Biotech Method

A Site-specific Integration Reporter System That Enables Rapid Evaluation of CRISPR/Cas9-mediated Genome Editing Strategies in CHO Cells

Nathaniel K Hamaker[1,2]

Kelvin H Lee[1,2]

[1]Delaware Biotechnology Institute, Newark, DE, USA

[2]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA

Correspondence: Dr. Kelvin H Lee, Department of Chemical and Biomolecular Engineering, University of Delaware, 590 Avenue 1743, 19713 Newark, USA

E-mail: KHL@udel.edu

Keywords: synthetic biology; biochemical engineering; biomolecular engineering; bioprocess development; CHO cells; CRISPR/Cas9, genome editing, site-specific integration

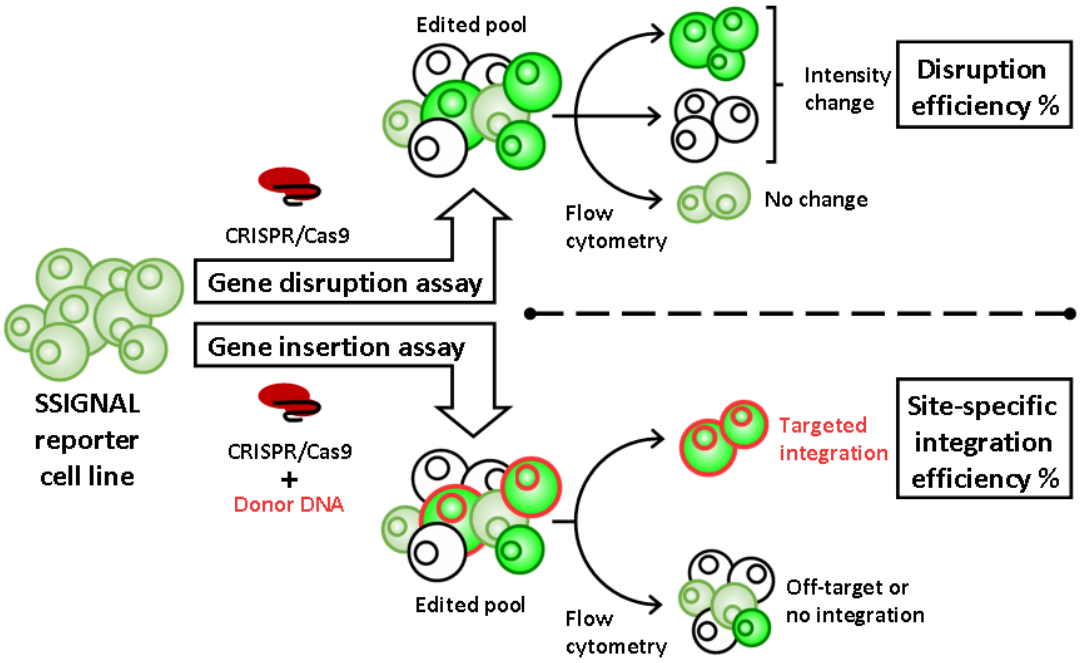
This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/biot.202000057.

Abbreviations: bGHpA, bovine growth hormone polyadenylation signal; **Cas9,** CRISPR-associated protein 9; **CHO,** Chinese hamster ovary; **CLD,** cell line development; **CRISPR,** clustered regularly interspaced short palindromic repeats; **DSB,** double-strand break; **eGFP,** enhanced green fluorescent protein; **FACS,** fluorescence-activated cell sorting; **gRNA,** guide RNA; **hCMV,** human cytomegalovirus major immediate-early promoter; **HDR,** homology-directed repair; **HMEJ,** homology-mediated end joining; **INDEL,** insertion or deletion; **KO,** knockout; **LP,** landing pad; **MMEJ,** microhomology-mediated end joining; **NHEJ,** non-homologous end joining; **PAM,** protospacer adjacent motif; **PEST,** sequence rich in proline, glutamic acid, serine, and threonine; **SSI,** site-specific integration; **SSIGNAL,** site-specific integration and genome alteration; **VCD,** viable cell density

Abstract

Targeted gene knockout (KO) and site-specific integration (SSI) are powerful genome editing techniques to improve the development of industrially relevant Chinese hamster ovary (CHO) cell lines. However, past efforts to perform SSI in CHO cells have been characterized by low efficiencies, necessitating tedious workflows to obtain a clone in which precise, rather than random, transgene integration has occurred. Moreover, numerous strategies have been proposed to boost SSI efficiency in mammalian cell types, yet few have been evaluated head to head or in combination to appreciably boost efficiencies in CHO. To enable systematic and rapid optimization of genome editing methods, we present the SSIGNAL (site-specific integration and genome alteration) reporter system, a tool which can be operated in either of two modes to analyze CRISPR (clustered regularly interspaced palindromic repeats)/Cas9 (CRISPR-associated protein 9)-mediated disruption activity alone, or in conjunction with SSI efficiency. The reporter system functions by using green and red dual-fluorescence signals to indicate genotype states

within four days following transfection, facilitating rapid data acquisition via standard flow cytometry instrumentation. In addition to describing the design and development of the system, we demonstrate two of its applications by first comparing transfection conditions to maximize CRISPR/Cas9 activity and subsequently assessing the efficiency of several promising SSI strategies. Due to its sensitivity and versatility, the SSIGNAL reporter system may serve as a tool to advance genome editing technology.



Gene knockout and site-specific integration are CRISPR/Cas9-mediated genome editing techniques that can be used to engineer recombinant CHO cell lines with improved biomanufacturing characteristics. Here, the SSIGNAL (site-specific integration and genome alteration) reporter system was developed to overcome low editing efficiencies and screening limitations by enabling rapid and simultaneous quantification of gene disruption and site-specific integration efficiencies. In this study, the authors explain how to implement the system and demonstrate its utility for evaluating promising genome editing strategies.

1 Introduction

Chinese hamster ovary (CHO) cells continue to be the expression platform of choice for the production of therapeutic glycoproteins, including monoclonal antibodies [1]. As an increasing number of therapeutic proteins enters the pipeline, genome editing tools have vastly expanded CHO cell line development (CLD) capabilities. Programmable endonuclease systems, mainly CRISPR (clustered regularly interspaced palindromic repeats)/Cas9 (CRISPR-associated protein 9), have been implemented to achieve site-specific integration (SSI) of transgenes as well as targeted gene knockout (KO) [2]. CRISPR/Cas9 is an RNA-guided nuclease that can be engineered to induce targeted DNA double-strand breaks (DSBs) within the genome. Gene KO occurs when DSB repair is mutagenic and produces nucleotide insertions or deletions (INDELs) that disrupt the reading frame of a protein coding sequence. This methodology has been used to disrupt a variety of native CHO genes to establish cell lines with customized traits, such as improved productivity [3], altered glycosylation patterns [4,5], and reduced levels of problematic host cell protein contaminants [6].

