEK293-FW cells.

To determine if the pIC3 plasmid had indeed integrated into the targeted genomic promoterless *Puro* gene, we amplified the predicted integration junctions by PCR as described earlier; the PCR products were ligated into pGEMT-EZ vectors, cloned in *E. coli* and sequenced. The clones from both the junctions resulted in insertions and deletions of up to 20 nucleotides (shown in italicized letters and dashes respectively in Table VI). Similar to the sequences obtained from HEK293-FW cells, many of the clones sequenced from the 2 junctions were identical.

This article is protected by copyright. All rights reserved

Discussion:

In this study, we have demonstrated for the first time that a CRISPR-Cas induced DSB can be used to insert a ~5 kb long foreign DNA sequence into a predetermined target site in the genome of mammalian cells via the NHEJ pathway. Several attempts have been made to improve the ratio of the HR to the NHEJ repair pathway owing to the former's high precision. Knocking out DNA ligase IV, a key component of the major NHEJ pathway in *Drosophila* greatly promotes HR directed repair (Beumer et al. 2008; Bozas et al. 2009). Similarly knockout of DNA ligase IV in *C*. *elegans*, which can tolerate the loss of this enzyme, also increased HR directed repair (Morton et al. 2006). However, NHEJ is still the dominant repair pathway in mammalian cells and has been efficiently used to insert DNA fragments into DSBs generated by ZFNs and TALENs. Cristea et al. (2013) showed that incorporating a nuclease target site into the donor plasmid followed by concurrent *in vivo* nuclease cleavage of the donor plasmid and the chromosome using ZFNs and TALENs in CHO cells resulted in the efficient integration of the donor plasmid into the chromosomal DSB. More recently the insertion of plasmid DNA into DSBs generated via CRISPR-Cas was demonstrated in zebrafish (Auer et al. 2014; Maresca et al. 2013). Here we used the simple and efficient CRISPR-Cas system to induce the DSB in the donor plasmid as well as the genome of HEK293 and CHO-K1 cells, which resulted in the integration of the donor plasmid into the chromosomal DSB.

For insertion at the *Puro* sequence, we achieved maximum efficiencies of 0.17% in HEK293 cells and 0.45% in CHO cells. However, some of the tested guide RNAs showed poor efficiencies in promoting NHEJ integration; thus several gRNA sequences should be assayed. Limitations on efficiency may be due to the need to target two DNA molecules, i.e., the genome and the plasmid; and the fact that only one of two possible insert orientations of the CMV promoter was being selected. We found that linearization of the plasmid *in vitro* prior to transfection was much less efficient than in vivo linearization by CRISPR-Cas. A lower yield of integration using a linearized version of the donor

This article is protected by copyright. All rights reserved

Accepted Preprint

plasmid has also been recently reported in other systems (Auer et al. 2014; Cristea et al. 2013). Possible reasons for this difference could be that transfection efficiency is higher for circular or super-coiled DNA and that *in vitro*-linearized DNA may be more prone to cellular nuclease activity (Cristea et al. 2013; Stuchbury and Munch 2010). Moreover, the main disadvantage associated with transfecting circular DNA – the inability to control exactly where cleavage occurs – is circumvented by this CRISPR-Cas9 approach.

Inverse PCR experiments on the 6-TG resistant CHO-K1 transfectants showed that the pFW plasmid became incorporated into off-target sites as well as the targeted sites. Blasting the gRNA sequences against the scaffolds obtained from these off-target sites yielded no sequence similarities, suggesting that these insertions may be occurring through random integrations or through rearrangements subsequent to targeted integration.

Sequencing results obtained from the 5' and the 3' junctions of the transfected HEK293-FW cells were somewhat surprising in that most of the sequences displayed identical indels, even when taken from independent transfections. In this regard it should be noted that the genomic and plasmid sequences targeted in these experiments are located within phiC31 Att P and Att B sites, respectively, reflecting a previous use of pFW (Arias et al., 2015). Although no phiC31 recombinase is present in these experiments, it is conceivable that phage attachment sites exhibit some sort of inherent peculiar recombinational behavior, perhaps related to their evolutionary history as substrates for site-specific recombination.

