DNA Research, 2017, 24(6), 609–621
doi: 10.1093/dnares/dsx029
Advance Access Publication Date: 30 June 2017
Full Paper

OXFORD

Full Paper

A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a 'self-cleaving' GFP-expression plasmid

András Tálas[1,2], Péter István Kulcsár[2,3,4], Nóra Weinhardt[2,3,4], Adrienn Borsy[2], Eszter Tóth[2,3], Kornélia Szebényi[2], Sarah Laura Krausz[1,2], Krisztina Huszár[2,3], István Vida[2,5], Ádám Sturm[2], Bianka Gordos[2], Orsolya Ivett Hoffmann[6], Petra Bencsura[2,3], Antal Nyeste[2,3], Zoltán Ligeti[2], Elfrieda Fodor[3], and Ervin Welker[2,3,*]

[1]School of Ph.D. Studies, Semmelweis University, Budapest, Hungary, [2]Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, Hungary, [3]Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary, [4]University of Szeged, Szeged, Hungary, [5]Institute of Organic Chemistry, Eötvös Loránd University, Budapest, Hungary, and [6]Animal Biotechnology Section, Ruminant Genome Biology Group, NARIC Agricultural Biotechnology Institute, Gödöllő, Hungary

*To whom correspondence should be addressed. Tel. +361 382 6610. Email: welker.ervin@ttk.mta.hu

Edited by Dr. Toshihiko Shiroishi

Received 29 November 2016; Editorial decision 27 May 2017; Accepted 7 June 2017

Abstract

The efficacies of guide RNAs (gRNAs), the short RNA molecules that bind to and determine the sequence specificity of the *Streptococcus pyogenes* Cas9 nuclease, to mediate DNA cleavage vary dramatically. Thus, the selection of appropriate target sites, and hence spacer sequence, is critical for most applications. Here, we describe a simple, unparalleled method for experimentally pre-testing the efficiencies of various gRNAs targeting a gene. The method explores NHEJ-cloning, genomic integration of a GFP-expressing plasmid without homologous arms and linearized in-cell. The use of 'self-cleaving' GFP-plasmids containing universal gRNAs and corresponding targets alleviates cloning burdens when this method is applied. These universal gRNAs mediate efficient plasmid cleavage and are designed to avoid genomic targets in several model species. The method combines the advantages of the straightforward FACS detection provided by applying fluorescent reporter systems and of the PCR-based approaches being capable of testing targets in their genomic context, without necessitating any extra cloning steps. Additionally, we show that NHEJ-cloning can also be used in mammalian cells for targeted integration of donor plasmids up to 10 kb in size, with up to 30% efficiency, without any selection or enrichment.

Key words: Cas9, CRISPR, gRNA testing, reporter assay, NHEJ-cloning

1. Introduction

The use of RNA-guided nucleases revolutionized our ability to modify complex genomes and to control gene expression networks, in the same manner that PCR revolutionized molecular biology at the end of the 20th century.[1–10] The specificity of Cas9 nucleases is determined in part by the so-called spacer sequence of the guide RNA (gRNA) associated to the protein and responsible for target recognition. This makes the Cas9 proteins easily reprogrammable by simply varying the sequence of the spacer.[1] This attractive feature of the Cas9 nucleases has been exploited for genome editing and gene expression modulation, as well as for generating disease and therapeutic models.[7,9,11–15]

The efficiencies of various gRNA spacers to result targeted DNA cleavage by Cas9 vary widely. While several sequence determinants of efficient target recognition for *Streptococcus pyogenes* Cas9 (SpCas9), have been revealed,[16–21] an adequate prediction of the efficiency of a given spacer sequence has not been achieved, partially because the activity of a gRNA may be modulated by the genomic context of the target: a factor difficult to predict.[22,23] Generally, several candidate spacer sequences are tested before choosing the appropriate ones for performing the desired genetic modifications. Existing methods to test the efficiency of a given spacer sequence commonly measure the sequence alterations acquired during the repair of the cleaved DNA.[24] Breaks in the DNA involving one or both strands are effectively repaired by the cell using one of the two main repair pathways: the homologous recombination (HR) or the more error-prone non-homologous end-joining (NHEJ).[25] NHEJ-dependent repair of double strand breaks frequently leads to small insertions or deletions (indels), the frequency of which can be explored to monitor the efficiency of Cas9 cleavage.[26–28] Accordingly, many methods rely on the assessment of indel frequencies that are usually determined by PCR amplification of the corresponding region followed by Surveyor/T7E1 nuclease assays,[29] high resolution melt analysis,[30,31] fluorescent PCR-capillary gel electrophoresis[32] and direct[33] or deep[11,12,34] sequencing. Alternatively, a large number of clones can be sequenced upon the PCR amplification of the regions of interest. The occurrence of HR events as a result of HR repair can also be monitored by analysing clones using PCR followed by restriction digestion[35] or by sequencing,[34] as well as by Southern blotting[36–38] of genomic DNA. As another approach, reporter assays are explored for detecting HR events: (I) where a fluorescent signal is measured as a result of recombination events correcting a truncated fluorescent protein;[39,40] (II) or where a fluorescent protein expression cassette is incorporated at the cleavage site - at the cost of the laborious construction of homologous arms.[41] In case of exploiting NHEJ repair, reporter assays are also frequently employed for monitoring indel events that alter the reading frame of a fluorescent protein, resulting in either loss or recovery of the fluorescent signal.[35,39,42] Reporter assays that enable the use of fluorescence activated cell sorting (FACS) to monitor the repair of double strand breaks are convenient, making genomic DNA isolation and PCR amplification -including the sometimes very tedious condition/primer optimization- unnecessary. They also provide more accurate estimations than Surveyor/T7E1 assays and are much less expensive than deep sequencing.[24] However, these approaches generally require extra cloning steps to become applicable to a particular task and lose the advantage of PCR-based methods of being capable of monitoring spacer efficiency in the genomic context of the particular targets.

Here, we introduce a reporter assay, which requires no additional cloning steps and is capable of testing spacer efficiencies on targets within their genomic context. The method explores NHEJ-cloning (NHEJ repair mediated integration) of a GFP-expression cassette to a target site that is cleaved by SpCas9. The key feature of this approach is the use of a 'self-cleaving' plasmid that enhances targeted integration.

Using SpCas9 and a self-cleaving plasmid, we also demonstrate a very convenient and effective way of inserting relatively large DNA fragments into the mammalian genome. NHEJ-cloning has been described in zebrafish, using transcription activator-like effector nuclease (TALEN) and SpCas9, and in mammalian cells, using zinc-finger nuclease (ZFN) and TALEN.[38,43,44] There have been a few attempts to use SpCas9 for NHEJ-cloning of larger DNA cassettes in mammalian cells as well,[45,46] however, they achieved integrations at a relatively low frequency, typically under 1% without selection or FACS-enrichment. Here, we achieved a frequency that is greater by over an order of magnitude. Thus, this approach facilitates the implementation of several genome engineering tasks in mammalian systems, making them less labor-intensive, especially so regarding the insertion of expression cassettes to safe-harbour locations in contrast with the use of random integration.

2. Materials and methods

2.1. Materials

All restriction enzymes, Klenow polymerase and T4 ligase were acquired from Thermo Scientific, Q5 and Phusion polymerases T7 endonuclease I, and Gibson Assembly Master Mix were from New England Biolabs Inc. Oligonucleotides were purchased from Microsynth AG and Sigma-Aldrich Co. and their sequences are listed in the Supplementary Data. Dulbecco's modified Eagle's Medium, foetal bovine serum, Turbofect, Lipofectamine 2000, GlutaMAX, penicillin, streptomycin and puromycin were acquired from Thermo Fisher Scientific, and trimethoprim (TMP) from Sigma-Aldrich Co. (T7883).

2.2. Vector constructions

Briefly, SpCas9-gRNA expressing plasmids (targeting human and mouse genes and also containing TL1 and TL2 spacers) were generated from the px330 vector (Addgene # 42230[3]), and cloning spacers into pmCherry_gRNA plasmid was conducted according to Hanhui Ma et al.[47] The construction of pmCherry_gRNA, pX330-Flag-ehSpCas9 and pX330-Flag-wtSpCas9 is described in (Kulcsár et al. 2017 submitted). The nuclease inactive SpCas9 (dCas9) expressing plasmid was created from px335 (Addgene #42335[3]) by using Body Double cloning method.[48] The self-cleaving plasmids were created by using Gibson assembly method. The self-cleaving plasmid containing a puromycin cassette (pSc1-puro) was constructed by using pLKO.1puro plasmid (Addgene #8453[49]). The self-cleaving plasmid with a degradation domain (pSc1-DD) was created from pBMN DHFR(DD)-YFP plasmid (Addgene #29325[50]). The *Prnp* gene targeting homologous recombination donor plasmid was created by using nested PCR. The *Rosa26* locus homologous recombination donor plasmid was created by using pDonor-MCS-rosa26 plasmid [Addgene #37200[51]]. For more detailed information see Supplementary Materials and Methods. The sequences of all plasmid constructs were confirmed by Sanger sequencing (Microsynth AG).

2.3. Cell culture and transfection

All cell lines, Neuro-2a (CCL-131), HeLa (CCL-2), NIH/3T3 (CRL-1658), and HEK293 (CRL-1573)—all from ATCC, were maintained

in Dulbecco's modified Eagle's Medium supplemented with 10% fetal bovine serum, 2 mM GlutaMAX, 100 U/ml penicillin, and 100 μg/ml streptomycin at 37 °C in a 5% CO2 in air and humidified atmosphere. All cell lines were checked regularly and before the experiments for mycoplasma infection.