SSI is a complementary genome editing technique that has emerged as a powerful CLD strategy by enabling the generation of production cell lines with more predictable performance and product quality, thus accelerating timelines. In the traditional CLD workflow, random integration of the transgene encoding the recombinant protein of interest leads to high levels of clonal variation due to positional effects. This heterogeneity necessitates the tedious screening of multiple clones to isolate those with suitable performance characteristics. In contrast, SSI results in transgene knock-in events at designated genomic loci, including pre-validated hot spots capable of supporting highly active and stable expression [7,8]. In practice, the location of a hot spot is often discovered empirically and then a master cell line is established through integration of a landing pad (LP) [9–11]. Typically, the LP is a sequence that expedites

retargeting by harboring a marker gene, promoter trap, and recombinase target site(s) [12]. This master cell line can be used as a platform to make reliable production cell lines by simply swapping the LP marker with genes encoding recombinant proteins of interest through recombinase-mediated cassette exchange [13]. Nuclease-mediated SSI methods remain essential for performing de novo LP targeting of loci chosen a priori as the presence of chromosomal recombinase target sites is a prerequisite for using recombinase systems. Like gene KO, SSI relies on the induction of targeted DSBs using engineered endonuclease systems, such as CRISPR/Cas9. However, SSI also requires that the nuclease be introduced simultaneously with donor DNA, which becomes incorporated at the cut site. Several DSB repair strategies can be leveraged to facilitate the targeted integration of donor DNA in CHO cells. These strategies differ according to their homology requirements and the native DSB repair pathway involved. For instance, homology-directed repair (HDR) relies on homology arms that flank the donor DNA, allowing the donor to serve as a repair template [14]. Microhomology-mediated end joining (MMEJ)-based methods also require homology, albeit relatively short (≤40 bp) [15]. In contrast, the non-homologous end joining (NHEJ) pathway can directly ligate together the free ends created at the chromosomal break site with the ends of the donor DNA, obviating the need for homology arms [16–18].

SSI efforts in CHO cells have followed closely behind genome editing advancements made in human and mouse systems, though with a considerably larger emphasis on recombinant protein expression applications. While early SSI experiments in CHO cells employed NHEJ-mediated SSI to introduce transgenes [16,18], approaches that couple CRISPR/Cas9 with HDR donors have largely become the most popular strategy to achieve SSI following seminal work adapting this methodology in CHO [14]. Still, efficient and precise genome editing remains a major challenge, hindering CLD efforts, due to the low frequency of

HDR events relative to the occurrence of undesired edits that corrupt the target site. Some improvements have been made, including a fluorescent enrichment method that resulted in a 3-fold increase in targeted integrants [19]. Additionally, meticulous dissection of DSB repair in CHO cells has led to the identification of endogenous genes that can be modulated to boost SSI competency [20]. Recently, CRISPR/Cas9-mediated linearization of plasmids has proven to be an especially successful strategy for HDR-mediated integration, even enabling multi-site knock-in events [21].

Regardless of the SSI methodology utilized, transfected cells must ordinarily undergo an extensive cloning and screening process to assess whether proper insertion has occurred. Until now, this obstacle has rendered it impractical to systematically explore the breadth and relative efficiency of the many available SSI strategies that have yet to be evaluated in CHO cells [22–26]. To overcome the limitations of current methods, we have developed a versatile tool dubbed the SSIGNAL (site-specific integration and genome alteration) reporter system that facilitates the systematic evaluation of genome editing techniques, including strategies for both CRISPR/Cas9-mediated gene KO and SSI. The system provides rapid genotype readouts based on a dual-fluorescence flow cytometry assay by: (1) exploiting the destabilizing properties conferred by a PEST (sequence rich in proline, glutamic acid, serine, and threonine) degron tag to modulate the brightness of an enhanced green fluorescent protein (eGFP) harbored on a chromosomal reporter LP, and (2) restricting expression of a promoterless mCherry gene carried on an exogenous donor to only be activated once it becomes properly integrated within the LP (Figure 1A). In the absence of donor DNA, CHO cells harboring the chromosomal reporter are used to conduct an eGFP disruption assay (Figure 1B) to measure CRISPR/Cas9 editing efficiency and determine conditions that are most suitable for DSB induction and gene KO. Mutagenic repair of DSBs produces INDELs that lead to either a PEST-KO or LP-KO genotype as indicated by a

change in cellular eGFP brightness. Alternatively, the system functions as an SSI reporter assay (Figure 1C) when donor DNA is co-transfected with CRISPR/Cas9 components. Expression of mCherry is limited to precise in-frame integration events and concomitant with PEST-KO, thus providing dual indicators with which to accurately determine SSI efficiency. Furthermore, PCR-based production of donor DNA grants control over homology arm lengths, which we employ to evaluate DSB repair strategies, namely HDR, MMEJ, and NHEJ. Here, we provide a methodology for using the SSIGNAL reporter system.

2 Materials and methods

2.1 Plasmid Construction for Reporter System Landing Pad

An LP sequence was designed *in silico* to encode the chromosomal component of the SSIGNAL reporter system, consisting of a destabilized eGFP (d4eGFP) [27] coding sequence expressed from a human CMV (hCMV) promoter with a bovine growth hormone polyadenylation signal (bGHpA). Crucial to the functionality of the reporter system, the *d4egfp* open reading frame included a linker sequence (between the 5' *egfp* and 3' *mODC-PEST* degron tag) that harbors an artificial CRISPR/Cas9 target sequence. The *d4egfp* gene and bGHpA sequence were ordered as gBlocks® (Integrated DNA Technologies (IDT), Coralville, IA) and assembled along with the hCMV promoter from pcDNA™3.1/Zeo(+) (Thermo Fisher Scientific, Waltham, MA) and backbone of pCR™-Blunt II-TOPO® (Thermo Fisher Scientific) using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs (NEB), Ipswich, MA). The assembled plasmid pLP1-cb4Gb was cloned into NEB® 5-alpha competent *E. coli* (NEB) and sequenced (DBI Sequencing Center, Newark, DE). Additionally, a second nearly identical construct (pLP1-cbGb) that differed based on the absence of the degron tag was assembled and cloned in the same manner.

2.2 Cas9/guide RNA (gRNA) Design and Plasmids for CRISPR

The following artificial CRISPR/Cas9 target sequence (LP-gRNA target in caps, protospacer adjacent motif (PAM) underlined) was present within the *d4egfp* linker mentioned above: GCAACGTTACAAGAACACCGcgg. The LP-gRNA target site was designed to favor active CRISPR/Cas9 targeting according to previously established criteria [28]. The BLAST server on CHOgenome.org (http://www.chogenome.org) [29,30] was used to confirm that the site does not exist in the CHO-K1 genome within 2 bp mismatches.

The Rosa26 locus was chosen as the integration site of the LP containing the chromosomal component of the reporter system. The syntenic region of the murine Rosa26 locus was previously identified in CHO [11]. To integrate the LP at the corresponding locus (accession: NW_003613637.1: 132,459), the following CRISPR/Cas9 target sequence (R26-gRNA target in caps, PAM underlined) was used: GAGGCTTTCTGGGAGATGGGcgg.

Plasmids expressing LP-gRNA and R26-gRNA were generated as previously described [6]. Briefly, gRNA expression cassettes specific to each target sequence were ordered as gBlocks® (IDT) and cloned using a Zero Blunt™ TOPO™ PCR Cloning Kit (Thermo Fisher Scientific). Constructs were verified by Sanger sequencing (DBI Sequencing Center). Cas9 expression plasmid (hCas9), constructed by the Church group [31], was purchased from Addgene (41815).

2.3 Donor Plasmid and PCR Fragment Generation for SSI Efficiency Assay

The sequence of all SSIGNAL reporter system primers and plasmids used for donor fragment generation are provided in Supporting Table 1 and Supporting Table 2, respectively.

To assess HDR efficiency of donor DNA contained on a circular vector, a set of gBlocks® (IDT) were assembled as described above to yield pDNR2HDR encoding mCherry downstream and in-

frame with a 2A self-cleaving peptide linker. The promoterless *2A-mcherry* unit was flanked by ~850 bp homology arms corresponding to sequences adjacent to the cut site of the LP-gRNA target on the LP. To assess NHEJ, MMEJ and HDR efficiency of linear donor DNA, *2A-mcherry* fragments with varying homology arm lengths (0 – 700 bp) were amplified by PCR from pDNR2HDR using Phusion® Hot Start Flex 2x Master Mix (NEB) according to manufacturer's instructions. Generally, three 50 µL PCR reactions were necessary to generate sufficient product for one transfection. Completed PCR reactions corresponding to the same amplicon were pooled and purified using a QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and eluted in 50 µL MaxCyte Electroporation Buffer (MaxCyte, Gaithersburg, MD).