Lastly, in transfections that used the gRNA combination of G1 and P1, the 3 nt PAM sequence as well as the 20 nt gRNA sequence of the plasmid were integrated into the 3' junctions and were correspondingly missing from the 5' junctions. This result suggests that Cas9 cleavage occurred at nt

This article is protected by copyright. All rights reserved

23 with this gRNA, just after the PAM sequence (Table IIIB) rather than after position 17 (Jinek et al. 2012); we are not aware of any previous report of such noncanonical cleavage. These results could also be explained by the insertion and deletion of CCCGGG sequences corresponding to gRNA positions 18 to 23, which we consider unlikely.

Overall, our results demonstrate the feasibility of CRISPR-Cas mediated gene insertion in HEK293 and CHO cells without the need of homologous sequence arms. While the efficiency (colonies per treated cell) of integration was not very high, it is high enough to make recombinant clone isolation without a selective marker feasible; these data encourage us to seek conditions that may improve on this frequency. Even as is, this system may prove useful for the genetic engineering of cultured mammalian cells for the production of high levels of recombinant proteins. For instance, one could target a highly active endogenous gene so as to expropriate its promoter for the production of a protein of interest. Although these ends could be achieved by homologous recombination, the use of NHEJ is simpler to implement as the plasmid vectors would be more easily constructed, or indeed extant plasmids could be used.

Acknowledgements

We thank Vincent Anquetil and Mauricio Arias for very helpful discussions and important assistance with these experiments. We are grateful to Mauricio Arias for the generous provision of the pFW plasmid and cell line HEK293-FW. This work was funded by ImClone Systems, a wholly owned subsidiary of Eli Lilly and Company. The authors declare no conflict of interest in this work.

This article is protected by copyright. All rights reserved

Accepted Preprint

References:

Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F. 2014. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res 24(1):142-53.

Bassett AR, Tibbit C, Ponting CP, Liu JL. 2013. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep 4(1):220-8.

Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D. 2008. Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc Natl Acad Sci U S A 105(50):19821-6.

Bozas A, Beumer KJ, Trautman JK, Carroll D. 2009. Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics 182(3):641-51.

Cacciatore JJ, Chasin LA, Leonard EF. 2010. Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the Dhfr-based CHO cell selection system. Biotechnol Adv 28(6):673-81.

Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, Liaw L, Cui X. 2010. Targeted genome modification in mice using zinc-finger nucleases. Genetics 186(2):451-9.

Carroll D. 2014. Genome engineering with targetable nucleases. Annu Rev Biochem 83:409-39.

Chang N, Sun C, Gao L, Zhu D, Xu X, Zhu X, Xiong JW, Xi JJ. 2013. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res 23(4):465-72.

Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31(3):230-2.

Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, Kuhn R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA and others. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819-23.

Cristea S, Freyvert Y, Santiago Y, Holmes MC, Urnov FD, Gregory PD, Cost GJ. 2013. In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. Biotechnol Bioeng 110(3):871-80.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31(9):822-6.

Gaj T, Gersbach CA, Barbas CF, 3rd. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol 31(7):397-405.

Galfre G, Milstein C. 1982. Chemical typing of human kappa light chain subgroups expressed by human hybrid myelomas. Immunology 45(1):125-8.

This article is protected by copyright. All rights reserved

Grant SG, Worton RG. 1989. Activation of the hprt gene on the inactive X chromosome in transformed diploid female Chinese hamster cells. J Cell Sci 92 (Pt 4):723-32.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. 2013. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194(4):1029-35.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O and others. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 31(9):827-32.

Huertas P. 2010. DNA resection in eukaryotes: deciding how to fix the break. Nat Struct Mol Biol 17(1):11-6.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31(3):227-9.

Iyama T, Wilson DM, 3rd. 2013. DNA repair mechanisms in dividing and non-dividing cells. DNA Repair (Amst) 12(8):620-36.

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. 2013. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Res 41(20):e188.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337(6096):816-21.

Joung JK, Sander JD. 2013. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14(1):49-55.

Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A 93(3):1156-60.

Lieber MR. 1999. The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. Genes Cells 4(2):77-85.

Mali P, Esvelt KM, Church GM. 2013a. Cas9 as a versatile tool for engineering biology. Nat Methods 10(10):957-63.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013b. RNA-guided human genome engineering via Cas9. Science 339(6121):823-6.

Maresca M, Lin VG, Guo N, Yang Y. 2013. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. Genome Res 23(3):539-46.

Morton J, Davis MW, Jorgensen EM, Carroll D. 2006. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in Caenorhabditis elegans somatic cells. Proc Natl Acad Sci U S A 103(44):16370-5.

