APPLICATION NOTE

Fluorescent tagging of an endogenous gene by homology-directed repair using Dharmacon Edit-R CRISPR-Cas9 reagents.

Authors

John A. Schiel
Maren M. Gross
Eldon T. Chou
Melissa L. Kelley
Anja van Brabant Smith

Revvity, Inc.

Abstract

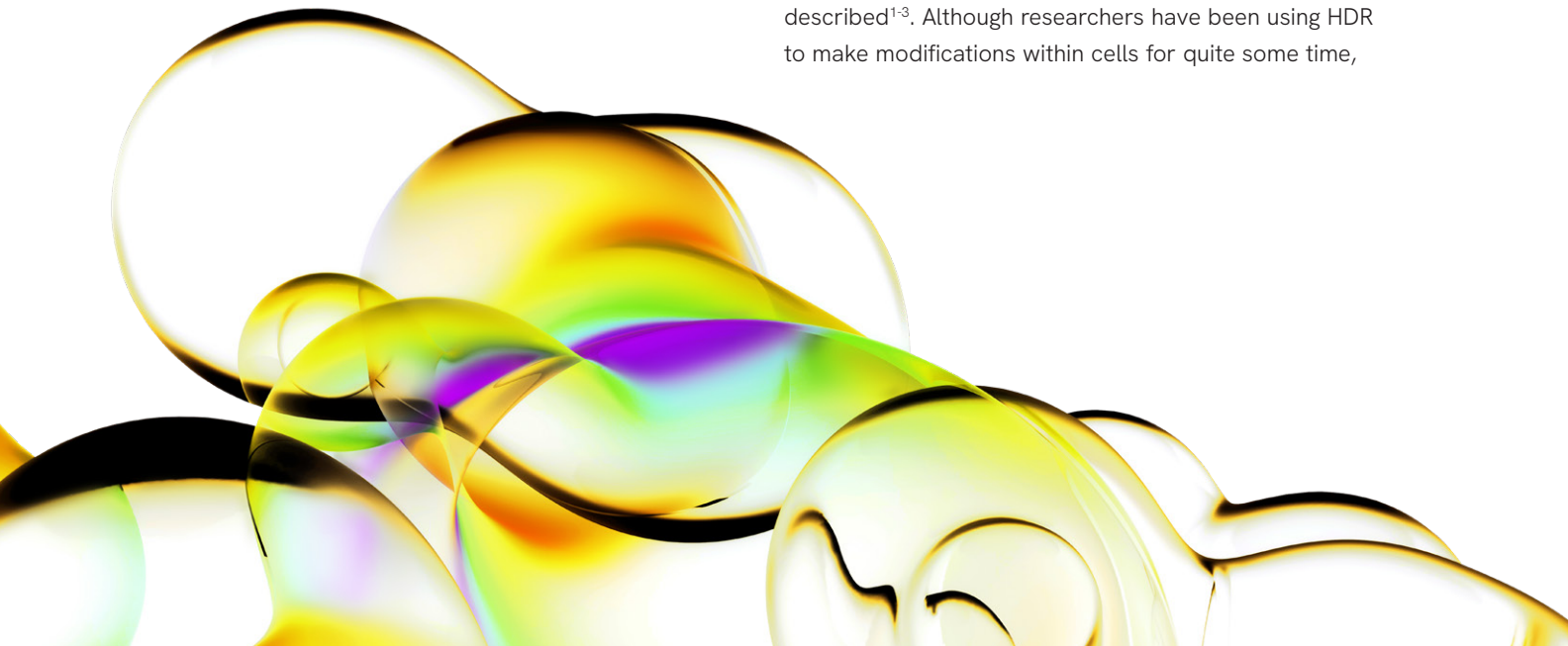
CRISPR-Cas9 has increased the accessibility of genome engineering due to its ease of use and ability to cause double-strand breaks (DSBs) at almost any locus of interest. The DSB that is created by Cas9, when guided to a specific genomic locus by crRNA and tracrRNA, can be repaired by various pathways, including homology-directed repair (HDR). The choice of repair pathway depends on numerous factors, including the presence of repair proteins and a repair donor template. Here we utilize the HDR pathway in conjunction with Dharmacon™ Edit-R™ CRISPR-Cas9 reagents and a DNA donor plasmid to create a fluorescent reporter-gene fusion with an endogenous gene target.

Keywords

Homology-directed repair, homologous recombination, non-homologous end joining, gene editing, genome engineering, Dharmacon Edit-R, Cas9, CRISPR, donor plasmid, donor DNA, homology arms, knockin, NHEJ, HDR

Introduction

When a double-strand break (DSB) is created in the genome, it can be repaired by several different repair pathways, predominantly non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ; also known as alt-NHEJ), and homology-directed repair (HDR). Detailed descriptions of these pathways have previously been described[1-3]. Although researchers have been using HDR to make modifications within cells for quite some time,



the HDR pathway has been characteristically inefficient, and has required the use of a donor construct containing several kilobases (kb) of homologous DNA to create precise genomic modifications. Since the advent of site-specific nucleases, we have seen HDR efficiencies increase through creation of targeted DSBs to force the cell to initiate DSB repair pathways.