2.4 CHO Cell Culture

Suspension-adapted CHO-K1 cells (ATCC, Manassas, VA) were maintained in 125 mL Erlenmeyer flasks (Corning, Corning, NY) containing 20-30 mL SFM4CHO medium (Hyclone Laboratories Inc., Logan, UT) and passaged at 3-4 day intervals. Cultures were agitated in a 37°C incubation shaker at 135 rpm (25 mm throw diameter) with an atmosphere of 5% CO_2 and 80% relative humidity. Three days prior to transfection, cells were inoculated in 500 mL Erlenmeyer flasks (Corning) with 75-80 mL medium at a viable cell density (VCD) of 0.4-0.6 x 10^6 cells mL^{-1}.

2.5 Generation of a Landing Pad CHO Cell Line

To examine genome editing efficiency in CHO under various disruptive and SSI conditions, we established an LP cell line that hosts the chromosomal component of the fluorescence-based reporter system. A null CHO-K1 cell line (ATCC) was previously adapted to serum-free, suspension culture [32]. A 2.3 kb fragment corresponding to the P^{hCMV}-*d4egfp-bGHpA* LP cassette was amplified from pLP1-cb4Gb using a primer pair that modified the template by adding 12 bp microhomology arms to both ends. The microhomology arms corresponded to sequences

adjacent to the cut site of the R26-gRNA target in the CHO genome. Briefly, 7.5 x 10^6 cells were co-transfected with plasmids expressing Cas9 and R26-gRNA and the PCR-amplified LP cassette at a molar ratio of 1:4:2 (Cas9:gRNA:LP, 34.5 µg mL^{-1} total concentration) with a Lonza Nucleofector™ 2b system using Kit V and program U-024 according to the manufacturer's protocol. The transfected cell pool was monitored by flow cytometry to determine when the fluorescent population had stabilized, at which point, eGFP-positive cells were sorted twice by fluorescence-activated cell sorting (FACS) using a FACSAria™ II flow cytometer (Becton Dickinson (BD), Franklin Lakes, NJ). Single-cells were isolated via the limiting dilution method and clonality was verified by inspection under an inverted microscope. Clones were screened by genomic PCR to identify those in which SSI of the LP had occurred at the Rosa26 locus. The clone denoted CHO-LP1R26A4 was found to be SSI-positive, and a working cell bank was established so that the line could be used for all subsequent super-transfection experiments described in the next section. In parallel, the same cell line development steps were performed using cells transfected with a P^{hCMV}-*egfp-bGHpA* LP cassette amplified from pLP1-cbGb as a control so that the cellular eGFP intensity conferred by the destabilized protein could be compared to that conferred by the standard protein.

2.6 Transfections for eGFP Disruption and SSI Efficiency Assays

Cas9, LP-gRNA and donor (pDNR2HDR) plasmids were purified using an EndoFree® Plasmid Maxi Kit (Qiagen) and resuspended in water at a final concentration of 5 µg $µL^{-1}$. For each transfection, 1 x 10^7 cells in exponential phase were harvested by centrifugation (180 g, 7 min) in a 15 mL conical tube (Fisher Scientific, Hampton, NH). Cells were resuspended in MaxCyte Electroporation Buffer with DNA to a total volume of 100 µL and transferred to an OC-100 cuvette (MaxCyte). All transfections were performed using a MaxCyte ATX® transfection system which has preloaded electroporation protocols for CHO cells referred to as "CHO" (lower

energy) and "CHO2" (higher energy). To examine the effect of electroporation conditions on disruptive genome editing (i.e. eGFP disruption), cells were co-transfected in duplicate with Cas9 and LP-gRNA plasmids at varying molar ratios and total concentrations (300 or 400 µg mL^{-1}) using either the "CHO" or "CHO2" protocol. For SSI experiments, cells were co-transfected in duplicate with Cas9 and LP-gRNA plasmids and pDNR2HDR or PCR-amplified donor DNA at a molar ratio of 1:4:2 (Cas9:gRNA:donor, 320-410 µg mL^{-1} total concentration) using the "CHO2" protocol. Negative control SSI experiments were performed using the same electroporation conditions to those described above but with either LP-gRNA plasmid or donor DNA left out of the transfected DNA mixture. Immediately following transfection, cells were transferred to the bottom of a 125 mL Erlenmeyer flask (Corning) and allowed to recover without agitation for 40 min in a 37°C incubator at 5% CO_2 and 80% relative humidity. Recovered cells were resuspended in 10 mL prewarmed SFM4CHO medium and cultures were maintained in an incubation shaker as described above. Two days after transfection, cultures were supplemented with fresh medium to a total volume of 15 mL. For time course data, flasks were passaged four days after transfection at a VCD of ~0.3-0.6 x 10^6 cells mL^{-1}.

2.7 Flow Cytometry and Cell Culture Analyses

Starting two days after transfection, cells were monitored by flow cytometry and measurements were taken every other day until day eight using a BD Accuri™ C6 Plus flow cytometer. For eGFP disruption experiments (i.e. Cas9 and LP-gRNA plasmids only), cells were analyzed 4 days post-transfection using a BD FACSAria™ II flow cytometer. Gates were established on the plot of eGFP fluorescence intensity (eGFP-A) vs side scatter (SSC-A) for unedited control cells (transfected with Cas9 plasmid only) to determine the eGFP-stable, eGFP-destabilized, and eGFP-negative population fractions for edited cells. The absolute eGFP disruption efficiency was calculated as the sum of eGFP-stable and eGFP-negative population fractions. For SSI experiments, a gate was

first applied to determine the eGFP-stable population fraction as before. Next, control cells (transfected with Cas9 and LP-gRNA plasmids only) were used to establish an mCherry-positive gate on the plot of mCherry fluorescence intensity (mCherry-A) vs eGFP-A. The absolute SSI efficiency was then calculated by counting the number of cells within the intersection of the eGFP-stable and mCherry-positive subsets and dividing by the bulk transfected cell count. For both flow cytometry methods, a gate was applied on forward scatter (FSC-A) and SSC-A to analyze 50,000 events (i.e. live cells) per sample by filtering out debris events on both flow cytometers. Scatter plot figures represent at least 10,000 events. The FACSAria™ II was equipped for detection of eGFP (488 nm laser, 510/10 emission). The Accuri™ C6 Plus was equipped for detection of eGFP (488 nm laser, 533/30 emission) and mCherry (488 nm laser, 670 longpass filter). For two-color experiments, compensation was applied to correct mCherry-A by subtracting a percentage from eGFP-A. All flow cytometry analysis was performed using Flowing Software v2.5.1 (http://www.flowingsoftware.com) and BD CSampler™ Plus Software v1.0.27.1. VCD and viability were measured by the trypan blue exclusion method using a Vi-Cell XR (Beckman Coulter, Brea, CA).

2.8 Statistical Analyses

Data were processed and figure plots were generated using MATLAB (http://www.mathworks.com/products/matlab), RStudio (https://www.rstudio.com), and Microsoft Excel. Data in bar graphs are presented as mean ± 95% c.i. from at least two independent experiments. Student's t-test or one-way ANOVA was used accordingly to compare genome editing efficiencies and viabilities. A p-value of < 0.01 was considered significant.