This article is protected by copyright. All rights reserved

Pan Y, Xiao L, Li AS, Zhang X, Sirois P, Zhang J, Li K. 2013. Biological and biomedical applications of engineered nucleases. Mol Biotechnol 55(1):54-62.

Pardo B, Gomez-Gonzalez B, Aguilera A. 2009. DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci 66(6):1039-56.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y and others. 2013a. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154(6):1380-9.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013b. Genome engineering using the CRISPR-Cas9 system. Nat Protoc 8(11):2281-308.

Ronda C, Pedersen LE, Hansen HG, Kallehauge TB, Betenbaugh MJ, Nielsen AT, Kildegaard HF. 2014. Accelerating genome editing in CHO cells using CRISPR Cas9 and CRISPy, a web-based target finding tool. Biotechnol Bioeng 111(8):1604-16.

Rouet P, Smih F, 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(12):8096-106.

Sedivy JM, Sharp PA. 1989. Positive genetic selection for gene disruption in mammalian cells by homologous recombination. Proc Natl Acad Sci U S A 86(1):227-31.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG and others. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343(6166):84-7.

Southern E. 2006. Southern blotting. Nat Protoc 1(2):518-25.

Stuchbury G, Munch G. 2010. Optimizing the generation of stable neuronal cell lines via pre-transfection restriction enzyme digestion of plasmid DNA. Cytotechnology 62(3):189-94.

Urlaub G, Chasin LA. 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc Natl Acad Sci U S A 77(7):4216-20.

Urlaub G, Kas E, Carothers AM, Chasin LA. 1983. Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. Cell 33(2):405-12.

Waaijers S, Portegijs V, Kerver J, Lemmens BB, Tijsterman M, van den Heuvel S, Boxem M. 2013. CRISPR/Cas9-targeted mutagenesis in Caenorhabditis elegans. Genetics 195(3):1187-91.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. 2013. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153(4):910-8.

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S and others. 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29(8):735-41.

This article is protected by copyright. All rights reserved

Yamane-Ohnuki N, Kinoshita S, Inoue-Urakubo M, Kusunoki M, Iida S, Nakano R, Wakitani M, Niwa R, Sakurada M, Uchida K and others. 2004. Establishment of FUT8 knockout Chinese hamster ovary cells: an ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. Biotechnol Bioeng 87(5):614-22.

Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell 154(6):1370-9.

Accepted Preprint

This article is protected by copyright. All rights reserved

Tables

Table I. CRISPR-Cas target sequences and primer sequences

Plasmid targets:

```
          _________P1__________PAM
GAATTGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGAATTCTAAGGGTTTCA
CTTAACCACGGTCCCGCACGGGAACCCGAGGGGCCCGCGCTTAAGATTCCCAAAGT
                                  PAM               P2
```

Genomic targets:

```
        ________G1_____________PAM ___________G2_________PAM
CTGGAGCAACACTCGAGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGTC
GACCTCGTTGTGAGCTCGGGGTTGACCCCATTGGAAACTCAAGAGAGTCAACCCCCAG
```

Primers:

Junction primers: (nucleotides in capital letters are complementary to pFW)

1) 5' Junction forward: attcaagggcccGATGAAAGTTGCCTGGAATGGA
2) 5' Junction reverse: tgtattgtcgaCTGATCCACTGATAGGATTACAAGT
3) 3' Junction forward: tgatatgggcccGGAGTTTGTTTTGGCACCAAAAT
4) 3' Junction reverse: aagtcagtcgacGTGGCGGTCCGGATCGA

Inverse PCR primers (see Fig. 3):

1) GCGCAGCGTGACCGCTA
2) CGGCGGGTGTGGTGGTT
3) GCGCGTCAGGTGGCACT
4) AAATGTGCGCGGAACCCCTA

Accepted Preprint

This article is protected by copyright. All rights reserved

Accepted Preprint

Table II. CRISPR-Cas mediated insertions at the Puro sequence in HEK293 cells

Transfection#	CRISPR gRNAs genomic DNA site	CRISPR gRNAs plasmid DNA site	Donor plasmid (pIC3) form[a]	Colonies[a]	Efficiency[c]
1	G1	P1	circular	160	0.08 %
2	G1	P2	circular	330	0.17 %
3	G2	P1	circular	12	0.01 %
4	G2	P2	circular	17	0.01 %
5	G1	-	linearized	42	0.02 %
6	G2	-	linearized	1	0.00 %
7	-	P2	circular	0	0.00 %
8	-	-	-	0	0.00 %

[a] Linearization was accomplished cleavage at a unique EcoRI restriction site in the plasmid.
[b] Puromycin-resistant colonies arising from a dish seeded with $2x10^5$ treated cells.
[c] Colonies per treated cell.