With the ability to efficiently target most genomic loci for genome engineering with CRISPR-Cas9 technologies, researchers are able to more efficiently introduce exogenous DNA into the genome. These changes are often insertions of single nucleotide polymorphisms (SNPs), corrections of single base mutations, insertions of fluorescent tags or reporter genes, or addition of genes or other pieces of DNA into a safe harbor site[4-6]. Depending upon the scientific question at hand, and the size of the DNA to be inserted within the genome, multiple factors will impact the efficiency of targeted HDR.

Required components: There are several components that need to be introduced into a cell for a CRISPR-Cas9 HDR experiment: Cas9 nuclease (plasmid, mRNA, recombinant protein, or a stable cell line), guide RNA (gRNA) (synthetic crRNA:tracrRNA, synthetic sgRNA, or sgRNA expressed from a plasmid), and a donor template (single-stranded DNA oligo or plasmid). Depending on the source of Cas9, gRNA, donor template, and cell type, the best delivery method should be determined and optimized. This optimization is essential to maximize the number of DSBs for sequence insertion with HDR as discussed further below.

Additionally, it is important to decide on the correct type of donor DNA template (single-stranded DNA oligo or plasmid), and the design of the homology arms for HDR. For short insertions such as small nucleotide (nt) changes (< 50 nt), synthetic donor oligos with 30-40 nt homology arms will be the most efficient method[7]. For large DNA insertions, such as creating a reporter fusion, donor plasmids with individual homology arm lengths between 500 and 1000 nt are required[6,9].

Selection of cut site by CRISPR-Cas9: HDR starts and ends with the creation and repair of a DSB. Therefore, it is important to ensure highly efficient generation of DSBs using the crRNA-guided Cas9 nuclease. For HDR experiments, selection of the crRNA target site is a balance between the efficiency of the DSB caused by the Cas9 nuclease and the distance of the DSB from the desired HDR insertion/alteration site. It has been reported that mammalian gene conversion tracts are extremely short; efficiency of insertion of foreign DNA into a DSB by HDR decreases by 17% when the insert is > 10 nt away from the cut site, and as much as 87% when the insert is 100 nt away from the cut site[8]. Therefore, it is important to assess the proximity of crRNAs to the insertion site and balance this with the crRNA cutting efficiency in order to determine the optimal crRNA for precise genome engineering.

Donor design considerations: When designing a donor plasmid, it is important to remember that Cas9 can target the donor template itself if it contains an intact protospacer sequence next to a protospacer adjacent motif (PAM). Therefore, in order to avoid Cas9 cutting the donor template, it is important to design the donor template with one or more of the following changes (Figure 1).

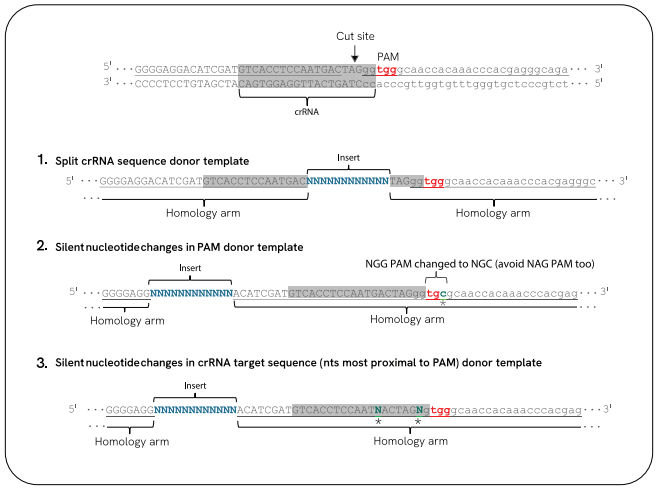


Figure 1: **Design homology arms that avoid additional rounds of cutting & NHEJ repair after insertion of desired sequence by HDR.** Three types of DNA donor oligo design considerations are presented to avoid Cas9 cutting. **1.** A DNA donor oligo with homology arms specific to the gene target (underlined) flanking the insert to be integrated via HDR (blue text).The insert separates all, or a part of, the crRNA target sequence from the PAM (red text). **2.** A DNA donor oligo with the right homology arm containing a nucleotide substitution (green text and *) in the PAM. **3.** A DNA donor oligo with the right homology arm containing two nucleotide substitutions (green text and *) in the crRNA target sequence.