3 Results and Discussion

3.1 eGFP Disruption Assay Measures CRISPR/Cas9 Activity

Existing genome editing reporter systems can generally be grouped according to whether they monitor intrachromosomal DSB repair outcomes or the integration of exogenous donor DNA. The former are well-suited for studying the activities of DSB repair pathways in their native context [33–35] and have also been utilized in the form of disruption assays as the gold standard for evaluating the activity of programmable nucleases [36]. The latter are designed for measuring the precise insertion of exogenous DNA into a chromosome [20]. The most versatile SSI reporter systems operate by exploiting native promoters as traps for knock-in reporter alleles to quantify SSI using donors designed for any DSB repair strategy [17,37]. Breaking the mold, the widely used Traffic Light Reporter and its derivatives were the first to enable monitoring of both mutagenic DSB repair and SSI [38,39]. However, the system is limited by the types of frameshift edits that can be directly quantified. Moreover, the reporter is only able to detect HDR-mediated SSI via gene correction, meaning the heterologous cargo size of the donor repair template cannot be functionally increased. From the existing foundation, we sought to build a reporter system capable of simultaneously measuring both gene KO and SSI in a manner that quantifies all productive edits (i.e. frameshift mutations) while maintaining flexibility with respect to the mechanisms that can be used to integrate the exogenous donor DNA.

To build our system, we first needed to establish a CHO cell line that harbors the chromosomal component of the SSIGNAL reporter. It was recently shown that the targeting efficiency of MMEJ-directed SSI was 10-17% in CHO cells after drug selection and single-cell cloning [40]. We modified the previously described approach by using PCR to make our MMEJ-donor rather than by cutting the vector with a programmable endonuclease to generate

microhomology arms. Here, the donor consisted of the LP cassette encoding destabilized eGFP with the artificial CRISPR/Cas9 target site upstream of the PEST sequence. CHO cells were transfected with LP donor DNA and CRISPR/Cas9 plasmids to facilitate targeted integration at the Rosa26 locus, which is a well-known 'safe harbor' for transgene expression in mammalian cells [41,42]. This locus was chosen because of its potential capacity for active and stable expression in CHO cells as well as to standardize positional effects when using the reporter system. eGFP-positive stable integrants, representing ~3% of the bulk pool seven days after transfection, were isolated by FACS and cloned without drug selection. Genomic DNA was extracted from the 17 clones with the highest eGFP fluorescence intensities so that both the left and right integration junctions could be verified by PCR to identify SSI-positive clones. While three clones generated incorrectly sized amplicons for one or both junctions, only one clone had both junction amplicons at the expected sizes (Supporting Figure 1), indicating proper LP integration at the intended site. Although an SSI event does not necessarily preclude the occurrence of off-target integration events within the same cell, the integration multiplicity was assumed to be low based on the transfection conditions used, and the SSI-positive clone was not tested to verify LP copy number. A cell line was expanded from the SSI-positive clone and used for all subsequently described SSIGNAL reporter system experiments.

To determine the efficacy of the reporter system as a tool to measure genome editing activity, we monitored the population dynamics of cells transfected with Cas9 and gRNA plasmids programmed to target the artificial linker sequence within the LP. While non-transfected cells had a uniform fluorescence intensity, the transfected cells split into three distinct subpopulations with differing fluorescence intensities (Figure 2A). Based on the finding that cells expressing destabilized eGFP were significantly dimmer than those expressing standard eGFP (Supporting Figure 2) due to reduced intracellular accumulation conferred by

the degron tag [27], we expected that CRISPR/Cas9-mediated genome editing of the LP would result in distinguishable fluorescence modes according to the type of disruption produced. Indeed, the flow cytometric readout supported the occurrence of the following three groupings: eGFP-destabilized, eGFP-stable, and eGFP-negative cells. eGFP-destabilized cells that remained at the same brightness characteristic of destabilized eGFP indicated a failure to receive both Cas9 and gRNA (i.e. no edit) or the occurrence of a non-disruptive in-frame INDEL (i.e. 3N-bp frameshift). eGFP-stable cells that became brighter indicated PEST-KO due to disruptive INDELs (i.e. 1N-bp and 2N-bp frameshifts), resulting in higher intracellular accumulation of eGFP which is stable in the absence of the degron tag. eGFP-negative cells that were no longer fluorescent indicated deletion events in any of the three reading frames that were sufficiently large to disrupt the functionality of eGFP entirely (i.e. LP-KO); large deletions have been previously characterized as a semi-regular outcome of CRISPR/Cas9-mediated editing in mammalian cells [43]. The three aforementioned genotype-phenotype interdependencies were used as a framework to interpret the results of the SSIGNAL reporter system, where eGFP-stable and eGFP-negative population fractions were summed to determine the absolute disruption efficiency as an indicator of CRISPR/Cas9 activity (see gating and calculation example in Supporting Figure 3). It should be noted that the disruption efficiency does not equal the total editing efficiency as the value does not include the unknown fraction of cells within the eGFP-destabilized population in which a non-disruptive in-frame INDEL occurred; however, such edits are typically considered unproductive in the context of a standard gene KO workflow. In addition to providing a comprehensive measure of productive CRISPR/Cas9-mediated editing events, the flow cytometric readout was obtained much faster than the time that would typically be necessary to clone and sequence transfected cells. Specifically, the absolute disruption efficiency became fixed by the fourth day after transfection (Supporting Figure 3B).

We next applied the reporter system to compare transfection parameters and determine which conditions achieve the highest genome editing activity. The following parameters were varied in combination: the molar ratio of Cas9 and gRNA plasmid transfected, the total DNA concentration dosed, and the electroporation protocol used. Notably, eGFP disruption efficiency only differed significantly between conditions with different electroporation protocols. The higher energy "CHO2" protocol gave the highest disruption efficiencies at 71-74%, whereas the efficiencies observed with "CHO" were lower by 10-15 percentage points (Figure 2B and Supporting Figure 3C). The levels of CRISPR/Cas9-mediated disruption achieved here were generally higher than those reported for native gene targets in unselected CHO cells [44–46], though target copy number and chromosomal context may account for this discrepancy [47]. Additionally, the collective results from all transfection conditions suggest that the cells were likely reaching a saturation limit with respect to plasmid expression since increasing both the ratio of gRNA to Cas9 and the total amount of loaded DNA failed to boost activity even when using the superior protocol. Cell viability across conditions was consistently high (97-98%) by 4 days post-transfection (Figure 2C) demonstrating that the MaxCyte electroporation system was not only effective for delivering sufficient DNA to achieve high levels of genome editing but also conducive to rapid cell recovery. Although only a subset of possible combinations was investigated here, the SSIGNAL reporter system could be readily applied to test disruption activity for other user-specific parameters to identify optimal KO conditions.

3.2 Fluorescent Reporter Is Suitable for Quantifying SSI Efficiency

Our second major aim for the reporter system was to enable sensitive and faithful quantification of SSI events. Attempts to perform targeted integration of transgene constructs in CHO cells are typically met with very low efficiencies, often comprising less than 10% of the bulk pool following transfection even when elaborate selection and enrichment strategies are

utilized [14,16,19,40]. To accurately detect these rare events, we constructed an HDR donor plasmid consisting of a promoterless mCherry gene flanked by homology arms to be used together with our LP cell line. Using the HDR donor plasmid as a template, linear donors with user-defined homology arm lengths from 0 to 700 bp could be generated via PCR. In conjunction with any arm length, the donor was assumed to be incapable of producing a fluorescent signal unless precisely integrated within the chromosomal reporter upon DSB induction and simultaneous disruption of the PEST degron tag (Figure 3A). With the expected SSI-positive phenotype known, the absolute SSI efficiency was calculated by dividing the number of cells that fell within gating criteria for both mCherry-positive cells and eGFP-stable cells (as determined using the same method as for the eGFP disruption assay) by the unselected bulk transfected cell count (see gating and calculation example in Supporting Figure 4). Given our single-laser flow cytometry setup, resolution of mCherry and eGFP-positive channels was limited, requiring both compensation and stringent gating criteria such that acceptable background within the mCherry-positive gate was set to 0.015% to accurately quantify rare SSI-positive cells (example data in Figure 3B). To ensure best performance and functionality of the SSIGNAL reporter system, we recommend a dual-laser approach as well as a large sample size of at least 50,000 cells.