This article is protected by copyright. All rights reserved

Table III. Plasmid-genome junction sequences in HEK293 cells[a]

3A: 5' Junctions

Transfection 1[b]: genomic gRNA G1, plasmid gRNA P1

Perfect 5' junction:	atcatgctggagcaacactcgagccccaaCCCGGGCGCGAATTCTAAGGGTTTCAACTTG
L2-1(2)	atcatgctggagcaacac----------AGAATTCTAAGGGTTTCAACTTG
L2-2	atcatgctgga----------AGAATTCTAAGGGTTTCAACTTG
L2-3	atcaggctggagcaacactcgagccccaa*TGCCATCCGTAAGAT*CCCGGGCGCGAATTCT
	AAGGGTTTCAACTTG

Transfection 2: genomic gRNA G1, plasmid gRNA P2

Perfect 5' junction:	acatcatgctggagcaacactcgagccccaaCGCGAATTCTAAGGGTTTCAACTTGTTTG
L4-1(3)	acatcatgctggagcaacac----------AGAATTCTAAGGGTTTCAACTTGTTTG

Transfection 3: genomic gRNA G2, plasmid gRNA P1

Perfect 5' junction:	aacactcgagccccaactggggtaacctttgagttctctcCCCGGGCGCGAATTCTAAGG
L6-1 (3)	aacac----------AGAATTCTAAGG

Transfection 4: genomic gRNA G2, plasmid gRNA P2

Perfect 5' junction:	agcaacactcgagccccaactggggtaacctttgagttctctcCGCGAATTCTAAGGGTT
L8-1	agcaacac----------GGAATTCTAAGGGTT
L8-2(2)	agcaacac----------AGAATTCTAAGGGTT

Transfection 5: genomic gRNA G1, *in vitro* linearized plasmid

Perfect 5' junction:	catgctggagcaacactcgagccccaaAATTCTAAGGGTTTCAACTTGTTTGGATAGTGA
L10-1(3)	catgctggagcaacac---------A*G*AATTCTAAGGGTTTCAACTTGTTTGGATAGTGA

3B: 3' Junctions

Transfection 1: genomic gRNA G1, plasmid gRNA P1

Perfect 3' junction:	GGTGCCAGGGCGTGCCCTTGGGCTCctggggtaacctttgagttctctcagttgggggtc
R2-2[c] (4)	GGTGCCAGGGCGTGCCCTTGGGCTCCCCGGG----------agttgggggtc

Transfection 2: genomic gRNA G1, plasmid gRNA P2

Perfect 3' junction:	CAGGGCGTGCCCTTGGGCTCCCCGGGctggggtaacctttgagttctctcagttgggggt
R4-1 (3)	CAGGGCGTGCCCTTGGGCTCCCCGGG----------agttgggggt

Transfection 3: genomic gRNA G2, plasmid gRNA P1

Perfect 3' junction:	ATTGGTGCCAGGGCGTGCCCTTGGGCTCcagttgggggtcgaaccaccatgaccgag
R6-1 (3)	ATTGGTGCCAGGGCGTGCCCTTGGGCTC-agttgggggtcgaaccaccatgaccgag

Transfection 4: genomic gRNA G2, plasmid gRNA P2

Perfect 3' junction:	CCAGGGCGTGCCCTTGGGCTCCCCGGGcagttgggggtcgaaccaccatgaccgag
R8-1 (2)	CCAGGGCGTGCCCTTGGGCTCCCCGGG-agttgggggtcgaaccaccatgaccgag
R8-2	CCAGGGCGTGCCCTTGGGCTCCCCGG----------*C*cgaaccaccatgaccgag

Transfection 5: genomic gRNA G1, *in vitro* linearized plasmid

Perfect 3' junction:	GCCCTTGGGCTCCCCGGGCGCGAATTctggggtaacctttgagttctctcagttgggg
R10-1	GCCCTTGGGCTC----------agttgggg
R10-2 (2)	GCCCTTGGGCTCCCCGGG----------agttgggg

[a] Genomic sequences are in lower case, donor plasmid sequences are in upper case, deletions are represented as dashes and inserted bases are italicized. The predicted junction sequences were based on the CRISPR-Cas system cutting 3 nucleotides before the PAM sequence with no indels. The number in parentheses is the number of *E. coli* clones with the indicated sequence.