1. Split the protospacer sequence (crRNA target sequence), or a portion of the protospacer sequence, and the PAM between the two homology arms of the donor plasmid so that the intended DNA insert is between these two elements;
2. Introduce silent SNPs in the *S. pyogenes* NGG PAM; or
3. Introduce silent SNPs in the crRNA target region most proximal to the PAM to prevent subsequent recognition by crRNA. (Figure 1)

Finally, it is not always possible to find an efficient and specific crRNA targeting sequence that is within ~ 100 nt of your intended insertion site. In these circumstances, the addition of a positive selection element to your donor plasmid, such as antibiotic selection, has been shown to aid in selection of the modified cells that contain the desired insertion[10].

Workflow overview: By using the recommendations outlined above, we have developed an experimental HDR workflow in which we use Cas9 mRNA, synthetic crRNA:tracrRNA, and a donor template to make precise HDR changes to a gene. Specifically, we demonstrate knockin of a GFP reporter to a gene target, creation of clonal cell lines, and confirmation of the precise insertion of GFP by Sanger sequencing and fluorescent microscopy.

The HDR workflow outlined here uses a lipid transfection method for delivery of the required components. Important steps for successful insertion of the desired sequence include:

1. Optimizing the creation of Cas9-mediated DSBs by identifying the crRNA with highest editing efficiency
2. Designing an effective DNA donor plasmid
3. Identifying the optimal amount of donor template that does not significantly decrease the amount of DSBs
4. Determining the optimal time point for analysis in order to avoid potential background
5. Using the appropriate assays to confirm precise insertion in clonal cell lines

Results

Optimizing creation of DSBs by selection of the most efficient crRNA: We set out to add a tag to the endogenous gene *SEC61B* (Gene ID: 10952) by inserting TurboGFP™ (Evrogen, Moscow, Russia) in-frame at the N-terminus of the protein. Two crRNA sequences (crRNA1-CCCUCAUCUCCAAUAUGGUA and crRNA2-CCAUACCAUAUUGGAGAUGA) nearest to the intended TurboGFP insertion site were chemically synthesized and transfected individually, each with Dharmacon Edit-R synthetic tracrRNA, into a U2OS-CAG-Cas9-integrated cell line. We characterized the editing efficiency of each crRNA to create DSBs in the absence of a donor plasmid by a mismatch detection assay using T7 endonuclease I (T7EI). The resulting percentage of insertions and deletions (indels) for the most active crRNA, crRNA1, was 40% (data not shown). The Cas9 cleavage site for crRNA1 was at the intended TurboGFP insertion site and split the protospacer motif to avoid future rounds of genomic cutting by Cas9, thus making this an ideal site for use with HDR (Figure 2A).

Design of DNA donor plasmid: The endogenous HDR machinery must recognize a template with homology to the target to commence precise repair[3]. We designed a TurboGFP donor plasmid flanked on each side with approximately 1000 nt homology arms for *SEC61B* (Figure 2B). Homology arms were generated by amplifying genomic DNA from the cell lysate of the experimental U2OS cell line. Therefore, any SNP(s) present in the PCR amplicon from this cell line will be included in the donor plasmid. Upon successful knock in, the TurboGFP insert will be placed in-frame at the N-terminus of *SEC61B* (Figure 2B).