2.4. Experiments with pSc1, pSc2 plasmids, and HDR assays

Cells were seeded at 1.5×10^5 cells/well in the case of HeLa, Neuro-2a (N2a) and HEK293 cells, and at 3×10^5 cells/well density for NIH/3T3 cells, on 6-well plates 1 day prior to transfection. Cells were transfected by 2 μg total DNA/well, combining 1 μg Cas9-gRNA vector and 1 μg GFP vector DNAs (pSc1, pSc2, pHR-PRNP, pHR-ROSA), and using either Turbofect at 3 μl/μg total DNA (in the case of HeLa, N2a, and HEK293 cells) or Lipofectamine 2000 at 4.5 μl/μg total DNA (in the case of NIH/3T3 cells). Transfected cells were passaged at a 1:10 ratio until confluency reached 80–100% and the excess cells were analysed by FACS. The passaging continued until the transient expression disappeared (typically 12–15 days post transfection). Transfection efficiency was determined on the second day after transfection by measuring the fraction of GFP positive cells (the values obtained were considered 100%) and was later used to normalize the corresponding sample data. All transfections were made in triplicates.

2.5. Coupled integration of the GFP and puromycin cassettes

Before performing the experiments, the lowest antibiotic concentration that kills all non-transfected N2a cells was determined by applying increasing concentrations of antibiotic, up to 3.1 μg/ml puromycin. A concentration of 0.7 μg/ml was the lowest that killed all cells after 2 weeks of selection. This concentration was used further for selection. N2a cells were co-transfected in four replicates with pSc1-puro and pten specific Cas9-gRNA plasmid, using the previously described protocol with the pSc1 plasmid. As a control for random integration, pSc1-puro was co-transfected with a plasmid harbouring an inactive SpCas9 cassette (dCas9). Non-transfected control cells were also used to monitor selection effectiveness. Puromycin selection was started 2 days after transfection and on days 2 and 12 post-transfection the number of GFP positive cells were measured by FACS.

2.6. Experiments with pSc1-DD

Transfection was performed similarly as above with pSc1 plasmid, except now 1 μg Cas9-gRNA, 300 ng pSc1-DD, and 300 ng mCherry expressing plasmids were used per well for transfection on 6-well plates. Transfection efficiencies were determined 2 days after transfection by measuring the fraction of mCherry positive cells (the values obtained were considered 100%) and was later used to normalize the corresponding sample data. TMP (1 μM) was added to the media 24 h before FACS analysis.

2.7. Experiments with eSpCas9

A gRNA and mCherry coding plasmid (pmCherry_gRNA) was co-transfected with a wtSpCas9 or eSpCas9 coding plasmid (pX330-Flag-ehSpCas9 or pX330-Flag-wtSpCas9, respectively) and with the pSc1 self-cleaving plasmid into N2a cells. Transfection was performed similarly as with the pSc1 plasmid above, except now 760 ng Cas9 coding, 760 ng pSc1 and 480 ng pmCherry_gRNA plasmid was used. Transfection efficiencies were determined 2 days after transfection by measuring the fraction of GFP positive cells (the values obtained were considered 100%) and was later used to normalize the corresponding sample data.

2.8. Isolation of genomic DNA and genomic PCR

Genomic DNA from 1×10^6 cells was isolated after stable GFP-cassette integration (typically 12–15 days after transfection) using PureLink™ Genomic DNA Mini Kit (Invitrogen) following the manufacturer's protocol or by following Gentra DNA purification protocol (Gentra Puregen Handbook, Qiagen). From the isolated genomic DNA PCR was conducted with Q5 polymerase (for PCR primers and conditions, see Supplementary Data).

2.9. TIDE and T7 endonuclease I assay

Genomic PCR products were gel excised, purified via NucleoSpin® Gel and PCR Clean-up kit and were Sanger sequenced by Microsynth AG. Indel efficiencies were analysed by TIDE web tool[33] (https://tide.nki.nl/) by comparing Cas9 treated and control samples.

Samples of 200 ng from the same PCR reactions that were analysed by TIDE were digested with five units of T7 endonuclease I for 2 h at 37 °C and were then analysed by agarose gel electrophoresis. Densitometry assessment of cleaved and uncleaved bands were conducted with ImageJ software and the following formula was used to determine InDel efficiency: $\text{InDel} = 100*(1 - (1 - (b+c)/(a+b+c))^{0.5})$, where *b* and *c* represent the cleaved products and *a* represents the uncleaved band.

2.10. Statistics

Differences between groups were tested for statistical significance in each experiment using one-way ANOVA test with Bonferroni's post-hoc test (IBM SPSS Statistics v.21).

2.11. Flow cytometry

Flow cytometry analysis was carried out on an Attune Acoustic Focusing Cytometer (Applied Biosystems by Life Technologies). Attune Cytometric Software was used for data analysis. In all experiments, a total of 10,000 events were acquired and viable single cells were gated based on side and forward light-scatter parameters. The GFP signal was detected using a 488 nm diode laser for excitation and a 530/30 nm filter for emission and the mCherry fluorescent signal was detected using a 488 nm diode laser for excitation and a 640LP filter for emission.

3. Results

3.1. In-cell cleavage of a GFP-plasmid facilitates monitoring of gRNA efficiencies

We aimed to compare the DNA double strand break mediating efficiencies of different gRNAs with SpCas9. To this end, we measured the fluorescence intensities associated with GFP expression cassettes captured by the NHEJ repair system at the corresponding sites of the double strand breaks. Here, our main concern was whether NHEJ-mediated integration would sufficiently be efficient to be discernible by fluorescence-detection from the random integration background of the GFP cassette. To test this, first, we co-transfected N2a mouse neuroblastoma cells with a circular GFP expression vector (pEGFP-C1 - referred to as GFP-plasmid) along with the SpCas9 and gRNA expressing vectors, and monitored the integration events by the

fluorescence of GFP after transfection. As a control, a nuclease inactive SpCas9 (dCas9) expressing plasmid was employed that allowed to monitor the decay of the fluorescent signal originating from transient expression and facilitated determination of targeted integration as being the signal detected above the background, non-targeted, random integration of the expression cassette.

Five spacer sequences targeting three genes in the mouse genome, three for *Prnp*, one for *Sprn* and one for *Piwil4* genes, were selected and the corresponding expression constructs with SpCas9 and gRNAs were co-transfected with the GFP-plasmid into N2a cells. Cells displaying GFP fluorescence were monitored by FACS until the signal derived from transient expression has decayed and the fluorescence stabilized. This occurred over a 12–15-day period in all experiments (Supplementary Fig. S1), depending on factors such as transfection efficiency, which could be determined by measuring the GFP fluorescence on the second day after transfection. Given that the level of stabilized fluorescence varies from experiment to experiment, all samples to be compared were tested side by side in these studies. In these experiments, however, after the transient expression decayed, none of the samples could be clearly distinguished from the dCas9 controls representing random integration (Supplementary Fig. S2A).

Since we expected linearized plasmids to be more readily integrated, next, we linearized the GFP-plasmid with a restriction enzyme and tested its integration similarly in N2a cells, and also, with an overlapping set of gRNAs in NIH/3T3 mouse fibroblast cells. Since the linearized DNA is more vulnerable to cellular nucleases, in order to decrease the possibility of digestion of the expression cassette before the targeted integration would take place, more than 1000 bp-long extra sequences were left at both sides of the GFP expression cassette when linearized by the restriction enzyme. The use of a linearized plasmid resulted in higher integrations in both cell lines (Supplementary Fig. S2B and C), however, similarly to the previous results using circular plasmid (Supplementary Fig. S2A), none of the samples provided signals over twice as large as the level of random integrations. This is in spite of that the majority of the employed gRNAs supported HR mediated targeted integration in N2a cells (unpublished results). These data, in line with the general perception, suggest that the efficiencies of capturing circular or pre-linearized exogenous DNA by NHEJ-repair at the sites of targeted cleavage are not generally sufficient to be clearly detectable above the level of random integration.

The efficiency of targeted integration of exogenous DNA by NHEJ was reported to be somewhat higher when the donor plasmid was linearized inside the cell by designer nucleases.[38,44,52] To exploit this possibility, we picked two efficient-gRNA expressing plasmids that were available in our laboratory (targeting *Ttn* and *Rbl2* genes—referred to as Ttn and Rb, respectively), and cloned each corresponding protospacer into the GFP-plasmid either preceding, or both preceding and following the GFP-cassette (Fig. 1A). These plasmids and the control (protospacerless) GFP-plasmid were each co-transfected with a vector containing an active SpCas9 and the corresponding gRNA (Rb or Ttn) or a vector containing inactive SpCas9. The numbers of fluorescent cells were counted as previously. The in-cell cleavage of plasmids resulted targeted integrations four to six-fold above the background level of random integration measured with circular, protospacerless GFP-plasmid and active SpCas9 (Fig. 1B).

The difference was less than two-fold between the two controls, when cleaving the genome without in-cell linearization and the random integration of a circular GFP-plasmid (Fig. 1B), similarly as we