To confirm the validity of the reporter system, a series of control experiments were performed by introducing donor DNA into cells with and without CRISPR/Cas9 components using the transfection conditions established above that had demonstrated high disruption efficiency. When gated against the negative control (i.e. no donor), conditions for which gRNA plasmid was omitted but that included linear donors without and with homology arms showed undetectable and nearly undetectable levels above background, respectively (Figure 3C). This result confirmed that the promoterless donor DNA design was effective as a non-functional

translational trap with a negligible rate of false positives due to unwanted transient expression or random integration. While transfected plasmid donor DNA in the absence of gRNA plasmid gave equivalent readouts to the linear donor controls (data not shown), the absolute SSI efficiency was 0.84 ± 0.03% when the HDR donor plasmid was introduced together with both CRISPR/Cas9 plasmids, indicating that the detected events were due to targeted integration into the LP. Here, the donor construct used was intended to imitate one that would be typical of an HDR-mediated SSI workflow, i.e. delivered in plasmid form with sizable homology arms (500-1000 bp) flanking the transgene. The measurement for HDR-mediated SSI efficiency provided by our reporter system was consistent with the efficiencies given or estimated from previous reports using CHO cells (~1%) [12,19].

3.3 Comparison of Repair Pathways for SSI: HDR Is Most Efficient

As a demonstration of the utility of the SSIGNAL reporter, we used the system to detect and quantify targeted integration events to compare the efficiencies of current genome editing strategies. The methods tested here differ according to the particular DSB repair pathway stimulated to achieve SSI, though all pathways yield the same knock-in genotype (Figure 3A). Moreover, it has been shown previously that the relative SSI efficiencies of these DSB repair strategies vary greatly depending on the cell line tested [37], but until now they had not been tested head to head in CHO cells. Usage of the different strategies was controlled via the length of homology arms generated on the PCR-amplified donor DNA fragments: 0 bp for NHEJ, 20 bp for MMEJ, and 250 bp or 700 bp for HDR. Comparable SSI efficiency values were calculated between day 4 and day 8 post-transfection, demonstrating the speed and reliability of the reporter system (Figure 3D). All conditions had comparable eGFP disruption efficiencies (Supporting Figure 5B), indicating that the observed differences in SSI efficiency were not caused by variations in CRISPR/Cas9 activity and presumably transfection efficiency. As NHEJ

and MMEJ are notoriously error-prone repair pathways especially when invoked during incorrect cellular contexts [48], we were curious to determine how many, if any, precise SSI events would be detected by our system when harnessing these pathways with the understanding that some on-target events may go undetected due to INDELs at the integration junctions. Although the efficiencies of both NHEJ and MMEJ-mediated SSI were comparably low (0.18% and 0.23%, respectively), they were statistically distinguishable from background and NHEJ-mediated SSI levels were consistent with previous reports in CHO [18]. Furthermore, HDR with linear donors was 0.65% and 1.90% efficient depending on homology arm length (250 bp and 700 bp, respectively). Targeting efficiency of HDR improved with increasing homology arm length as documented previously [49], though additional testing is necessary to determine a more exact functional dependence between these quantities in CHO cells. The NHEJ-0 and HDR-700 donors span the range of SSI efficiencies documented in this work, yet neither yielded significant levels when their corresponding controls lacking gRNA plasmid were tested (Figure 3C), underscoring the necessity of CRISPR/Cas9-induced DSBs within the mechanism of repair-dependent targeted integration strategies.

In prior works, the use of HDR-mediated SSI with linear donor DNA, sometimes referred to as homology-mediated end joining (HMEJ), was accomplished via nuclease-mediated cleavage and release of circular plasmids in vivo [21,37,50,51]. Nonetheless, we observed a 2.3-fold improvement in SSI efficiency when using a PCR-amplified linear HDR donor compared with a plasmid HDR donor (both with similar homology arm lengths), which matches previously documented efficiency gains when comparing the two donor formats [21,50,51]. Together, our findings show the functionality of the reporter system and promote the use of linearized HDR donors to improve SSI efficiency relative to other DSB repair strategies. Admittedly, the SSIGNAL reporter system is as of yet unable to quantify the occurrence of random integration

events, so differences in the total integration efficiency (i.e. off-target plus targeted) among conditions cannot be wholly discounted when considering which genome editing strategy is superior.

4 Concluding Remarks

We developed the SSIGNAL reporter system to facilitate rapid detection and quantification of both gene disruption and SSI events. This reporter system expands upon the capabilities of existing genome editing reporters by facilitating simultaneous monitoring of both of these genome editing techniques while providing more detailed editing readouts and maintaining compatibility with donor DNA formats that leverage diverse DSB repair strategies. From transfection to flow cytometric readout, assays take only four days to complete without requiring single-cell cloning or deep sequencing of the target locus. As demonstrated, the method can be applied to evaluate transfection conditions for effective gene KO by comparing CRISPR/Cas9 activity levels. Furthermore, we used the reporter system to test several current approaches for targeted integration to determine the most efficient DSB repair strategy in CHO cells. Ultimately, the development of robust SSI protocols may be implemented to improve long-term stability of transgenes for better CHO cell culture process predictability and control. Although not demonstrated in this study, the system can be used with other programmable endonuclease tools besides CRISPR/Cas9 and also allows for co-expression of auxiliary factors to modulate the activity of endogenous DSB repair pathways. Together, our results demonstrate the utility and sensitivity of the SSIGNAL reporter system and suggest that it may function as a valuable CLD tool by helping to further the advancement of genome editing in CHO cells.

Acknowledgement

We thank Dr. James Brady and Dr. Weili Wang for critical discussions and access to MaxCyte instrumentation; the University of Delaware DNA Sequencing & Genotyping Center; and Dr. Chen-Yuan Kao for assistance with FACS. This work was supported in part by the National Science Foundation (grant 1736123). N.H. was supported in part by Award Number T32GM008550 from the National Institute of General Medical Sciences. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. This work was supported, in part, by financial assistance award 70NANB17H002 from U.S. Department of Commerce, National Institute of Standards and Technology.

Conflict of interest

The authors declare no financial or commercial conflict of interest.

5 References

[1] G. Walsh, *Nat. Biotechnol.* **2018**, *36*, 1136.

[2] J.S. Lee, L.M. Grav, N.E. Lewis, H.F. Kildegaard, *Biotechnol. J.* **2015**, *10*, 979.

[3] A. Ritter, T. Rauschert, M. Oertli, D. Piehlmaier, P. Mantas, G. Kuntzelmann, N. Lageyre, B. Brannetti, B. Voedisch, S. Geisse, T. Jostock, H. Laux, *Biotechnol. Bioeng.* **2016**, *113*, 2433.

[4] G. Byrne, S.M. O'Rourke, D.L. Alexander, B. Yu, R.C. Doran, M. Wright, Q. Chen, P. Azadi, P.W. Berman, *PLoS Biol.* **2018**, *16*, 1.

[5] T. Amann, A.H. Hansen, S. Kol, G.M. Lee, M.R. Andersen, H.F. Kildegaard, *Biotechnol. J.* **2018**, *13*, e1800111.

[6] J. Chiu, K.N. Valente, N.E. Levy, L. Min, A.M. Lenhoff, K.H. Lee, *Biotechnol. Bioeng.* **2017**, *114*, 1006.