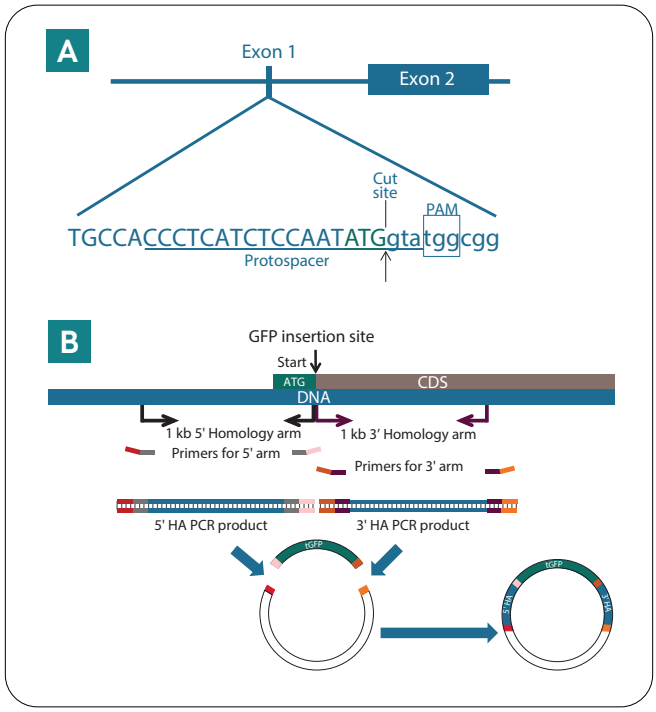


Figure 2. **Target gene for insertion of TurboGFP through a donor plasmid and the HDR pathway**. **A.** Illustration of the first two exons of *SEC61B*. Introns are indicated by lines and lowercase sequence. Exons are indicated by blue boxes and uppercase sequence. The ATG start codon is in green text. The protospacer sequence is represented by the underlined sequence. **B.** Primers were designed to amplify 1 kb 5′ and 3′ homology arms for insertion of GFP on the N-terminus of *SEC61B* and subsequently cloned into a vector along with TurboGFP (tGFP) to generate the donor plasmid. Primers used for construction of the donor plasmid, on average, were 45 nts in length and designed whereby half the primer sequence was specific to the area needed to amplify the homology arm (5′ arm gray arrows and bars; 3′ arm purple arrows and bars) and the other half was specific to the donor plasmid backbone (5′ arm red bars; 3′ arm orange bars). HA = homology arm.

Determination of assay time point: Optimization of the time interval between transfection and detection of HDR events will be dependent on the experiment-specific reagents. As controls, donor plasmids containing TurboGFP with or without homology arms were transfected without Cas9, crRNA, or tracrRNA into U2OS cells. While the TurboGFP plasmid without homology arms did not result in any GFP expression, approximately 10% of the cells transfected with the donor plasmid containing homology arms for *SEC61B* displayed cellular GFP expression 24 hours after transfection (Figure 3A). Expression of GFP from the donor plasmid is due to the presence of endogenous promoter region sequences for *SEC61B* in the ~ 1000 nt 5′ homology arm[11]. By monitoring the transfections over the course of seven days after the initial transfection, GFP expression from the donor plasmid diminishes each day post-transfection, with the nadir occurring seven days post-transfection (Figure 3B). Thus, in order to avoid this background GFP signal from the donor plasmid, subsequent GFP detection assays were performed at seven days post-transfection.

Detection of TurboGFP integration into the target locus: Various methods can be used to assess integration of exogenous DNA into a target locus. The methods of detection that we tested were high-content imaging, flow cytometry, PCR, and Sanger sequencing. Below we describe the detection of TurboGFP insertion at the desired locus following transfection of Cas9 mRNA, synthetic crRNA:tracrRNA, and donor plasmid into U2OS cells.

Detection of TurboGFP integration and expression by high-content imaging: U2OS cells were imaged at seven days post-transfection in order to determine expression of TurboGFP and correct localization of the target gene. SEC61B is part of the endoplasmic reticulum (ER) membrane and can also be observed in the post-ER compartment. Cells containing TurboGFP-tagged SEC61B were detected and shown to co-localize with antibody stained SEC61B (Figure 4), supporting in-frame insertion of the TurboGFP.

Detection of TurboGFP expression by flow cytometry: U2OS cells were analyzed by flow cytometry seven days post-transfection to determine the percentage of the population expressing TurboGFP. We determined that the donor plasmid-only control (no Cas9 mRNA, crRNA, or tracrRNA) contained 0.13% TurboGFP positive cells (Figure 5A), whereas the cells transfected with all required components for HDR insertion (Cas9 mRNA, crRNA:tracrRNA, and donor plasmid) contained 5.06% TurboGFP-positive cells (Figure 5B). Additionally, we grew 135 individual, flow cytometry-sorted, TurboGFP-positive cells in 96-well plates and determined that 98.5% (N=133) of those cell clones displayed the correct localization of SEC61B. These results are in agreement with our previous observation that expression of TurboGFP from the donor plasmid is only present in a very small fraction of the total TurboGFP-expressing cells when analyzed at seven days post-transfection.

Detection of donor plasmid integration by PCR amplification and Sanger sequencing: Primers that were optimized for the mismatch detection assay were used to amplify genomic DNA from lysate of untreated U2OS cells and TurboGFP-*SEC61B* positive and negative (determined by FACS analysis) clonal lines (Figure 6). PCR amplicons for the untreated control and TurboGFP-negative cell clone were detected at the expected length of 507 bp (clones UN and 1), whereas positive integration of TurboGFP at *SEC61B* was detected at the expected size of 1242 bp (clones 2-4; Figure 6A). In some cases, additional bands were also observed indicating indel formation by the NHEJ pathway in different alleles (clones 3 and 4; Figure 6A).

PCR amplicons from two individual GFP-positive cells (clones 3 and 4; Figure 6A) were cloned into a plasmid, transformed, and single bacterial colonies were sequenced. While GFP-positive clone 3 did have TurboGFP insertion in at least one allele, another allele did not have TurboGFP integrated and contained a 107 bp deletion (Figure 6A and C). Interestingly, GFP-positive clone 4 did have TurboGFP insertion in at least one allele, while another allele contained an incomplete copy of TurboGFP (Figure 6A and D).