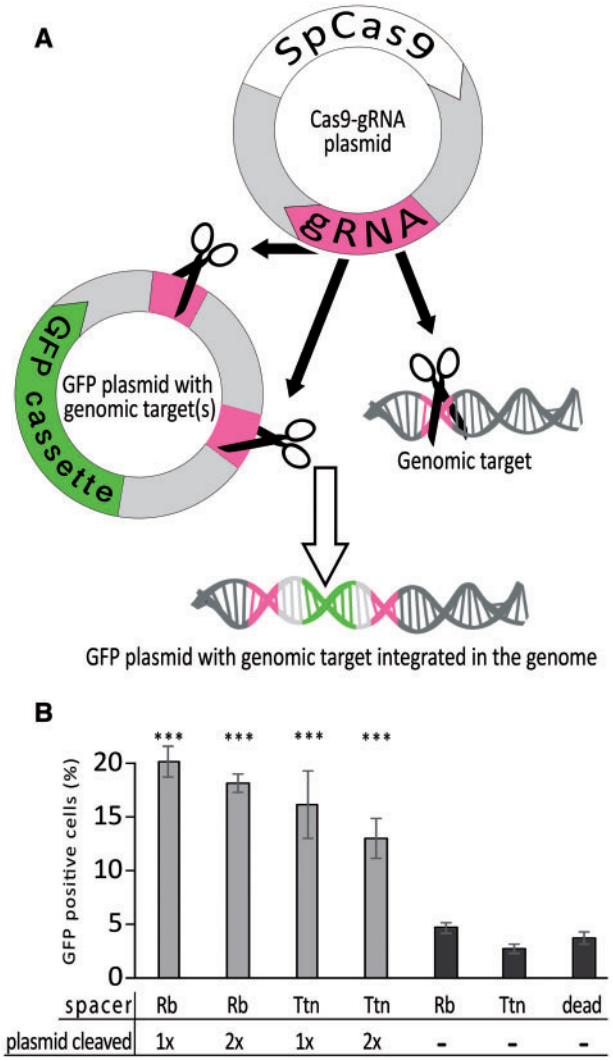

Figure 1. in-cell cleavage of the GFP-plasmid facilitates targeted integration. (A) Scheme of the genomic integration of an in-cell cleavable plasmid. The target sequence of the gRNA (magenta) from the plasmid containing the SpCas9 is present both in the genome (magenta target in the grey double helix) and in the GFP-cassette containing plasmid - either preceding or at both ends (the latter case is shown in the figure by two magenta boxes) of the GFP-cassette (green). SpCas9 cleaves both the genome and the GFP-plasmid. During repair of the broken genomic DNA, the opened plasmid will efficiently integrate into the genome at the cleavage site even in the absence of homologous arms. The protospacer on the GFP-plasmid has to be redesigned and re-cloned for every new genomic target in order to achieve opening of the plasmid. (B) in-cell linearization of the circular plasmid increases the efficiency of the genomic integration. Two different genes, *Rbl2* and *Ttn* were chosen and targeted by corresponding gRNAs (Rb and Ttn, respectively), to test the effect of in-cell linearization of the GFP-plasmid. Bars show the percentages of cells harbouring stably integrated GFP-cassette after co-transfection by an expression vector for both SpCas9 and either Rb or Ttn gRNAs and a circular GFP-plasmid that either contains one (Rb 1x, Ttn 1x) or - as on scheme A - two (Rb 2x, Ttn 2x) copies of the respective targets (light grey bars). As controls (dark grey bars), an inactive SpCas9 with the GFP plasmid (dead) and to measure the backgrounds without in-cell cleavage, an active SpCas9 with Rb or Ttn targets combined with the circular plasmids were used (Rb -, Ttn -). The values were compared to their respective circular control with the active nuclease. Bars show the means ± S.D. of percentages measured in $n = 3$ independent transfections, which were normalized to the transfection efficiency (% GFP positive cells measured on 2nd day after transfection: ranging between 87 and 91%). ***$P<0.001$.

found earlier using different spacers (Supplementary Fig. S2A). Cleaving at both sides of the GFP-cassette may lead to more in-cell linearized plasmids, and so it may be expected to result in higher integrations. However, we detected no increased fluorescence with the plasmids that contain two targets thus, for the further experiments we cleaved only at one site of the plasmids.

3.2. Construction of self-cleaving plasmids with universal target-spacer pairs

Although the above results are encouraging and indicate that the integration of in-cell linearized plasmids without homology arms are effective in these cells, such a design does not allow the measurement of the actual integration to the SpCas9 cleaved genomic sites. This is because the same gRNA mediates the cleavage of the genome and of the plasmid, and thus, the two events are coupled and it is not possible to assess the background originating only from the random integration of the in-cell linearized plasmid independently of the genome cleavage.

Furthermore, the efficiency of targeted integration when using the same target sequence both in the genome and in the GFP-plasmid (Fig. 1A) depends on both, possibly differing, effectiveness of cleaving the target in its genomic context and out of its genomic context on the GFP-plasmid. To allow a more comparable and adequate estimation of cleavage activities for various gRNAs on their targets in their genomic context, the use of one common GFP-plasmid that contains the same target sequence ensuring identical in-cell cleavage for all gRNAs to be tested, and which is also independent of the tested targets, would be the most adequate. In addition, such an approach would eliminate the need for the laborious cloning of each corresponding protospacer to the GFP-plasmid for each spacer to be tested. The practicality of such generally applicable plasmid would be further extended if the corresponding common gRNA for the in-cell cleavage of the plasmid was also placed into the GFP-plasmid. Further we refer to plasmids constructed in this way as 'self-cleaving', since such plasmids, expressing the gRNA with its own target, are programmed to initiate their own cleavage (Fig. 2A).

To prepare such effective GFP-plasmids for general use, we computer-generated a pool of 30,000 random spacer sequences and picked the ones that have the least predicted off-targets in the mouse genome as assessed by e-CRISP Cas9 construct designer[53] and had the best predicted efficiency by sgRNA Designer.[16,21] From these, two candidates were picked that according to the CasFinder[54] have no targets in the human genome and in 11 frequently employed model organisms (mouse, rat, cat, dog, pig, cow, chicken, zebrafish, *C. elegans*, *Drosophila*, and *S. cerevisiae*) thus, reserving the potential to be used more generally for self-cleaving GFP-plasmids in different cell lines/organisms (although, we do not explore these possibilities here). We refer to these as TL (targetless) spacers, TL1 and TL2; these two spacer sequences were used to generate two 'self-cleaving' GFP-plasmids (referred to as pSc1 and pSc2, respectively - Addgene #80436, #80437) containing both the TL spacer/gRNA and the corresponding protospacer on one vector (Fig. 2A).

We tested the targeted integration of the two self-cleaving plasmids in combination with either the Ttn or Rb-gRNA in comparison with the Ttn and Rb cleaved in-cell linearized plasmid (Fig. 2B, indicated as Rb-, Ttn-InCell), as earlier. As a control, TL-gRNA bearing plasmids were co-transfected with either of the two self-cleaving plasmids; thus, the GFP-plasmid is linearized in-cell without the genome being targeted. These experiments demonstrate that the self-cleaving plasmids achieve comparable integrations with the two in-cell

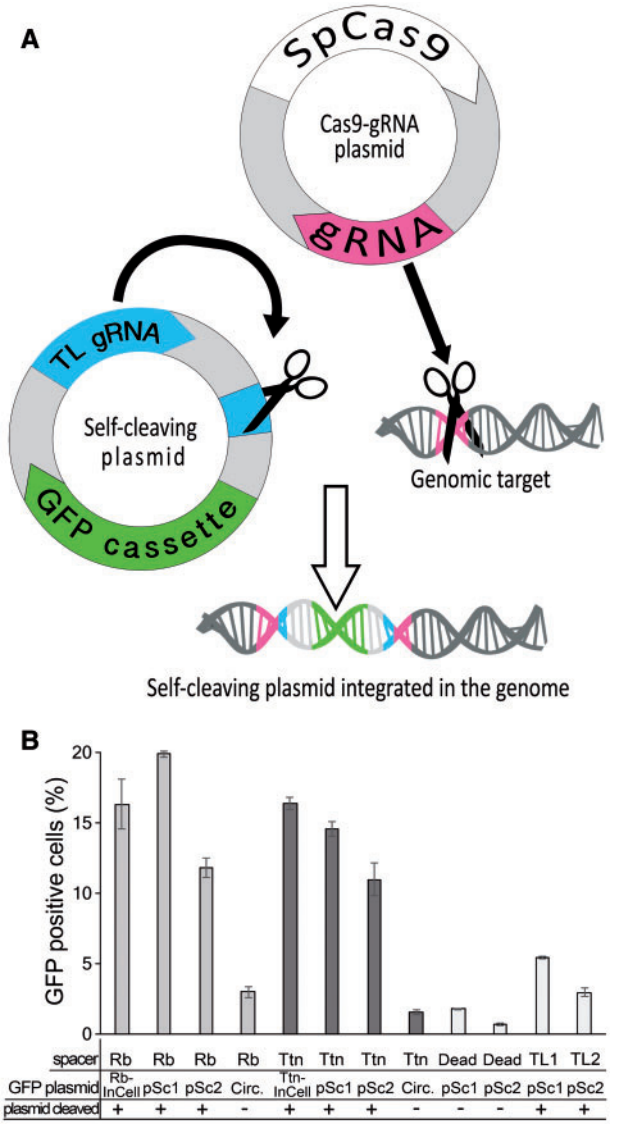

Figure 2. (A) Scheme of the genomic integration of a self-cleaving plasmid. The target sequence of the gRNA (magenta) expressed from the SpCas9 plasmid is present only in the genome (magenta target in the grey double helix). The circular plasmid that contains a GFP-box (green) harbours a genomic-targetless gRNA (TL gRNA) and its target sequence (blue boxes). During repair of the broken genomic DNA, the opened plasmid will efficiently integrate into the genome at the cleavage site even in the absence of homologous arms. Such construction of the GFP-plasmid does not require retailoring the protospacer for every new genomic target and can be universally used. (B) The efficiency of targeted integration when using self-cleaving GFP-plasmid in N2a cells. The percentages of the cells harbouring a stably integrated GFP-cassette are shown. The *Rbl2* or *Ttn* genes were targeted by the corresponding gRNAs, Rb or Ttn, respectively. Cells were co-transfected by the expression vector for both wild type SpCas9 and a respective gRNA, and either of the self-cleaving circular GFP-plasmids pSc1 or pSc2 or the circular GFP-plasmid containing the Rb or Ttn protospacer (Rb-InCell, Ttn-InCell) or without containing any target sequence (Rb or Ttn, Circ). To measure the random integration background of the self-cleaving plasmids, they were co-transfected with a vector expressing active SpCas9 and the corresponding TL gRNAs; thus, the genome is not targeted, but the self-cleaving plasmids are linearized (TL1, TL2). As an additional negative control, an inactive SpCas9 was used along with either of the two self-cleaving plasmids (dead). Bars show the means ± S.D. of percentages measured in $n = 3$ independent transfections, which were normalized to the transfection efficiency (% GFP positive cells measured on 2nd day after transfection: ranging between 94 and 96%).

cleaved GFP-plasmids in case of both genome-targeting (Ttn or Rb) gRNAs tested. This suggests that both gRNAs containing TL spacers (TL-gRNAs) support as efficient plasmid DNA cleavage as the Rb or Ttn spacers do. Thus, these self-cleaving plasmids seem to facilitate efficient integration, while providing a more adequate assessment of efficiencies of different gRNAs.