[7] N.K. Hamaker, K.H. Lee, *Curr. Opin. Chem. Eng.* **2018**, *22*, 152.

[8] J.K. Cheng, A.M. Lewis, D.S. Kim, T. Dyess, H.S. Alper, *Biotechnol. J.* **2016**, *11*, 1100.

[9] L. Zhang, M.C. Inniss, S. Han, M. Moffat, H. Jones, B. Zhang, W.L. Cox, J.R. Rance, R.J. Young, *Biotechnol. Prog.* **2015**, *31*, 1645.

[10] S.A. O'Brien, K. Lee, H.-Y. Fu, Z. Lee, T.S. Le, C.S. Stach, M.G. McCann, A.Q. Zhang, M.J. Smanski, N. V. Somia, W.-S. Hu, *Biotechnol. J.* **2018**, *13*, e1800226.

[11] L. Gaidukov, L. Wroblewska, B. Teague, T. Nelson, X. Zhang, Y. Liu, K. Jagtap, S. Mamo, W.A. Tseng, A. Lowe, J. Das, K. Bandara, S. Baijuraj, N.M. Summers, T.K. Lu, L. Zhang, R. Weiss, *Nucleic Acids Res.* **2018**, *46*, 4072.

[12] M.C. Inniss, K. Bandara, B. Jusiak, T.K. Lu, R. Weiss, L. Wroblewska, L. Zhang, *Biotechnol. Bioeng.* **2017**, *114*, 1837.

[13] S. Turan, C. Zehe, J. Kuehle, J. Qiao, J. Bode, *Gene* **2013**, *515*, 1.

[14] J.S. Lee, T.B. Kallehauge, L.E. Pedersen, H.F. Kildegaard, *Sci. Rep.* **2015**, *5*, 8572.

[15] T. Sakuma, S. Nakade, Y. Sakane, K.-I.T. Suzuki, T. Yamamoto, *Nat. Protoc.* **2015**, *11*, 118.

[16] S. Cristea, Y. Freyvert, Y. Santiago, M.C. Holmes, F.D. Urnov, P.D. Gregory, G.J. Cost, *Biotechnol. Bioeng.* **2013**, *110*, 871.

[17] K. Suzuki, Y. Tsunekawa, R. Hernandez-Benitez, J. Wu, J. Zhu, E.J. Kim, F. Hatanaka, M. Yamamoto, T. Araoka, Z. Li, M. Kurita, T. Hishida, M. Li, E. Aizawa, S. Guo, S. Chen, A. Goebl, R.D. Soligalla, J. Qu, T. Jiang, X. Fu, M. Jafari, C.R. Esteban, W.T. Berggren, J. Lajara, E. Nuñez-Delicado, P. Guillen, J.M. Campistol, F. Matsuzaki, G.-H. Liu, P. Magistretti, K. Zhang, E.M. Callaway, K. Zhang, J.C.I. Belmonte, *Nature* **2016**, *540*, 144.

[18] R. Bachu, I. Bergareche, L.A. Chasin, *Biotechnol. Bioeng.* **2015**, *112*, 2154.

[19] J.S. Lee, L.M. Grav, L.E. Pedersen, G.M. Lee, H.F. Kildegaard, *Biotechnol. Bioeng.* **2016**, *113*, 2518.

[20] S. Bosshard, P.-O. Duroy, N. Mermod, *DNA Repair (Amst).* **2019**, *82*, 102691.

[21] S.W. Shin, J.S. Lee, *Biotechnol. Bioeng.* **2020**, 1.

[22] V.T. Chu, T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, R. Kühn, *Nat. Biotechnol.* **2015**, *33*, 543.

[23] B. Gu, E. Posfai, J. Rossant, *Nat. Biotechnol.* **2018**, *36*, 632.

[24] R. Jayavaradhan, D.M. Pillis, M. Goodman, F. Zhang, Y. Zhang, P.R. Andreassen, P. Malik, *Nat. Commun.* **2019**, *10*, 1.

[25] H.A. Rees, W.H. Yeh, D.R. Liu, *Nat. Commun.* **2019**, *10*,.

[26] C.D. Yeh, C.D. Richardson, J.E. Corn, *Nat. Cell Biol.* **2019**, *21*, 1468.

[27] X. Li, X. Zhao, Y. Fang, T. Duong, C. Fan, C. Huang, S.R. Kain, X. Jiang, *J. Biol. Chem.* **1998**, *273*, 34970.

[28] J.G. Doench, E. Hartenian, D.B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B.L. Ebert, R.J. Xavier, D.E. Root, *Nat. Biotechnol.* **2014**, *32*, 1262.

[29] S. Hammond, M. Kaplarevic, N. Borth, M.J. Betenbaugh, K.H. Lee, *Biotechnol. Bioeng.* **2012**, *109*, 1353.

[30] B.G. Kremkow, J.Y. Baik, M.L. MacDonald, K.H. Lee, *Biotechnol. J.* **2015**, *10*, 931.

[31] P. Mali, L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville, G.M. Church, *Science* **2013**, *339*, 823.

[32] K.N. Valente, A.M. Lenhoff, K.H. Lee, *Biotechnol. Bioeng.* **2015**, *112*, 1232.

[33] A.J. Pierce, R.D. Johnson, L.H. Thompson, M. Jasin, *Genes Dev.* **1999**, *13*, 2633.

[34] W.Y. Mansour, S. Schumacher, R. Rosskopf, T. Rhein, F. Schmidt-Petersen, F. Gatzemeier, F. Haag, K. Borgmann, H. Willers, J. Dahm-Daphi, *Nucleic Acids Res.* **2008**, *36*, 4088.

[35] K. Kostyrko, N. Mermod, *Nucleic Acids Res.* **2015**, *44*,.

[36] D. Reyon, S.Q. Tsai, C. Khayter, J.A. Foden, J.D. Sander, J.K. Joung, *Nat. Biotechnol.* **2012**, *30*, 460.

[37] X. Yao, X. Wang, X. Hu, Z. Liu, J. Liu, H. Zhou, X. Shen, Y. Wei, Z. Huang, W. Ying, Y. Wang, Y.-H. Nie, C.-C. Zhang, S. Li, L. Cheng, Q. Wang, Y. Wu, P. Huang, Q. Sun, L. Shi, H. Yang, *Cell Res.* **2017**, *27*, 801.

[38] M.T. Certo, B.Y. Ryu, J.E. Annis, M. Garibov, J. Jarjour, D.J. Rawlings, A.M. Scharenberg, *Nat. Methods* **2011**, *8*, 671.

[39] R. Kuhar, K.S. Gwiazda, O. Humbert, T. Mandt, J. Pangallo, M. Brault, I. Khan, N. Maizels, D.J. Rawlings, A.M. Scharenberg, M.T. Certo, *Nucleic Acids Res.* **2016**, *42*, e4.

[40] T. Sakuma, M. Takenaga, Y. Kawabe, T. Nakamura, M. Kamihira, T. Yamamoto, *Int. J. Mol. Sci.* **2015**, *16*, 23849.

[41] B.P. Zambrowicz, A. Imamoto, S. Fiering, L.A. Herzenberg, W.G. Kerr, P. Soriano, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 3789.

[42] S. Irion, H. Luche, P. Gadue, H.J. Fehling, M. Kennedy, G. Keller, *Nat. Biotechnol.* **2007**, *25*, 1477.

[43] M. Kosicki, K. Tomberg, A. Bradley, *Nat. Biotechnol.* **2018**, *36*,.

[44] C. Ronda, L.E. Pedersen, H.G. Hansen, T.B. Kallehauge, M.J. Betenbaugh, A.T. Nielsen, H.F. Kildegaard, *Biotechnol. Bioeng.* **2014**, *111*, 1604.