[c] The CCCGGG sequence may represent an unexpected cleavage point for Cas9 at the 3' end of the PAM sequence.

[b] See Table II for transfection numbering.

This article is protected by copyright. All rights reserved

Accepted Preprint

Table IV. CRISPR-mediated insertions at the *Hprt* locus in CHO-K1 cells

Transfection	gRNAs targeting the genomic DNA site	gRNAs targeting the plasmid DNA site	Donor plasmid	Colonies
1	pHprt1	pFW1	pFW	+++[a]
2	pHprt1	pFW2	pFW	++
3	pHprt2	pFW1	pFW	++
4	pHprt2	pFW2	pFW	++
5	pHprt1	-	pFW	+
6	pHprt2	-	pFW	+
7	-	pFW1	pFW	0
8	-	-	pFW	0
9	-	-	-	0

[a] Colony counts were estimated by microscopic examination.

This article is protected by copyright. All rights reserved

Table V. CRISPR-Cas mediated insertions at the Puro sequence s in CHO-K1 cells

Transfection	CRISPR gRNAs genomic DNA site	CRISPR gRNAs plasmid DNA site	Donor plasmid	Colonies[a]	Efficiency[b]
1	G1	P2	pIC3	2700	0.45 %
2	G1	-	pIC3	505	0.08 %
3	-	P2	pIC3	42	0.01 %
4	G1	-	-	54	0.01 %
5	-	-	pIC3	30	0.01 %
6	-	-	-	29	0.01 %

[a] Puromycin-resistant colonies arising from a dish seeded with 6 x10^5 treated cells.
[b] Colonies per treated cell.

Accepted Preprint

This article is protected by copyright. All rights reserved

Table VI. Plasmid-genome junction sequences at the *Puro* sequence in CHO cells[a]

5' Junctions: genomic gRNA G1, plasmid gRNA P2	
Perfect junction:	acatcatgctggagcaacactcgagccccaaCGCGAATTCTAAGGGTTTCAACTTGTTTG
Clone L1-1 (2	acatcatgctggagcaacactcgagccccaaCGCGAATTCTAAGGGTTTCAACTTGTTTG
Clone L1-2	acatc--------------------------CGCGAATTCTAAGGGTTTCAACTTGTTTG
Clone L1-3	acatcaggctggagcaacactcgagccccaaa*CGCGAATTCTAAGGGTTTCAACTTGTTTG*
3' Junctions: genomic gRNA G1, plasmid gRNA P2	
Perfect junction:	GGGCGTGCCCTTGGGCTCCCCGGGctggggtaacctttgagttctctcagttgggggt
Clone R1-1[b] (3)	GGGCGTGCCCTTGGGCTCCCCGGGctggggtaacctttgagttctctcagttgggggt
Clone R1-2[b]	GGGCGTGCCCTTGGGCTCCCCGGG*GCCGAGCGCAGAAGTGGTCCTGCAACTTTATC*
	*CGC*ctggggtaacctttgagttctctcagttgggggtcg

[a]Genomic sequences are in lower case, donor plasmid sequences are in upper case, deletions are represented as dashes and inserted bases are italicized. The predicted junction sequences were based on the CRISPR-Cas system cutting 3 nucleotides upstream of the PAM sequence with no indels. The number in parentheses is the number of *E. coli* clones with the indicated sequence.

Accepted Preprint

This article is protected by copyright. All rights reserved

Accepted Preprint

Figure Legends

Figure 1. Promoter complementation selection. gRNAs were designed to target the region upstream of the promoter-less *Puro* gene provided by pFW (grey rectangles) in the HEK293 genome (solid line) and the region downstream of the CMV promoter on the pIC3 plasmid (white rectangles). The gRNA target sequences in the genome and in the plasmid are shown by thick vertical black lines representing the 2 different sequences at each of the 2 target sites. This assay ensures that only site-specific integration in proper orientation can give rise puromycin resistance. Protospacer adjacent motif (PAM) bases are in grey and the expected DSBs 3 nucleotides upstream to the PAM are shown as dotted lines.