The two self-cleaving GFP-plasmids pSc1 and pSc2 were further used to test 11 spacers, targeting 7 different mouse genes [*Pten*,[55] *Rbl2*, *Tp53*,[55] *Prnp*, *Sprn*, *Piwil2*, and *Piwil4* (Fig. 3A and B)]. Targeted integration is assessed here as the fluorescence *above* the background level that derives from the random integration of the in-cell cleaved self-cleaving plasmid. For this control sample the gRNA, which is identical to the corresponding TL-gRNA of the self-cleaving plasmid is placed to the SpCas9 plasmid; thus, the GFP-plasmid is linearized in-cell without the genome being targeted. The random integration of these in-cell-cleaved GFP cassettes is usually about 4–5%, which is higher than the random integration of a circular (1–2%) or even a linear (2–3.5%) plasmid. The nearly 2% increment likely come from an increased propensity of the in-cell cleaved plasmid for random integration, although a contribution also from off-target cleavage activity of the TL guides cannot be ruled out. Nevertheless, these values—although, may vary among cell types—are sufficiently low to allow the detection of targeted integration above this background level fluorescence.

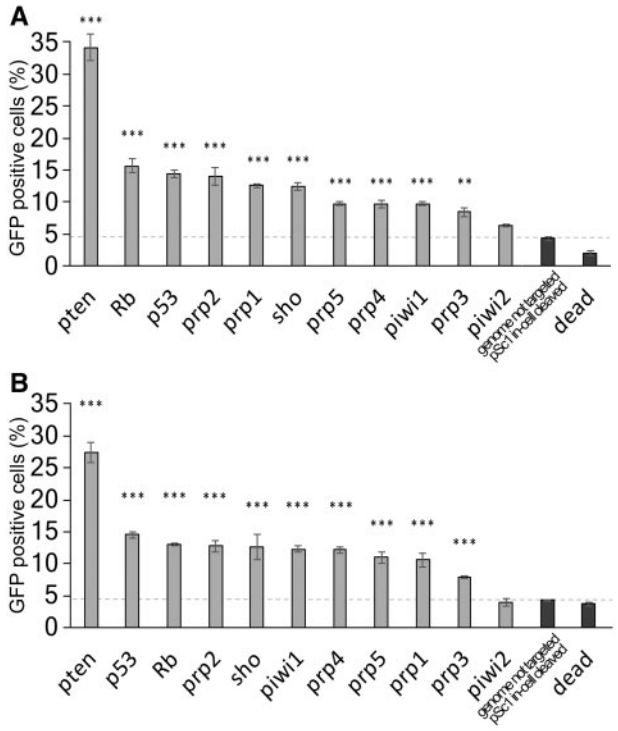

Figure 3. The efficiency of different spacers to mediate targeted integration when used with self-cleaving GFP-plasmids in N2a cells. The efficiencies of different spacers in targeted integration using the pSc1 (A) or pSc2 (B) self-cleaving GFP-plasmid are assessed. The percentages of the cells harbouring stably integrated GFP-cassette were measured for 11 different spacers (pten, Rb, p53, prp1-5, sho, piwi1-2) targeting seven genes (*Pten*, *Rbl2*, *Tp53*, *Prnp*, *Sprn*, *Piwil4*, *Piwil2*) as indicated by the light grey bars, when co-transfecting with either the circular pSc1 (A) or pSc2 (B) self-cleaving GFP-plasmids. To measure the random integration background of the self-cleaving plasmids, they were co-transfected with a vector expressing active SpCas9 and the corresponding TL gRNA; thus, the genome is not targeted, but the self-cleaving plasmids are linearized (genome not targeted, Sc1 in-cell cleaved; A and B, indicated by the dashed lines). Bars show the means ± S.D. of percentages measured in $n = 3$ independent transfections, which were normalized to the transfection efficiency (% GFP positive cells measured on 2nd day after transfection: ranging between 94 and 98%). Values obtained are compared to the control where the linearized self-cleaving GFP-plasmid (genome not targeted, pSc1 in-cell cleaved) was used. ***$P<0.001$.

Several spacers gave rise to targeted integration (above the random levels of 4%) in combination with either of the two self-cleaving plasmids, pSc1, or pSc2 (Fig. 3A or 3B, respectively). Experiments with the least effective spacer piwi2, (indicated as 'piwi2', Fig. 3) resulted in no clearly detectable targeted integrations that would exceed random integration of an in-cell cleaved plasmid. By contrast, the pten spacer (indicated as 'pten', Fig. 3) was by far the most effective with both self-cleaving plasmids mediating targeted integration of around 25%. The two self-cleaving GFP-plasmids supported targeted integrations in a similar fashion in combination with the 11 spacers. To test the presence of targeted integrations, we performed PCR using one primer complementary to the end of the plasmid and another one complementary to the genomic locus, on DNA samples purified from the stably transfected cell populations. Testing on the *Pten* locus agarose gel electrophoresis analysis revealed that the plasmids were incorporated at both orientations to the target site (Supplementary Fig. S3). Direct sequencing of the PCR products in case of integration at a different locus (*Tp53*), demonstrated that a considerable fraction of the integrations took place by error-free NHEJ repair (Supplementary Fig. S4) as it had also been reported earlier for blunt ended NHEJ-cloning.[38,46]

In another experiment, additional seven gRNAs (targeting the mouse *Vim* and *Dcun1d2* genes) were tested along with those targeting *Pten*, *Rbl2*, and *Ttn*, with both the wild type and the enhanced SpCas9 (eSpCas9) (Supplementary Fig. S5). While for Rb and Ttn both nucleases showed comparable efficiencies and for four other gRNAs (all that target *Dcun1d2* and with vim1 spacer) neither of the nucleases showed efficient cleavage compared to their controls, with another four wtSpCas9 produced higher rates of targeted integrations while eSpCas9 showed low or no activity, which most probably is attributable in part to the higher specificity of eSpCas9.

3.3. The universal self-cleaving GFP-plasmid approach compares to an HR-based fluorescence reporter method and it is also effective in NIH/3T3, HeLa, and HEK293 cells

We compared the relative efficiencies of four prion protein gene-targeting spacers to result either HR or NHEJ mediated targeted integration of a GFP-cassette flanked by either 1 kb-long homology arms or no homologous sequences, respectively. The latter is present in the pSc1 plasmid (Fig. 4). We selected genomic targets whose cleavage sites are situated within a few base pairs in order to minimize the effect of the distances between cleavage sites of SpCas9 and the beginning of the homologous region, on the efficiency of HR-mediated targeted integration. Two of the cleavage sites (prp2, prp3) were situated in a few nucleotide distances from the end of the upstream and two other (prp4, prp5) from the beginning of the downstream HR arms, respectively. Both HR and NHEJ mediated integrations indicated that both target pairs can be efficiently cleaved by SpCas9. Interestingly, in case of the upstream site gRNAs (prp2 and -3) prp2 seems to be more effective than prp3 when assessed by NHEJ-mediated integration that is not apparent in the HR-mediated assay.

In another experiment, we selected and cloned 5 spacers targeting the *Rosa26* locus in the mouse genome (one of them, rosa5 spacer

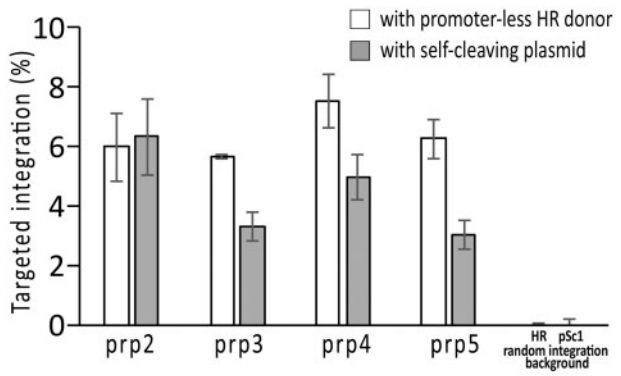

Figure 4. Efficiency of integration of the self-cleaving plasmid to the genome is comparable to that of a HR-mediated integration. The percentages of the cells harbouring a stably integrated GFP-cassette (without normalization to the transfection efficiency) are shown for four different spacers targeting the *Prnp* gene (prp2-5). Cells were co-transfected by an active SpCas9 and the respective gRNA expressing vector along with either of the following GFP-plasmids: a homologous recombination (HR) donor GFP-plasmid (empty bars) containing a promoter-less GFP gene with 1 kb homologous arms for the *Prnp* gene at both sides of the GFP-cassette; or the self-cleaving GFP-plasmid pSc1 (grey bars). As a control for the HR mediated integrations, the HR donor plasmid was co-transfected with an inactive SpCas9 expression vector (dead Cas9). The values obtained for this control (0.1%) indicated as *random integration background with HR*, was subtracted from all corresponding sample values to determine the extent of targeted integrations. For experiments with the self-cleaving plasmid, the control used was the self-cleaving plasmid pSc1 co-transfected with a vector expressing active SpCas9 and the TL1 gRNA and thus, the pSc1 plasmid is linearized in-cell but the genome is not targeted (indicated as *random integration background with pSc1 donor*). The value obtained for this control (5%) was subtracted from all corresponding sample values to determine the extent of net targeted integrations. Values were not normalized for transfection efficiency in either experiments. Bars show the means ± S.D. of percentages measured in $n = 3$ independent transfections.

previously reported in ref.[56]). The five gRNAs seem to mediate comparable cleavage efficiencies (about 10–20%) by both approaches (Supplementary Fig. S6), although the target sites were not so carefully selected as in case of Figure 4 and thus, the relative position of the targets may have influenced their efficiency in the HR-mediated assay.