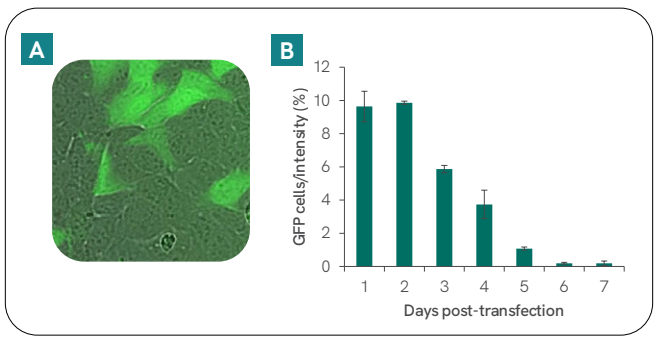


Figure 3. **Characterizing TurboGFP expression from a donor plasmid.** The donor plasmid for insertion of TurboGFP at the N-terminus of *SEC61B* (without Cas9, crRNA, or tracrRNA) was transfected into U2OS cells and TurboGFP expression was determined by fluorescent microscopy and flow cytometry. **A.** Fluorescent imaging at three days post-transfection shows expression of GFP. **B.** Flow cytometry-based detection of GFP positive cells at one to seven days post-transfection.

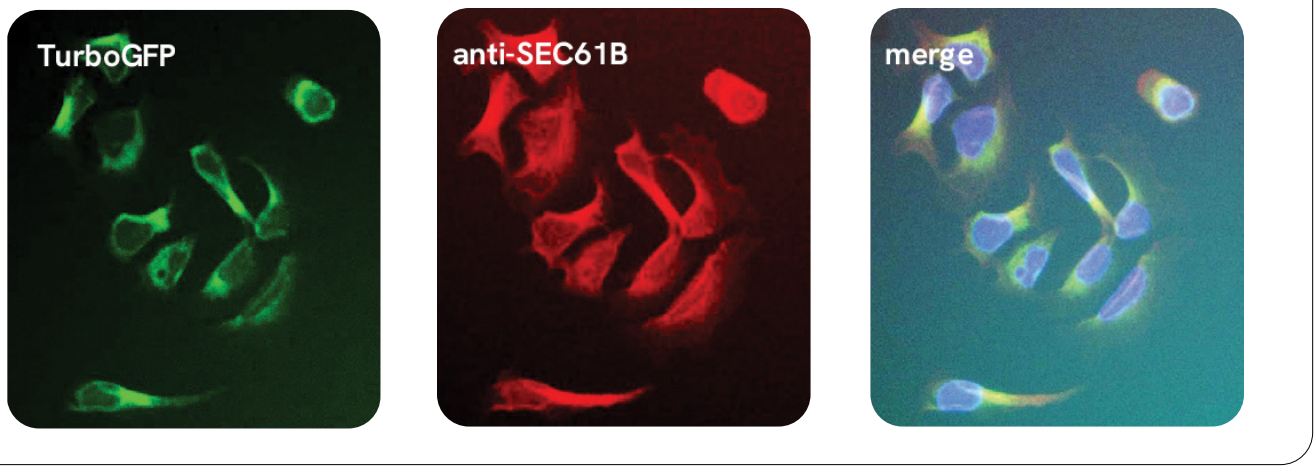


Figure 4. Co-localization of endogenous SEC61B with TurboGFP-tagged SEC61B in GFP-positive cell clones. A TurboGFP-positive SEC61B U2OS cell line was fixed and co-stained with a primary antibody to confirm correct TurboGFP localization.