Figure 2. Integration of pIC3 into the HEK293 genome. A) The map expected for pIC3 integrated into the targeted site in the HEK293 genome is shown. Cleavage sites for the restriction enzyme NsiI used for the Southern blot experiment are indicated. The probe used against the *Puro* CDS is shown as a stippled line, the pre-existing pFW plasmid as grey rectangles, the incoming pIC3 as white rectangles, and genomic DNA as a solid line. B) Southern blot performed on HEK293-FW cell DNA using the *Puro* probe. The expected band at 1782 bp can be seen in the untransfected control lane. DNA from transfection 1 and transfection 3 yielded the 6075 bp band expected for pIC3 integration. Transfections 1 and 3 differed only in the gRNA used to target the genomic DNA sequence: G1 and G2 plasmids were used respectively.

Figure 3. Inverse PCR strategy used to determine the chromosomal insertion site of pFW plasmids targeted into the *Hprt* gene in CHO K1 cells. Genomic DNA of pooled transfectants was digested with a restriction enzyme that cuts pFW within ~300 to ~500 nt from the targeted plasmid site and at unknown sites in the genomic flanks. The DNA fragments were then circularized with T4 DNA ligase. The flanking genomic sequences were amplified by nested PCR using primers 2 and 3 followed by primers 1 and 4.