HR-mediated integration is currently the only approach used in fluorescence-based methods that are capable of testing the genomic targets in their genomic context (GFP-disruption and analogous assays are restricted to targets aimed at the GFP sequences). However, the HR-mediated method is much more laborious due to the need to construct the homologous arms. Therefore, it is only worth to apply it for screening gRNA-efficiencies if the eventual application is to employ the same homologous arms. Thus, the advantages of our approach over other fluorescence-based methods for gRNA testing are apparent and it may be more relevant to compare to PCR-based methods what we present later in the manuscript.

Further experimentation on NIH/3T3 mouse fibroblast cells with one of the self-cleaving plasmids and seven spacers of the set previously tested in N2a cells, targeting five genes (*Pten*, *Rbl2*, Tp53, *Prnp*, and *Piwil4*) resulted in sufficiently high targeted integrations (i.e. for the majority of the spacers clearly measurable above the background 5%) (Supplementary Fig. S7A). The most effective spacer, pten resulted in close to 15% targeted integration, whereas piwi2 showed no detectable targeted integration. Thus, these results recapitulate those seen earlier with these spacers in N2a cells and suggest that our approach is applicable in NIH/3T3 mouse cells as well.

The TL spacers have neither targets nor predicted off-targets in the human genome, allowing us therefore, to test the applicability of our approach on human cells as well. The commonly used human cell line HeLa was selected, using one of the two self-cleaving plasmids (pSc1). We could not use most of the mouse gene-targeting spacers tested earlier in N2a cells on the corresponding human genomic sites; however, we recognized that the pten spacer also targets the human *PTEN* gene and *PTENP1* pseudogene. In addition, two other spacers that target the human *PRNP* gene were designed. While the in-cell linearized plasmid background remained at 1.4%, the two prion protein targeting spacers and the pten spacer mediate targeted integrations (i.e. above background level) from 5% to 15%, respectively (Supplementary Fig. S7B). Additionally, four previously reported active gRNAs[57] were tested in HEK293 cells targeting the *LMNA* and *PML* genes along with the gRNA targeting the *PTEN* gene. All spacers mediated effective targeted integrations (Supplementary Fig. S7C). Altogether, these results on different cell lines (N2a, NIH/3T3, HeLa, and HEK293) suggest that NHEJ-mediated targeted integration in combination with these self-cleaving plasmids is sufficiently high above random integration to exploit it for testing gRNA efficiency more generally.

3.4. Shortening detection time with a degradation domain fused to GFP

Although these experiments demonstrate the effectiveness of our approach requiring little effort and cost, the long time-frame necessary for obtaining the readout limits its use. Our experiments suggested that the long decay time of the transient signal originates from the high stability of GFP protein molecules that are present after the disappearance of the plasmid DNA. To avoid this effect, ideally, GFP expression should be induced after the degradation of the transfected plasmid DNA in the cell, ensuring that the GFP expression originates only from the integrated copies of the expression cassette. We achieved a similar outcome by fusing a degradation domain (DD), a destabilized mutant of *E. coli* dihydrofolate reductase (ecDHFR) protein to GFP that drives faster degradation of the fluorescent protein. The degradation domain is stabilized in the presence of trimethoprim (TMP), preventing enhanced degradation.[50] Adding TMP to the cells after the transfected plasmid DNA has disappeared allows the detection of the GFP signal originating only from proteins expressed from the integrated copies.

Fusing the degradation domain either to the N- or to the C- terminus of GFP resulted similar expression profiles (data not shown), thus, we selected the N-terminal fusion for further experiments and placed it to a self-cleaving plasmid (pSc1-DD, Addgene #80439). Three gRNAs tested previously, Rb, p53, and piwi2 (the first two representing efficient, whereas the third poor gRNAs) were exploited to adjust the condition of the modified assay employing pSc1-DD. The gRNA-Cas9 plasmids were co-transfected with both the pSc1-DD plasmid and an mCherry coding plasmid to N2a cells. The latter allows for monitoring the transfection efficiencies. To determine the earliest time point that is appropriate for distinguishing between poor and active gRNA-s, GFP positive cells were measured daily from the 2nd to the 10th day post transfection (data not shown). TMP was added to all samples 24 h before FACS analysis. The gRNAs tested previously with the pSc1 and pSc2 plasmids (Fig. 3), now with pSc1-DD showed similar activities. This approach resulted in the successful reduction of the assay time from 12–15 to 6–8 days

(Fig. 5). The activities of the gRNAs are clearly distinct (i.e. the poor and good gRNAs can be easily distinguished) on the 6th day post transfection, although the fluorescence stabilizes only from the 8th day (data not shown).

3.5. The approach is also validated by two PCR-based methods: TIDE and T7 endonuclease I assay

Ten spacer sequences targeting six genes in the mouse genome (four for *Prnp*, two for *Pten*, and one for *Dcun1d2, Rbl2*, *Tp53*, and *Piwil4* genes), were selected for a comparison of this approach and indel detection methods such as Tracking Indels by Decomposition (TIDE[33]) and T7 endonuclease I assay. These 10 gRNA-Cas9 plasmids were used for co-transfection with the pSc1-DD plasmid into N2a cells. Genomic DNA from the FACS-analysed cells was extracted on the 8th day post transfection and the SpCas9 targeted sites were PCR amplified and indels were detected by both TIDE and T7 endonuclease I assay. All approaches, this reporter assay, TIDE and T7 endonuclease I assay identified the same five gRNAs as being the more efficient ones (pten, pten2, p53, Rb, and prp2) and the same five gRNAs to result low or no cleavage activities (prp8, prp9, prp10, piwi2, and dcun1d2) (Fig. 6). However, as expected, the relative efficiencies among the gRNAs slightly differed, likely reflecting the differences in the outcomes of the competing repair events they monitor.

3.6. In-cell cleavage of a linear plasmid does not recapitulate the same level of integration as in-cell cleavage of a circular plasmid

It is not clear why in-cell cleavage of circular plasmids promotes more efficient integration as compared to transfection of linear, i.e.: *out-of-cell cleaved* ones (we refer to this as pre-linearized plasmids). One explanation could be that linear DNA is more prone to degradation and a large proportion of the transfected DNA is digested before its targeted integration can take place. Alternatively, the end of the linear molecule might become integration incompetent with time, for example by concatemerization. It is also possible that NHEJ ligation may preferentially take place among DNA ends that are liberated at the same time. To get more insight into this problem we compared NHEJ-mediated targeted integration of pre-linearized plasmids either cleaving or not cleaving them again in-cell. in-cell cleavage of the pre-linearized plasmids apparently does not increase targeted integration whereas in-cell linearization of the circular plasmid (resulting in one fragment) facilitated twice as many targeted integrations as pre-linearization (Fig. 7).

Since in-cell double cleavage (where the restriction enzyme linearized plasmid is cleaved both upstream and downstream the GFP-cassette), may effectively triple the number of DNA molecules with free ends, it might adversely affect the integration efficiency of the fragment containing the GFP-cassette having to compete for the available genomic DNA breaks. To account for this effect, we repeated the experiments, now using controls of pre-linearized plasmids digested into one, two or three pieces. The linear plasmids were also cleaved again once or twice in-cell generating either two or three fragments, respectively (Supplementary Fig. S9). Pre-linearization of the plasmids into two or three pieces before transfection decreased targeted integration, as expected. in-cell cleavage of the pre-linearized plasmid into two or three pieces facilitated targeted

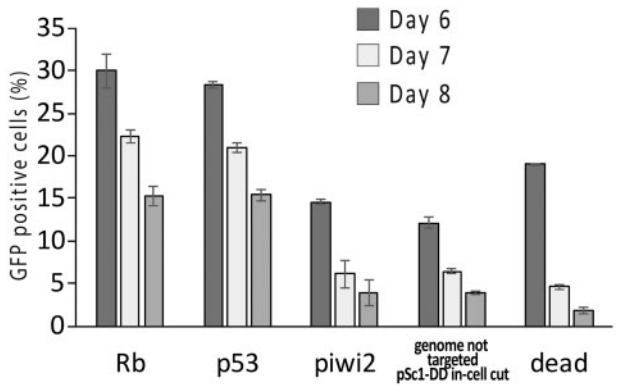

Figure 5. Shortening detection time with a degradation domain fused to GFP. The percentages of the cells harbouring stably integrated GFP-cassette were measured between 6 and 8 days after transfection when a self-cleaving plasmid with a degradation domain-fused GFP (pSc1-DD) was co-transfected with three formerly used (e.g. Fig. 3) SpCas9-gRNA coding plasmids (targeting the *Rbl2*, *Tp53* and *Piwil4* genes) and an mCherry coding plasmid (to track transfection efficiency) to N2a cells. To stabilize the degradation domain, trimethoprim was added to the cells 24 h before FACS analysis. As controls, the pSc1-DD plasmid was co-transfected with either (I) a nuclease inactive SpCas9 coding plasmid (dead); or (II) with an active SpCas9 and the TL1 gRNA expressing vector and thus, the pSc1 plasmid is linearized in-cell but the genome is not targeted (indicated as *genome not targeted pSc1-DD in cell cut*). Bars show the means ± S.D. of percentages measured in $n = 3$ independent transfections, which were normalized to the transfection efficiency [% mCherry positive cells (ranging between 44 and 52%) measured on the 2nd day after transfection].