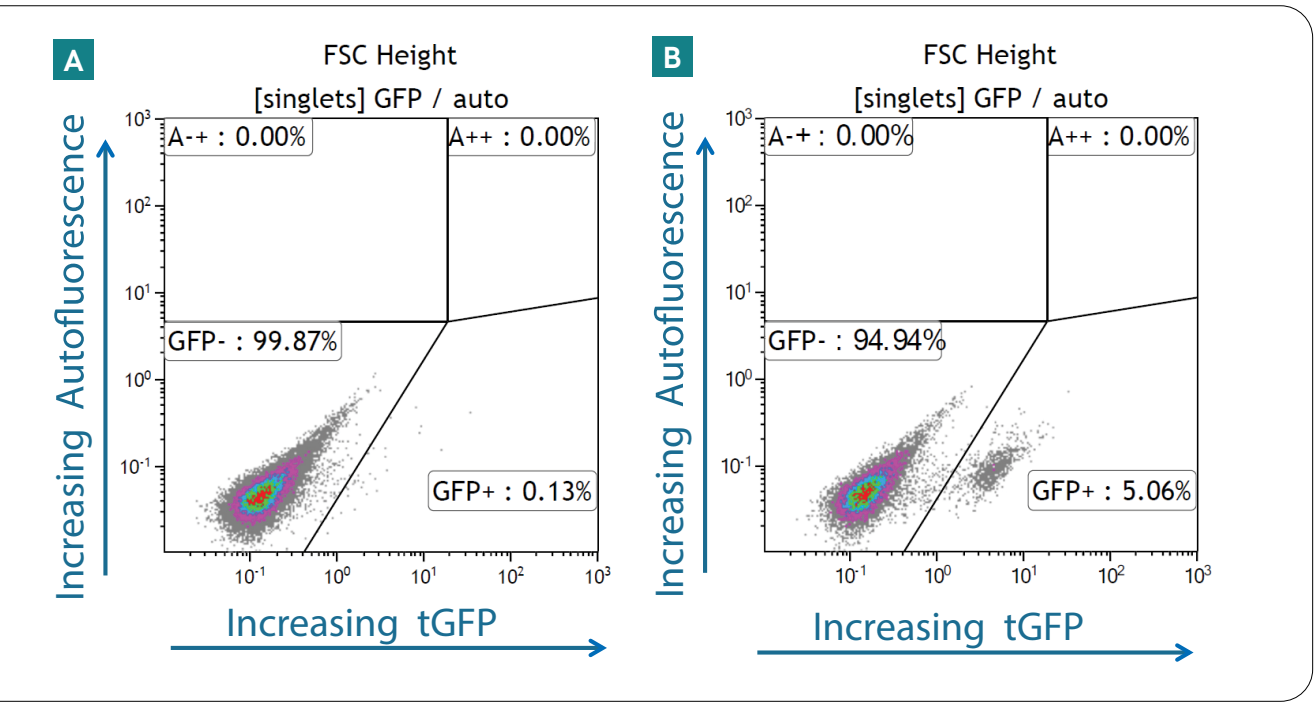


Figure 5: **FACS detection of GFP-positive cells from plasmid donor transfected cells.** U2OS cells were analyzed by FACS so that viable, single GFP-positive cells were gated, quantified, and sorted. GFP intensity is measured along the x-axis and auto fluorescence detection at 575 nm is measured along the y-axis. **A.** *SEC61B* donor only transfected cells (~ 25,000 cells measured). **B.** *SEC61B* HDR transfected cells (~ 18,000 cells measured). A-+ and A++ represent unquantified quadrants containing zero FACS detected cells.