This article is protected by copyright. All rights reserved

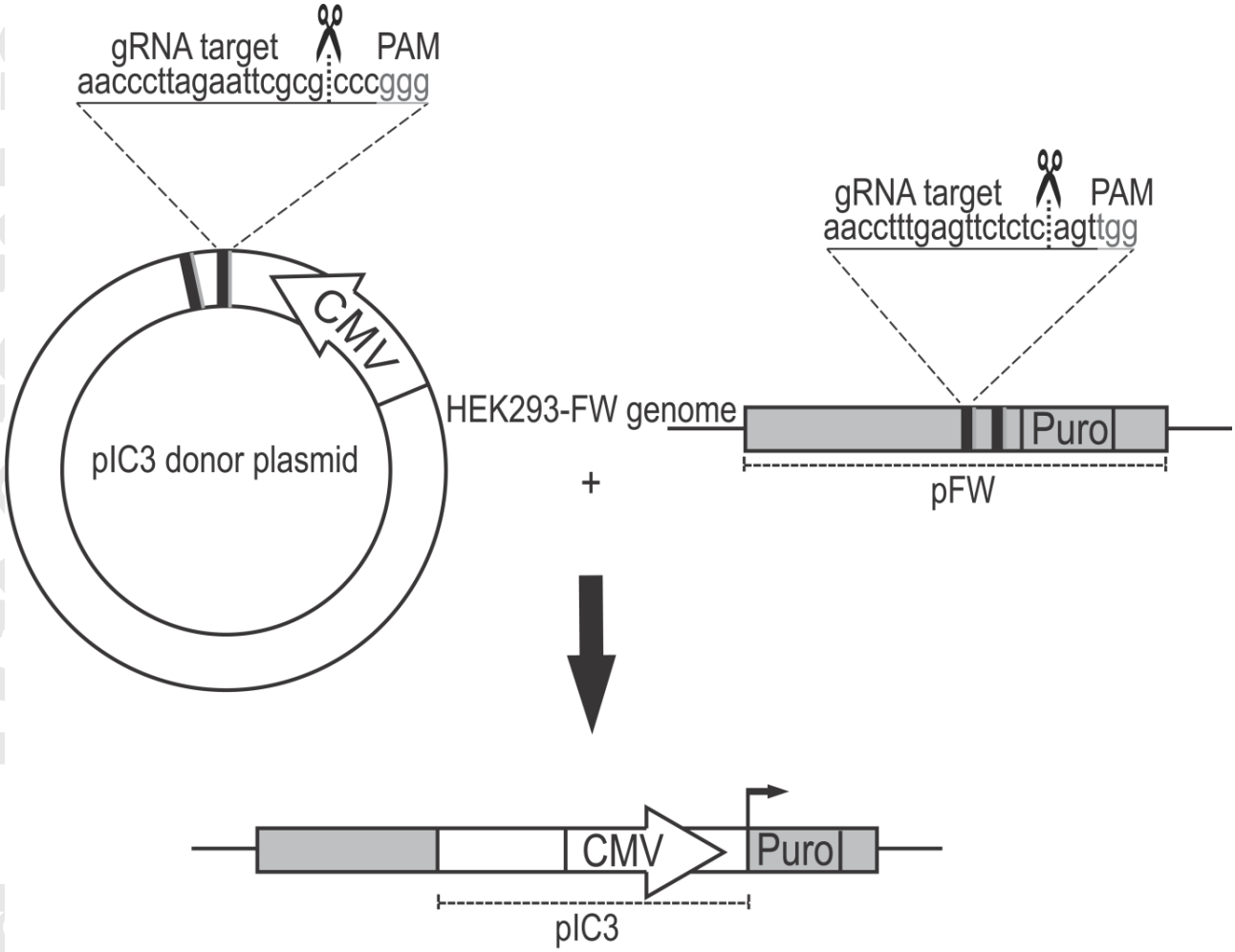


Figure 1

Accepted manuscript

This article is protected by copyright. All rights reserved

Accepted Preprint