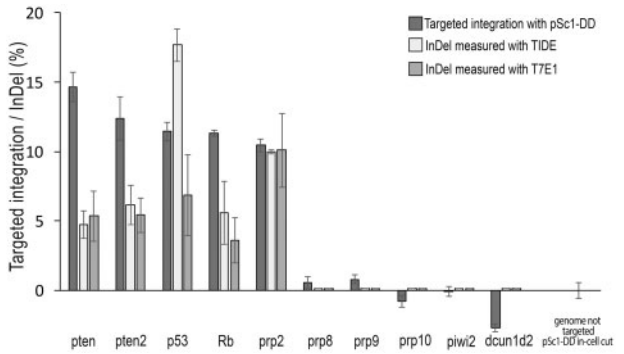

Figure 6. Validation of the method by TIDE and T7 endonuclease I assay. The activities of 10 gRNA-s are shown that were measured with the degradation domain-fused GFP bearing self-cleaving plasmid (pSc1-DD, dark grey boxes) and with indels that were detected in the same samples with TIDE (light grey boxes) and T7 endonuclease I assay (medium grey boxes). The percentages of the cells harbouring a stably integrated GFP-cassette are shown for 10 different spacers targeting the (*Prnp, Pten, Dcun1d2, Rbl2*, *Tp53*, and *Piwil4* genes). Cells were co-transfected by an active SpCas9 and the respective gRNA expressing vector along with the degradation domain bearing GFP-self-cleaving plasmid, pSc1-DD (dark grey bars). As a control the pSc1-DD was co-transfected with an active SpCas9 and the TL1 gRNA expressing vector and thus, the pSc1-DD plasmid is linearized in-cell but the genome is not targeted (indicated as *genome not targeted, pSc1-DD in-cell cut*). The GFP positive cells were analysed by FACS 8 days after transfection. Bars show the means ± S.D. of percentages measured in $n = 3$ independent transfections which were normalized to the transfection efficiency [% mCherry positive cells (ranging between 44 and 52%) measured on the 2nd day after transfection]. After FACS measurements, the three parallel samples were mixed, genomic DNA was isolated and three parallel PCR reactions were carried out from each sample. Indels were measured by both TIDE and T7 endonuclease I assay on each PCR reaction. The agarose gel electrophoresis analysis of T7EI assays are shown in Supplementary Fig. S8.

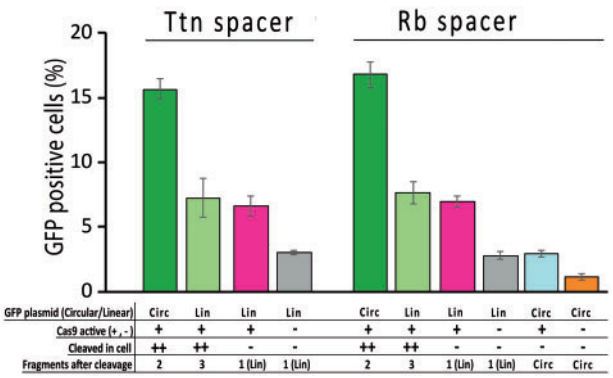

Figure 7. Integration efficiency of the in-cell cleaved pre-linearized plasmid is lower than that of the in-cell cleaved circular plasmid in N2a cells. The percentages of the cells harbouring stably integrated GFP-cassette are shown when targeting two different genes, *Ttn* or *Rbl2* with the spacers Ttn or Rb, respectively, as indicated. In case of either of the targets, the cells were co-transfected by an expression vector for nuclease active SpCas9 (+) and the respective gRNA (Rb or Ttn) along with: (I) a circular GFP-plasmid containing two matching protospacers (++) before and after the GFP-cassette (dark green bars; as seen formerly in Fig. 1A); (II) a pre-linearized GFP-plasmid containing two matching protospacers (++) before and after the GFP-cassette (light green bars); (III) a pre-linearized GFP-plasmid containing no protospacers (magenta bars). For both spacers as a negative control a nuclease inactive SpCas9 (-) and gRNA expressing plasmid was co-transfected with a pre-linearized GFP-plasmid (grey bars). For the in-cell linearized plasmid, as a control, a circular plasmid with no matching protospacer was co-transfected with Rb spacer and SpCas9 expressing vector, therefore, the plasmid cannot be linearized in the cell (sky blue bar). As an additional control a circular GFP-plasmid was co-transfected with inactive SpCas9 nuclease expressing vector (orange). Bars show the mean ± standard deviation (S.D.) of percentages measured in $n = 3$ independent transfections which were normalized to the transfection efficiency (% GFP positive cells measured on 2nd day after transfection: ranging between 95 and 99%).

integration twice as much as pre-linearization of the plasmids into two or three pieces before transfection. However, no significant differences were seen when the pre-linearized plasmid was cut once or twice in-cell (Supplementary Fig. S9). Thus, in-cell cleavage of pre-linearized plasmid failed to increase targeted integration to a similar level as the in-cell cleavage of a circular DNA. These experiments suggest that the major reason for the increased targeted integration of in-cell cleaved circular plasmids in N2a cells is the higher resistance to nucleases, although these experiments do not rule out a positive contribution from a synchronized in-cell cleavage.

3.7. NHEJ-cloning employing SpCas9 and a self-cleaving plasmid is superior to random integration

Targeted integration of plasmids as large as 5 kb by NHEJ-cloning using SpCas9, had also been demonstrated in mammalian cells, although, with limited efficiency.[45] We used a self-cleaving plasmid to assess if NHEJ-cloning is capable of targeted integration of 10 kb plasmids with sufficient efficiency. A 5 kb cassette was cloned to the self-cleaving plasmid (resulting in 10805 bp final size) and its integration was tested with the most effective spacer (pten) used in these studies. We found 30% and 18% targeted integration (without normalization for the transfection efficiency i.e.: without upscaling to 100%) of the 10 kb plasmid harbouring a GFP-cassette in N2a and HeLa cells, respectively (Supplementary Fig. S10A and B). These experiments suggest that the size of the plasmid in this range is not a limiting factor for NHEJ-cloning in these cell lines.

The efficiencies we achieved here for targeted integrations are almost an order of magnitude higher than those reported earlier in HEK293 cells without using a self-cleaving plasmid.[45] To see if the differences are attributable only to the different cell lines (N2a, NIH/3T3, HeLa vs. HEK293, and CHO) used, we tested our approach on HEK293 cells as well. The experiments presented on Supplementary Fig. S10C demonstrate that by using self-cleaving plasmids (either 5 or 10 kb large), efficient integrations are achievable in HEK293 cells, showing more than an order of magnitude higher efficiencies than seen in earlier studies [7% targeted integration here (Supplementary Fig. S10C) vs. 0.17% in Ref. 45]. These results were confirmed later exploiting more targets (Supplementary Fig. S7C).

Our experiments suggest that NHEJ-cloning using SpCas9 nuclease with a self-cleaving plasmid could be a superior alternative to random integration of DNA cassettes. To establish an easy way for integration of DNA cassettes to safe-harbour locations, we identified 5 spacers targeting the *Rosa26* locus in the mouse genome, as discussed earlier (Supplementary Fig. S6).

We also examined whether NHEJ-cloning of a self-cleaving plasmid would result in coupled integration of two expression cassettes harboured on the same plasmid. This is a major concern exploiting random integration where selection for an antibiotic does not always result in expression of the target gene located on the same plasmid. N2a cells were co-transfected with two plasmids: one is a self-cleaving plasmid containing both a GFP and a puromycin expression cassette (pSc1-puro, Addgene #80438) and the other is a plasmid for SpCas9 and gRNA expression targeting the *PTEN* gene. The fraction of cells exhibiting GFP fluorescence was counted by FACS after 10 days of puromycin selection when the transient GFP expression had been decayed. These experiments demonstrated about 90% coupled integration of the puromycin and the GFP expression cassettes by NHEJ-cloning of a self-cleaving plasmid. By contrast, this value is 50% for coupled random integration in the same experimental setup but without cleaving the plasmid and the genome (Supplementary Fig. S11). This latter experiment resulted in two orders of magnitude less resistant cells indicating the higher efficiency of NHEJ-cloning.

4. Discussion

In the absence of accurate prediction tools for the efficiency of Cas9 spacers, suitable gRNAs can only be selected reliably by experimental testing. Although several approaches exist to measure the activity of SpCas9-gRNAs,[24] there is an absence of simple fluorescence-based methods that would be able to assess the effectivity of gRNAs on targets in their genomic context while ensuring an easy and reliable detection by FACS. Here, we introduced a method that fulfils this gap and offers a simple alternative to existing approaches for testing gRNA efficiencies prior to executing more complex applications like generating specifically modified cell lines, transgenic animals or performing multiplex genome engineering tasks. As it requires only the ready gRNAs, our method can be applied by simply co-transfecting the Cas9-gRNA plasmids with a self-cleaving reporter plasmid. Thus, it does not require extra cloning steps for testing a particular gRNA in contrast to other fluorescent reporter-based methods.[35,39,40,42] In Table 1, we compared the self-cleaving plasmid reporter method to other FACS- and PCR-based methods. Although, the exact numbers for some of the parameters may vary from lab-to-lab making an absolute comparison difficult, considering optimal instances for each, the self-cleaving method presents numerous advantages over the other methods. FACS-based methods are inherently

Table 1. Comparison of methods that can detect Cas9 nuclease induced DNA mutations

	PCR based methods			FACS-based assays		
	TIDE[33]	T7E1	NGS	GFP disruption	EGxxFP[40]	self-cleaving plasmid
Readout time	2–5 days	2–3 days	7–21 days	3–7 days	2 days	6–8 days
Hands-on time/sample	3 h	3 h	3 h	40 min	30 min	40 min
Optimization needed	Yes	Yes	Yes	No	No	No
Genomic context	Yes	Yes	Yes	Yes	No	Yes
Any target	Yes	Yes	Yes	GFP targets only	Yes	Yes
Extra cloning steps	No	No	No	No	Yes	No
Cost/sample	$25	$18	$70	$7	$6	$7

The self-cleaving *plasmid* method is compared to some of the PCR- and FACS-based assays' most relevant attributes, which can detect Cas9 nuclease induced mutations. The cost and time calculations are available in detail in Supplementary Data. Favourable conditions are marked with white and less favourable conditions are marked with grey backgrounds.

less labor-consuming than PCR based methods. As far as optimization is concerned, here only a new cell line may present a need for optimization, which for our technique refers to adjusting the parameters of a FACS device. However, if the transfection efficiency is also measured when a PCR-based approach is applied (generally and most easily by exploiting a fluorescent protein and FACS), no extra optimization is necessary, in this sense, for our method. By contrast, PCR-based approaches require some optimization of the primers and PCR condition for each target region selected. Thus, it seems to be a favourable alternative approach for those who do not mind the longer time-frame in exchange for spending less effort on gRNA testing.