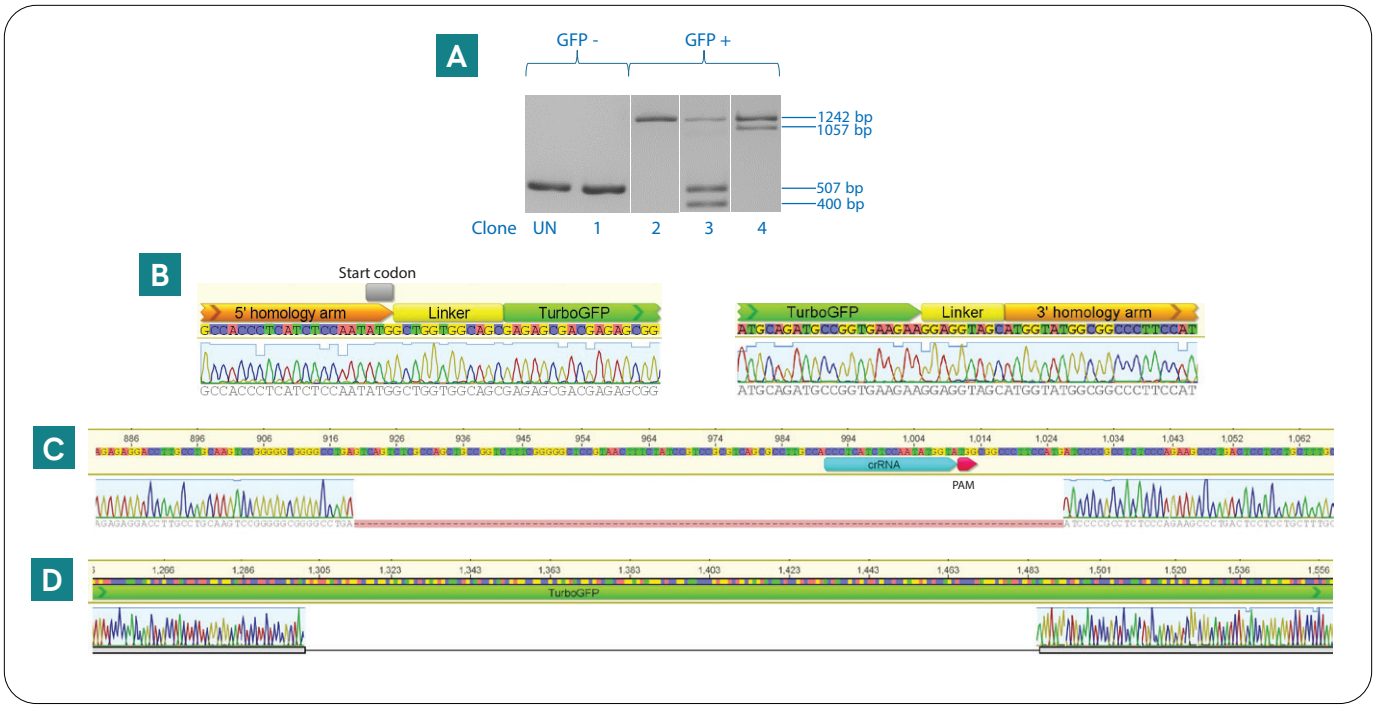


Figure 6: **PCR-based detection of TurboGFP integration. A.** PCR was performed across the intended insertion site for single cell clones of TurboGFP-*SEC61B* to detect a shift in the size of the PCR amplicon due to the TurboGFP insertion. UN = untreated. **B.** Sanger sequencing results of each TurboGFP integration junction in TurboGFP-*SEC61B* clone 2. **C.** Sanger sequencing results showing the indel from the 400 bp amplicon from TurboGFP-*SEC61B* clone 3. **D.** Sanger sequencing results showing the deletion of 185 bp (resulting in the 1057 bp band) from TurboGFP-*SEC61B* clone 4.

Discussion

Here we have demonstrated that fluorescent tagging of an endogenous protein can be accomplished with the Dharmacon Edit-R CRISPR-Cas9 genome engineering platform and a donor plasmid template. We inserted TurboGFP in-frame with the endogenous *SEC61B* gene and selected individual clonal cell lines, which were verified by Sanger sequencing. Transfection optimization is one of the most important experimental factors when performing HDR experiments since the HDR machinery will only repair a fraction of the DSBs created by Cas9 nuclease. It is also necessary to perform a donor-only transfection to evaluate false-positive effects from the donor; in the example provided, GFP-positive background was detected from the donor plasmid alone and not due to true genomic insertion of GFP. Importantly, GFP expression from a plasmid donor alone will diminish over time as the plasmid is degraded and diluted during cell culture maintenance.

When choosing a DNA donor for CRISPR-Cas9 HDR experiments, one must consider the size of the exogenous DNA to be inserted. In our experience, short inserts (< 50 nt) with DNA oligos can be up to three times more efficient then large inserts (e.g., GFP reporter) with donor plasmids. Additional biological factors may also contribute to variable HDR insertion, such as genomic accessibility, DSB and insert site distance, gRNA activity, and essential gene function.

The Dharmacon CRISPR design tool and the Dharmacon Edit-R HDR donor designer can assist in identifying an efficient guide RNA target site and proper design of the donor template.

Interestingly, when characterizing the HDR clonal cell lines, we detected expression of TurboGFP by fluorescent imaging; however, when amplifying the integration site, we detected smaller bands that differed from both the expected unintegrated PCR amplicon size and positive integration PCR amplicon size. Smaller bands from the unintegrated PCR amplicons likely occurred through the NHEJ pathway, while the integrated truncated version of TurboGFP might have originated from other repair pathways such as the microhomology-mediated end joining pathway. These results underscore the necessity of generating many clonal cell lines and fully characterizing each clonal cell line before deciding to progress to future experiments with your desired knockin cells.

In conclusion, we have discussed and shown best practices for HDR workflows, including crRNA site selection, DNA donor design, transfection optimization, assay time point assessment, and detection of successful integration using several methods resulting in precise tagging of an endogenous protein, SEC61B, with TurboGFP. These methods and recommendations are applicable to other HDR-mediated insertions of large DNA into different genomic locations.

Materials and methods

Tissue culture: All U2OS cells were maintained in standard growth medium per manufacturer's recommendations (ATCC, Cat #HTB-96).

Transfection: U2OS and U2OS-CAG-Cas9-integrated cells were seeded in a 96-well plate at 10,000 cells per well one day prior to transfection. Dharmacon Edit-R synthetic crRNAs (Dharmacon, custom synthesis) and tracrRNA (Cat #U-002000-20) were individually resuspended in 10 mM Tris-HCl (pH 7.5) to a concentration of 100 µM. crRNA and tracrRNA were combined at equimolar ratio and diluted to 2.5 µM using 10 mM Tris-HCl (pH 7.5). A final concentration of 25 nM crRNA:tracrRNA complex (25 nM of each crRNA and tracrRNA) was used for transfection. Cells were transfected with 200 ng of Dharmacon Edit-R Cas9 Nuclease plasmid (Cat #U-005100-120) or 200 ng Dharmacon Edit-R Cas9 Nuclease mRNA (Cat #CAS11195) using 0.3 µL/well of DharmaFECT™ Duo transfection reagent (Dharmacon Cat #T-2010-03) for co-transfections with crRNA:tracrRNA complex, with or without 200 ng donor plasmids.