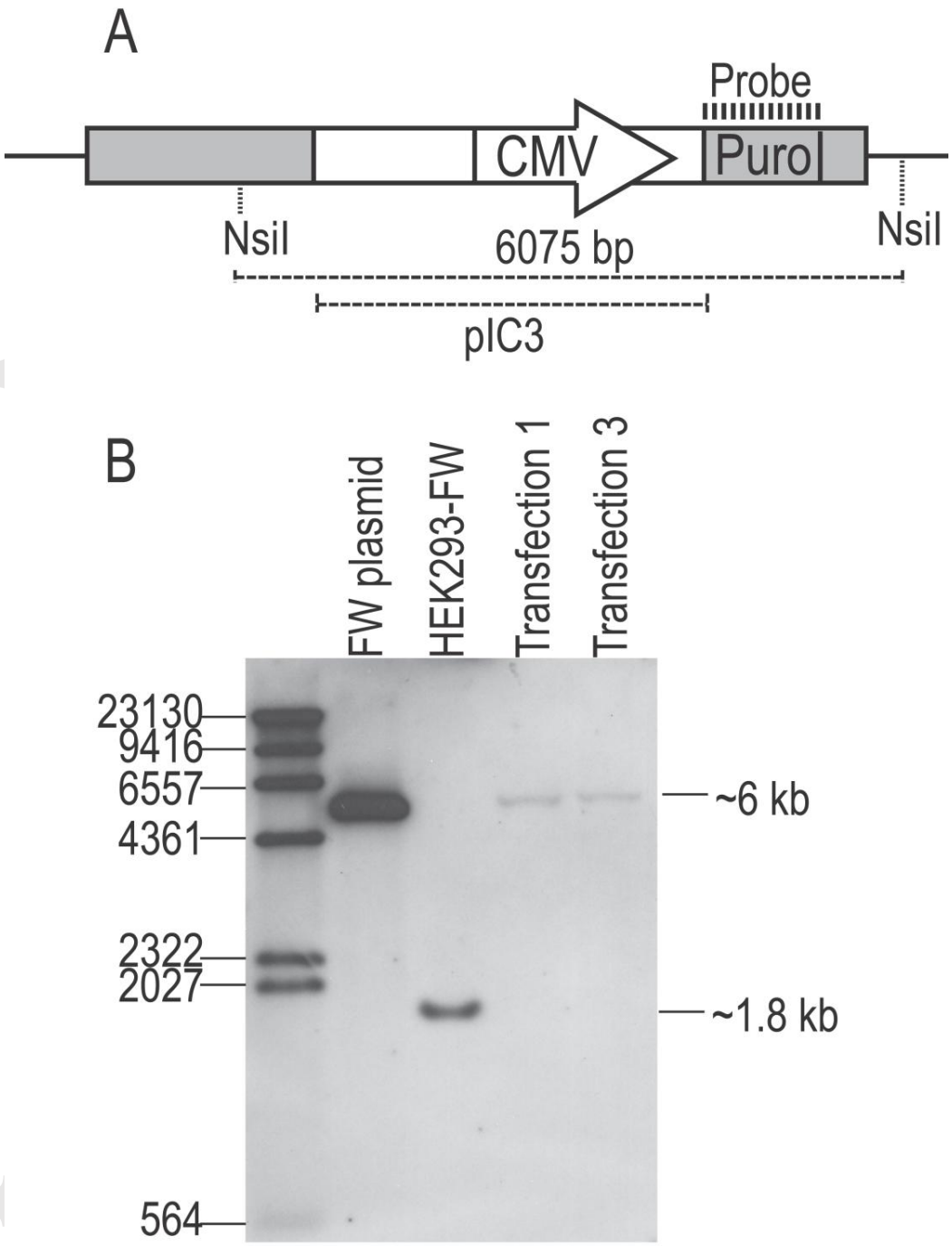


Figure 2

This article is protected by copyright. All rights reserved

Accepted Preprint

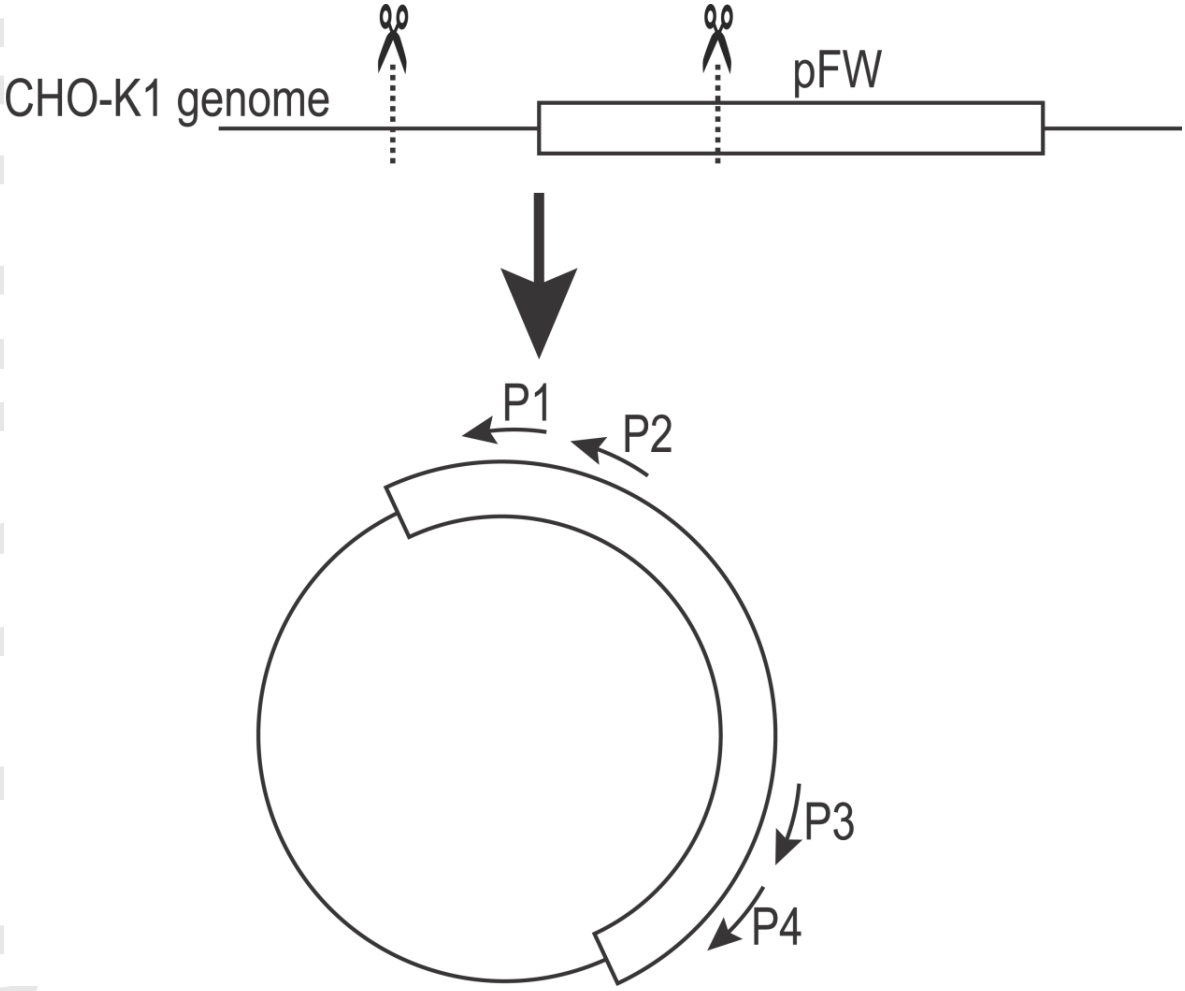


Figure 3

This article is protected by copyright. All rights reserved