Another built-in advantage of the method is that the transfection efficiencies can easily be determined by simply measuring the GFP fluorescence of a portion of each population on the second day post transfection. While the transfection efficiency influences the readout of any gRNA-testing method to a great extent, it is generally not monitored in most of the studies using either PCR- or fluorescence-based methods.[58,59]

The targeted integration of the GFP-cassette can be assessed only after the transient expression has decayed. The use of an inducible degradation domain fused to GFP decreased this time to 6–8 days, a time-period that is less than what is usually needed for deep sequencing, but it is slightly longer than that required for measuring indel frequency by Surveyor/T7E1 assay (Table 1).[24] The extent of the random integration also seems to vary with the condition of the cells; this makes necessary the use of appropriate negative controls such as non-targeted integration of the in-cell cleaved plasmid, in each experiment.

The method measures the targeted integration to sites of genomic cleavage identified by the gRNA. Thus, it does not distinguish between on-target and off-target cleavage. However, off-target cleavage can generally be kept low with the careful design of the spacer sequence and using improved SpCas9 nucleases with reduced off-target effect, such as eSpCas9 (Supplementary Fig. S5) or SpCas9-HF.[60,61]

The extent and nature of the correlation between generating double strand breaks and integrating exogenous DNA by NHEJ repair is difficult to assess since there is no easily available method to directly measure the formation of double strand breaks in the cells. Most approaches are based on detecting those outcomes of the repairs of double strand DNA breaks that result in altered sequences.[24] However, the exact correlation between the number of double strand breaks and detectable repair events (i.e. repairs resulting in altered sequences) are not known for any of the approaches designed to monitor it. As such, they depend on many factors, for instance, the actual DNA sequence at the break and the position and presence of microhomologies that are thought to greatly influence the ratio of error-free to non-error-free NHEJ repair.[62–64] The conditions and factors that influence the ratio of HR to NHEJ mediated repair are also difficult to account for.[25] In addition, both main repair systems are capable of capturing foreign DNA at the break points, although the factors that may influence this are also poorly understood, for instance the ratio of NHEJ-mediated integration and indel generating NHEJ repair. In addition, the actual methods used might not accurately measure the frequency of mutations generated by the repair system. For example, it is not fully understood how the distribution and nature of the mutations in the indel population affect the outcome of the Surveyor/T7E1 nucleases assays.[65,66] Furthermore, PCR-based methods generally ignore larger deletions or insertions of which frequencies might also influence the number of smaller indels detected.[24] Thus, it is not apparent that the relative efficiencies of spacers would be identical when assessed with two different methods. Nevertheless, all approaches are likely to be suitable for a rough estimation to distinguish very poor from efficient spacers.

Here, we measured NHEJ mediated and HR mediated targeted integration using the same sets of gRNAs (Fig. 4, Supplementary Fig. S6). The gRNAs were designed to minimize the positional effects of the target position in relation to the homologous arms. Although there might be several other factors difficult to account for, this allowed a more direct comparison of the gRNA efficiencies to induce DNA repair by the two repair systems. The method was also compared to both TIDE and T7 endonuclease I assays that measure indel frequencies induced by SpCas9 cleavage employing an identical set of gRNAs. These experiments confirmed that our approach is capable of distinguishing poor and efficient spacer sequences and is a useful, effortless alternative for screening for more efficient gRNAs.

The method absolutely requires in-cell cleavage of the GFP-plasmid to be integrated. Occasionally, we were able to detect targeted integration of pre-linearized plasmids above the background random integration with some spacers (Fig. 7); however, these were neither reproducible nor robust enough to base a method on. The underlying exact mechanism for the integration of the in-cell cleaved plasmids is not well understood. To distinguish between alternative scenarios, we in-cell re-cleaved the pre-linearized plasmid (Fig. 7, Supplementary Fig. S9). These experiments seem to indicate that the major factor for the difference is that circular DNAs are more resistant to cellular DNases than the linear ones are. However, it is still

not clear that by increasing the linear DNA pool in the cell, available for integration to the competing double strand breaks, why the in-cell cleaved plasmid is preferred to be integrated (at proportionally higher rates) to nuclease-generated over to endogenous DSBs (Fig. 1B) when one compares this to circular or pre-linearized plasmids (Supplementary Fig. S2).

The key step of the method is to use a self-cleaving plasmid that provides both controlled and standardized in-cell cleavage and enhanced integration of the plasmid. Another important feature lies in the use of targetless, low-background and effective spacers. The former characteristic was achieved by exploring some 30000 spacers and it is necessary for the low background we observed, as well as for general applicability. The latter probably increases the sensitivity of our method, that is in accord with Auer et al. showing that more efficient cleavage of the donor plasmid results in higher targeted integration in zebrafish.[38]

Our work also demonstrates an efficient way for targeted integration of DNA cassettes in mammalian cells without laborious construction of homologous arms although without base pair precision. Based on earlier studies it was not clear if NHEJ-cloning is also effective in mammalian cells following SpCas9 cleavage that generates blunt ends. In attempts to enhance its efficiency, a few approaches used compatible overhangs or edited microhomology (8 bases-long) to the DSB site.[43,44,52,67] Auer et al., building on the high activity of NHEJ in zebrafish embryos,[68–70] demonstrated high efficiency of NHEJ-cloning without any overhangs or edited microhomology.[38] However, when SpCas9 was used in mammalian cells for NHEJ-cloning of a donor plasmid without any overhangs or edited microhomology, none of the attempts demonstrated highly efficient targeted integration (0.17% and 0.45% or about 1%, in HEK and CHO or HAP1 cells, respectively).[45,46] Here, we successfully demonstrated that NHEJ-cloning coupled with a self-cleaving donor plasmid without any pre-selection or FACS-enrichments can be effectively applied (with up to 20–30% efficiency of integration) to mammalian cells as well.

A recent study demonstrated targeted integration of expression cassettes amplified by PCR, and reached targeted integration of mCherry-containing cassettes between 1 and 5%.[59] this is in agreement with our results presented here using linear plasmids (Fig. 7; Supplementary Figs S2 and S9). Interestingly, using the smaller and brighter protein Clover, about 20% integrations were reported suggesting that transfection of a linear donor DNA less than 2 kilobase, might integrate more efficiently. While it is very promising, in the absence of measuring the actual transfection efficiencies, the widely different amounts of DNA used for transfections of samples and controls, and in the absence of monitoring the decay of the transient fluorescence with an appropriate inactive dCas9 control, it is difficult to assess the real impact of these results. Another work demonstrated homology independent integration, donor DNA knock-in not only in cell culture model systems but *in vivo* as well.[71]

NHEJ-cloning is a useful tool for targeted integration when there is no need for absolute precise joining of the DNA ends or when a selection can be applied for the proper joining. Since the DNA fragment can be inserted in two orientations and in three frames, only some of the targeted integration events would result from in frame incorporation of the donor DNA, depending on whether precise NHEJ dominates over error-prone NHEJ or whether the actual sequences prefer certain deletions/insertion over others. Thus, it seems that it is generally not efficient enough for easy protein tagging without selection. Even so, a few studies reported successful application of NHEJ-cloning to tag endogenous proteins using fluorescent protein or other selectable markers.[46,67,72] Nevertheless, it might be more suitable for IRES tagging that is without much sensitivity for base pair-precise insertion and thus, about every second targeted integration (one of the two orientations of the inserted DNA) is appropriate for proper tagging, provided that sufficiently high activity gRNAs are employed. It is important to note that, using only one site for in-cell linearization also results in the incorporation of bacterial, 'junk' DNA. In applications where avoiding the integration of 'junk' DNA is preferable, a more precise integration of the target gene/DNA cassette can also be achieved by employing the design on Figure 1A, using two cleavage sites as also demonstrated by Nakade et al. and Lackner et al.[46,72] A specifically attractive approach, CRISPaint was developed recently, by which both the integration of the plasmid-cleaving gRNA into the target genomic site and, by applying DNA minicircles, the insertion of bacterial DNA can be avoided.[73]

NHEJ-cloning is especially attractive for its potential as a superior substitute for random integration, to incorporate DNA cassettes to safe-harbour sites in the genome—avoiding the positional effects of the insertion—frequently encountered using random integration and the laborious work for constructing homologous arms. It can be more effective by at least one order of magnitude than random integration and provides stronger coupling between target and marker genes (Supplementary Fig. S11), facilitating a more reliable identification of the proper clones. Another attractive application of it is the creation of loss-of-function alleles by the easy targeting of DNA at endogenous loci. Using self-cleaving plasmids to disrupt genes is comparably effective to out-of-frame indel-generation methods with the added advantage of the invalidated alleles being labeled.

5. Conclusion

We demonstrated a versatile, yet simple approach for the identification of the effective gRNAs using a fluorescent reporter assay with self-cleaving plasmids. The method is capable of testing targets in their genomic context and requires no extra cloning steps. While it needs about 4 days longer to get the results, the method requires considerably less hands-on time and it certainly will be favoured by researchers who wish to avoid the use of the 'gold-standard' T7/Surveyor assays. This method might generally be used for other RNA-guided designer nucleases as well, such as Cas9-s from other species and Cpf1 nucleases. Additionally, we showed that NHEJ-cloning can also be used in mammalian cells for targeted integration of donor plasmids up to 10 kb in size, with up to 30% efficiency, without any selection or enrichment. Its further advantages as compared to random integrations include being much more effective and being able to provide stronger coupling of two DNA cassettes, such as the one to be integrated and a selection marker.

Acknowledgements

We would like to thank György Várady and Izabel Patik for helping with the FACS measurements, Lucia Illy, Judit Tálas and Zsófia Molnár for critical reading and suggestions to the manuscript and Ildikó Pulinka Szűcsné, Judit Szűcs and Luca Xénia Turgyán for technical assistance. This work was supported by the Hungarian Scientific Research Fund [K-82090, PD-111964].

Competing interests

The authors declare no competing interests.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at DNARES online.

Funding

This work was supported by the Hungarian Scientific Research Fund [K-82090, PD-111964].

Availability

The self-cleaving plasmids used in this study are available at Addgene: pSc1 (#80436), pSc2 (#80437), pSc1-puro (#80438) and pSc1-DD (#80439). Sequences of all other constructs are listed in Supplementary Data and are available upon request.