Cell imaging: Live cell colonies were imaged with an IN Cell Analyzer 2200 Imaging System (GE Healthcare) after replacing the culture medium with imaging medium (HBSS, 20 mM HEPES, 16.8 mM D-glucose, pH 7.2). A 10x objective was used to identify TurboGFP-positive cells and then reimaged with a 20x objective for localization analysis. For immunofluorescent microscopy, cells were fixed with 4% paraformaldehyde (Electron Microscopy Services, Cat #15710) for 15 minutes at room temperature, washed twice with PBS, then incubated in permeabilization buffer (PBS containing 200 mg saponin, 2% FBS, and 100 µg BSA; Sigma, Cat #A3059) for 30 minutes at room temperature. Endogenous SEC61B was stained with a polyclonal rabbit antibody (ProteinTech Group, Cat #15087-1-AP) at 1:50 in permeabilization buffer for 1 hour, washed twice in PBS, followed by a secondary antibody (Alexa 568 goat anti-rabbit; Invitrogen, Cat #A11010) incubation at 1:250 dilution in permeabilization buffer for 30 minutes at room temperature. All cells were stained with Hoechst 33342 (Molecular Probes, Cat #H-3570) in PBS at a 1:2500 dilution for 10 minutes followed by two final washes in PBS.

Flow cytometry: U2OS cells were trypsinized, resuspended in cell sorting medium at 10 million cells/mL, and stored on ice until sorting. Cells were sorted on a Moflo XDP 100 cell sorting instrument by the Flow Cytometry Core at The University of Colorado Cancer Center [Cancer Center Support Grant (P30CA046934)]. Sorted cells were placed into tubes and 96-well plates, using FBS enriched medium (FBS:U2OS culture medium at 1:1 ratio).

Isolation of individual cell clones: A Moflo XDP100 flow cytometer was used to sort single TurboGFP-positive cells into individual wells of five 96-well plates for colony expansion. Penicillin-Streptomycin (2%; HyClone, Cat #SV30010) was added to normal U2OS growth medium as a precaution from exposure to the open air during sorting. Cells were maintained in culture and visually observed for GFP fluorescence for three weeks.

Genomic DNA isolation, PCR, and mismatch detection assay: Genomic DNA was isolated 72 hours post-transfection by direct lysis of the cells in Phusion™ HF buffer (Thermo Scientific, Cat #F-518L), proteinase K (Thermo Scientific, Cat #EO0491) and RNase A (Thermo Scientific, Cat #EN0531) for 20 minutes at 56 °C followed by heat inactivation at 95 °C for 5 minutes. PCR was performed with Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Cat #F530S) and primers flanking the cleavage/insertion sites. PCR products (500 ng) were treated with T7EI (NEB, Cat #M0302L) for 25 minutes at 37 °C and the samples were separated on a 2% agarose gel. Percent editing in each sample was estimated using ImageJ software (NIH, imagej.nih.gov/ij, 1997-2014).

Chromatograms were analyzed with Geneious Version 8.1.3, created by Biomatters (geneious.com)

References

1. J. Chapman, M. Taylor, et al., Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell.* 47(4), 497-510 (2012).
2. K. Chiruvella, Z. Liang, et al., Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect.* Biol. 5(9592), 1-21 (2013).
3. R. Ceccaldi, B. Rondinelli, et al., Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* 26(1), 52-64 (2016).
4. P. Mali, L. Yang, et al., RNA-guided human genome engineering via Cas9. *Science.* 339(6121), 823-826 (2013).
5. H. Yang, H. Wang, et al., One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell.* 154, 1370-1379 (2013).
6. M. Ratz, I. Testa, et al., CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. *Sci. Rep.* 5, 1-6 (2014).
7. B. Elliott, C. Richardson, et al., Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell Biol.* 18(1), 93-101 (1998).
8. J. Schiel, E. Chou, et al., Homology-directed repair with Dharmacon Edit-R CRISPR-Cas9 reagents and single-stranded DNA oligos. GE Healthcare, Dharmacon 2015; dharmacon.horizondiscovery.com/uploadedFiles/Re-sources/hdr-dna-oligo-text-appnote.pdf.
9. A. Hendel, E. Kildebeck, et al., Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep.* 7(1), 293-305 (2014).
10. S. Mansour, K. Thomas, et al., Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature.* 336(6197), 348-352 (1988).
11. R. Dreos, G. Ambrosini, et al., The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids* Res. 43 (Database issue), D92-D96 (2014).

Revvity, Inc.
940 Winter Street
Waltham, MA 02451 USA
www.revvity.com

For a complete listing of our global offices, visit www.revvity.com