[45] L.M. Grav, J.S. Lee, S. Gerling, T.B. Kallehauge, A.H. Hansen, S. Kol, G.M. Lee, L.E. Pedersen, H.F. Kildegaard, *Biotechnol. J.* **2015**, *10*, 1446.

[46] J. Shin, N. Lee, Y. Song, J. Park, T.J. Kang, S.C. Kim, G.M. Lee, B.K. Cho, *Biotechnol. Bioprocess Eng.* **2015**, *20*, 825.

[47] R.M. Daer, J.P. Cutts, D.A. Brafman, K.A. Haynes, *ACS Synth. Biol.* **2017**, *6*, 428.

[48] R. Scully, A. Panday, R. Elango, N.A. Willis, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 698.

[49] O. Baker, S. Tsurkan, J. Fu, B. Klink, A. Rump, M. Obst, A. Kranz, E. Schröck, K. Anastassiadis, A. Francis Stewart, *Nucleic Acids Res.* **2017**, *45*, 8105.

[50] J.-P. Zhang, X.-L. Li, G.-H. Li, W. Chen, C. Arakaki, G.D. Botimer, D. Baylink, L. Zhang, W. Wen, Y.-W. Fu, J. Xu, N. Chun, W. Yuan, T. Cheng, X.-B. Zhang, *Genome Biol.* **2017**, *18*, 35.

[51] M. Zhao, J. Wang, M. Luo, H. Luo, M. Zhao, L. Han, M. Zhang, H. Yang, Y. Xie, H. Jiang, L. Feng, H. Lu, J. Zhu, *Appl. Microbiol. Biotechnol.* **2018**, *102*, 6105.

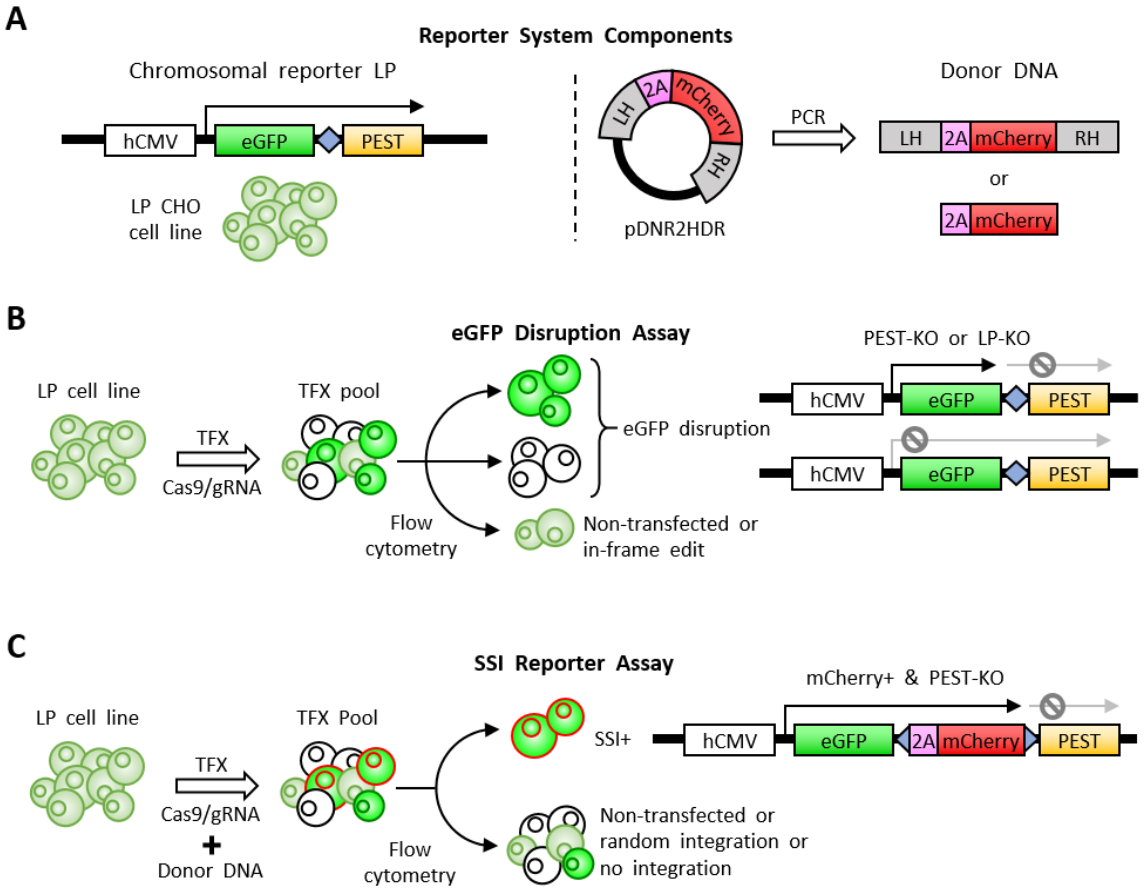


Figure 1. The SSIGNAL (site-specific integration and genome alteration) reporter system. A) Components of the reporter system. The landing pad (LP) CHO cell line contains the chromosomal reporter element, which is an expression cassette for destabilized eGFP. *PEST* encodes a degron tag that is fused to *eGFP* by a linker that harbors an artificial CRISPR/Cas9 target (blue diamond). Black arrow, expressed sequence. Donor DNA, carrying a promoterless *2A-mCherry* gene flanked by homology arms (LH/RH) with user-defined lengths, is generated via PCR from a plasmid template (pDNR2HDR). B) Schematic of eGFP disruption assay and readout. Cells with PEST-knockout (KO) or LP-KO phenotypes indicate mutagenic repair of targeted double-strand breaks, providing a measure of CRISPR/Cas9 activity. Grayed out arrow, sequence that is not expressed due to KO. C) Schematic of site-specific integration (SSI) reporter assay and readout. Cells with both PEST-KO and mCherry-positive phenotypes indicate an SSI event.