Author's contributions

A.T. and E.W. conceived and designed the experiments, A.T., P.I.K., N.W., A.B, K.S., E.T., S.L.K., B.G., A.S., O.I.H, Z.L., performed the experiments and all authors analysed the data. A. T., E. F. and E.W. wrote the manuscript.

References

1. 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**, 816–21.
2. Mali, P., Yang, L., Esvelt, K. M., et al. 2013, RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–26.
3. Cong, L., Ran, F. A., Cox, D., et al. 2013, Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–23.
4. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. 2013, RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.
5. Gilbert, L. A., Larson, M. H., Morsut, L., et al. 2013, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442–51.
6. Shalem, O., Sanjana, N. E., Hartenian, E., et al. 2014, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–7.
7. Zalatan, J. G., Lee, M. E., Almeida, R., et al. 2015, Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–50.
8. Gilbert, L. A., Horlbeck, M. A., Adamson, B., et al. 2014, Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*, **159**, 647–61.
9. Qi, L. S., Larson, M. H., Gilbert, L. A., et al. 2013, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–83.
10. Ramakrishna, S., Dad, A.-B. K., Beloor, J., Gopalappa, R., Lee, S.-K. and Kim, H. 2014, Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res.*, **24**, 1020–7.
11. Mali, P., Aach, J., Stranges, P. B., et al. 2013, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.*, **31**, 833–8.
12. Ran, F. A., Hsu, P. D., Lin, C.-Y., et al. 2013, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–9.
13. Dow, L. E. 2015, Modeling Disease In Vivo With CRISPR/Cas9. *Trends Mol. Med.*, **21**, 609–21.
14. Kim, H. and Kim, J.-S. 2014, A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.*, **15**, 321–34.
15. Ramakrishna, S., Cho, S. W., Kim, S., et al. 2014, Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat. Commun.*, **5**, 3378.
16. Doench, J. G., Hartenian, E., Graham, D. B., et al. 2014, Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol* **32**, 1262–7.
17. Wang, T., Wei, J. J., Sabatini, D. M. and Lander, E. S. 2014, Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–4.
18. Chari, R., Mali, P., Moosburner, M. and Church, G. M. 2015, Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods*, **12**, 823–6.
19. Moreno-Mateos, M. A., Vejnar, C. E., Beaudoin, J.-D., et al. 2015, CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods*, **12**, 982–8.
20. Wong, N., Liu, W. and Wang, X. 2015, WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.*, **16**, 1–8.
21. Doench, J. G., Fusi, N., Sullender, M., et al. 2016, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol* **34**, 184–91.
22. Hinz, J. M., Laughery, M. F. and Wyrick, J. J. 2015, Nucleosomes inhibit Cas9 endonuclease activity in vitro. *Biochemistry*, **54**, 7063–6.
23. Horlbeck, M. A., Witkowsky, L. B., Guglielmi, B., et al. 2016, Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife*, **5**, e12677.
24. Hendel, A., Fine, E. J., Bao, G. and Porteus, M. H. 2015, Quantifying on- and off-target genome editing. *Trends Biotechnol.*, **33**, 132–40.
25. Chapman, K. M., Medrano, G. A., Jaichander, P., et al. 2015, Targeted germline modifications in rats using CRISPR/Cas9 and spermatogonial stem cells. *Cell Rep.*, **10**, 1828–35.
26. Dow, L. E., Fisher, J., O'Rourke, K. P., et al. 2015, Inducible in vivo genome editing with CRISPR-Cas9. *Nature Biotechnol.*, **33**, 390–4.
27. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F. 2013, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–308.
28. Cho, S. W., Kim, S., Kim, J. M. and Kim, J.-S. 2013, Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, **31**, 230–2.
29. Qiu, P., Shandilya, H., D Alessio, J. M., O Connor, K., Durocher, J. and Gerard, G. F. 2004, Mutation detection using Surveyor^{TM} nuclease. *Biotechniques*, **36**, 702–7.
30. Dahlem, T. J., Hoshijima, K., Jurynec, M. J., et al. 2012, Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.*, **8**, e1002861.
31. Wang, X., Wang, Y., Wu, X., et al. 2015, Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–8.
32. Ramlee, M. K., Yan, T., Cheung, A. M., Chuah, C. T. and Li, S. 2015, High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. *Sci. Rep.*, **5**, 15587.
33. Brinkman, E. K., Chen, T., Amendola, M. and van Steensel, B. 2014, Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, **42**, e168.
34. Yang, L., Guell, M., Byrne, S., et al. 2013, Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.*, **41**, 9049–61.
35. Urnov, F. D., Miller, J. C., Lee, Y.-L., et al. 2005, Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–51.
36. Wagner, J. C., Platt, R. J., Goldfless, S. J., Zhang, F. and Niles, J. C. 2014, Efficient CRISPR-Cas9-mediated genome editing in *Plasmodium falciparum*. *Nat. Methods* **11**, 915–8.
37. Wang, H., Yang, H., Shivalila, C. S., et al. 2013, One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910–8.
38. Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. and Del Bene, F. 2014, Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.*, **24**, 142–53.

39. Certo, M. T., Ryu, B. Y., Annis, J. E., et al. 2011, Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods*, **8**, 671–6.
40. Mashiko, D., Fujihara, Y., Satouh, Y., Miyata, H., Isotani, A. and Ikawa, M. 2013, Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci. Rep*., **3**, 3355.
41. Hockemeyer, D., Soldner, F., Beard, C., et al. 2009, Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol*., **27**, 851–7.
42. Kim, H., Um, E., Cho, S.-R., Jung, C., Kim, H. and Kim, J.-S. 2011, Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods*, **8**, 941–3.
43. Orlando, S. J., Santiago, Y., DeKelver, R. C., et al. 2010, Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology. *Nucleic Acids Res*., **38**, e152.
44. Cristea, S., Freyvert, Y., Santiago, Y., et al. 2013, In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng*., **110**, 871–80.
45. Bachu, R., Bergareche, I. and Chasin, L. A. 2015, CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. *Biotechnol. Bioeng* **112**, 2154–62.
46. Lackner, D. H., Carré, A., Guzzardo, P. M., et al. 2015, A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat. Commun*., **6**, 10237.
47. Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S. A., Zhang, S. and Pederson, T. 2015, Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. USA*, **112**, 3002–7.
48. Tóth, E., Huszár, K., Bencsura, P., et al. 2014, Restriction enzyme body doubles and PCR cloning: on the general use of type IIs restriction enzymes for cloning. *PloS One*, **9**, e90896.
49. Stewart, S. A., Dykxhoorn, D. M., Palliser, D., et al. 2003, Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA*, **9**, 493–501.
50. Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D. and Wandless, T. J. 2010, A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol*., **17**, 981–8.
51. Perez-Pinera, P., Ousterout, D. G., Brown, M. T. and Gersbach, C. A. 2012, Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res*., **40**, 3741–52.
52. 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**, 539–46.
53. Heigwer, F., Kerr, G. and Boutros, M. 2014, E-CRISP: fast CRISPR target site identification. *Nat. Methods*, **11**, 122–3.
54. Aach, J., Mali, P. and Church, G. M. 2014, CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes. *bioRxiv*, 005074.
55. Xue, W., Chen, S., Yin, H., et al. 2014, CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380–4.
56. Fujii, W., Kawasaki, K., Sugiura, K. and Naito, K. 2013, Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Res*., **41**, e187.
57. Pinder, J., Salsman, J. and Dellaire, G. 2015, Nuclear domain 'knock-in'screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res*., **43**, 9379–92.
58. Li, K., Wang, G., Andersen, T., Zhou, P. and Pu, W. T. 2014, Optimization of genome engineering approaches with the CRISPR/Cas9 system. *Plos One*., **9**, e105779.
59. Geisinger, J. M., Turan, S., Hernandez, S., Spector, L. P. and Calos, M. P. 2016, In vivo blunt-end cloning through CRISPR/Cas9-facilitated non-homologous end-joining. *Nucleic Acids Res*., **44**, e76–e76.
60. Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X. and Zhang, F. 2016, Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–8.
61. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., et al. 2016, High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–5.
62. Mladenov, E. and Iliakis, G. 2011, Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat. Res*., **711**, 61–72.
63. Bae, S., Kweon, J., Kim, H. S. and Kim, J.-S. 2014, Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods*, **11**, 705–6.
64. McVey, M. and Lee, S. E. 2008, MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*., **24**, 529–38.
65. Vouillot, L., Thélie, A. and Pollet, N. 2015, Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3*, **5**, 407–15.
66. Huang, Y., Zhao, S., Shi, M., Liu, J. and Liang, H. 2012, Microchip electrophoresis coupled with on-line magnetic separation and chemiluminescence detection for multiplexed immunoassay. *Electrophoresis*, **33**, 1198–204.
67. Sakuma, T. 2015, Front-line of genome editing technology for animal cell engineering. *BMC Proceed.,* **9**, O1.
68. Liu, J., Gong, L., Chang, C., Liu, C., Peng, J. and Chen, J. 2012, Development of novel visual-plus quantitative analysis systems for studying DNA double-strand break repairs in zebrafish. *J. Genet. Genomics*, **39**, 489–502.
69. Dai, J., Cui, X., Zhu, Z. and Hu, W. 2010, Non-homologous end joining plays a key role in transgene concatemer formation in transgenic zebrafish embryos. *Int. J. Biol. Sci*., **6**, 756.
70. Hagmann, M., Bruggmann, R., Xue, L., et al. 1998, Homologous recombination and DNA-end joining reactions in zygotes and early embryos of zebrafish (*Danio rerio*) and *Drosophila melanogaster*. *Biol. Chem*., **379**, 673–82.
71. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., et al. 2016, In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144–9.
72. Nakade, S., Tsubota, T., Sakane, Y., et al. 2014, Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun*., **5**, 5560.
73. Schmid-Burgk, J. L., Höning, K., Ebert, T. S. and Hornung, V. 2016, CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun*., **7**, 12338.