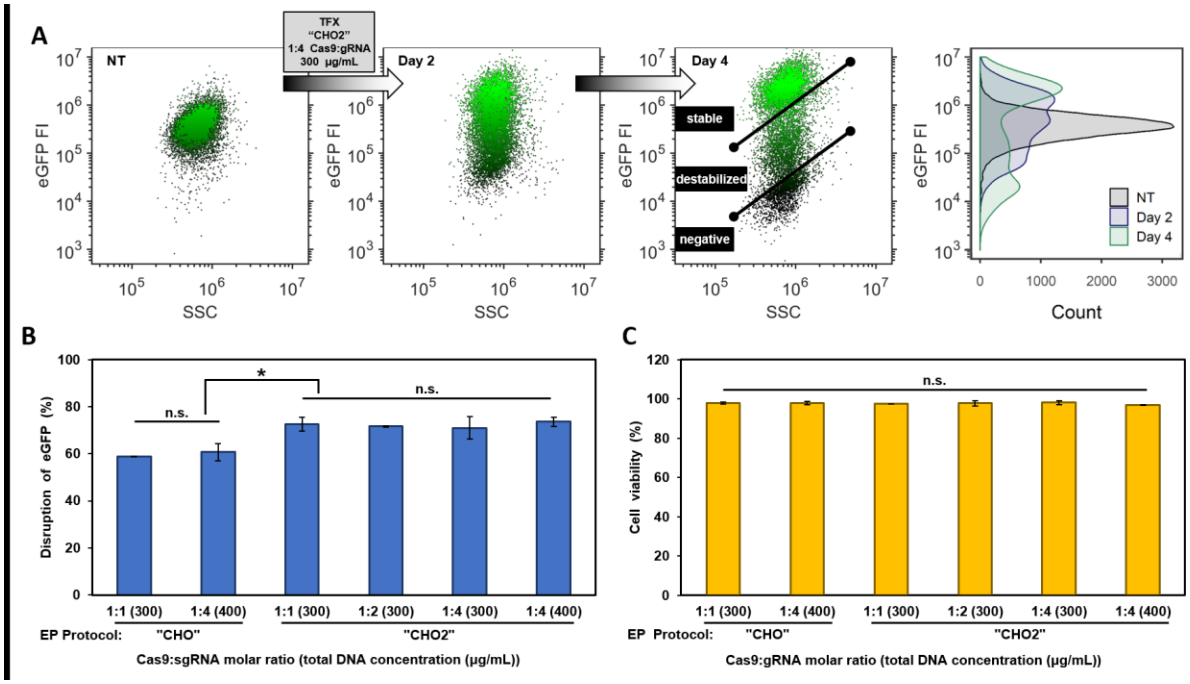


Figure 2. Measurement of CRISPR/Cas9-mediated genome editing using eGFP disruption assay. A) Representative flow cytometry scatter plots of CHO-LP1R26A4 cells before and after transfection with Cas9 and gRNA plasmids. Density graph provides overlay of eGFP intensities as a function of cell count shown in scatter plots. Mock gating scheme used to determine absolute disruption efficiency is presented, where eGFP-stable and eGFP-negative fractions are summed to calculate efficiency (refer to Supporting Figure 3 for extended gating details and example data). Transfection (TFX) conditions noted in gray box. NT, non-transfected. FI, relative fluorescence intensity. SSC, side scatter. FI and SSC reported in arbitrary units. Plots represent at least 10,000 live cells. B) Disruption efficiencies as determined by flow cytometric analysis (raw data provided in Supporting Figure 3C) and (C) culture viabilities for specified transfection conditions. Measurements taken 4 days post-transfection. Bar graphs represent mean of two independent experiments. *P<0.01 by one-way ANOVA (n.s., not significant). Error bars, 95% c.i.

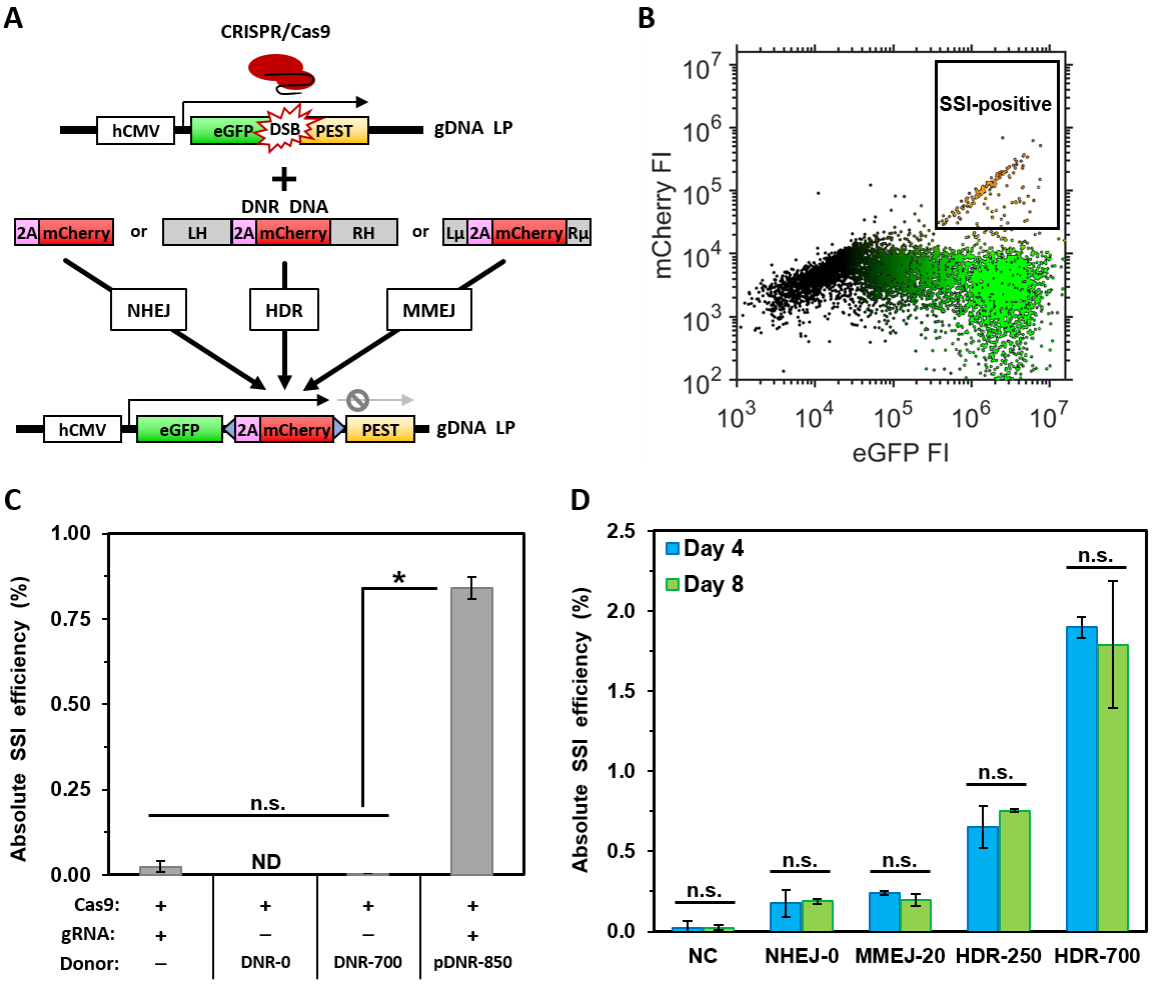


Figure 3. Evaluation of targeted integration methods using the SSIGNAL reporter system. A) Schematic of genotype produced from SSI event. Following CRISPR/Cas9-mediated induction of a targeted double-strand break (DSB) within the landing pad (LP), donor (DNR) DNA can be integrated precisely by using one of the native DSB repair pathways depending on the length of homology arms used. The non-homologous end joining (NHEJ) pathway is harnessed by using donor DNA that lacks homology with the target site. Homology-directed repair (HDR) and microhomology-mediated end joining (MMEJ) are stimulated by the presence of relatively long (LH/RH) and short (Lμ/Rμ) homology arms, respectively. Black arrow, expressed sequence; grayed out arrow, sequence that is knocked out of frame. B) Representative flow cytometry

scatter plot of CHO-LP1R26A4 cells after co-transfection with Cas9, gRNA, and donor plasmids. Mock gating represents strategy to identify SSI-positive based on both mCherry and eGFP fluorescence intensity (FI) criteria (refer to Supporting Figure 4 for extended gating details and example data). FI reported in arbitrary units. Plot represents at least 10,000 live cells. C) Control SSI conditions with or without specified transfection components. Efficiencies shown are determined from measurements taken 8 days post-transfection. Donor DNA was introduced in linear or plasmid (pDNR) form. Donor numbers correspond to the length of their homology arms. Bar graphs represent mean of two independent experiments. *P<0.01 by one-way ANOVA (n.s., not significant). Error bars, 95% c.i. D) SSI efficiencies of different repair strategies with linear donors. Pathway usage is based on the length of homology arms on the donor DNA as indicated by the donor number (NHEJ, 0 bp; MMEJ, 20 bp; HDR, 250 bp or 700 bp). NC, no donor gating control. ND, not detected. Bar graphs represent mean of two independent experiments with measurements taken at day 4 and 8 post-transfection compared using Student's t-test (extended statistical analysis provided in Supporting Figure 5A). Error bars, 95% c.i.