



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Comparison of surrogate reporter systems for enrichment of cells with mutations induced by genome editors

Citation for published version:

He, Z, Shi, X, Liu, M, Sun, G, Proudfoot, C, Whitelaw, C, Lillico, S & Chen, Y 2016, 'Comparison of surrogate reporter systems for enrichment of cells with mutations induced by genome editors' *Journal of Biotechnology*, vol. 221, pp. 49-54. DOI: 10.1016/j.jbiotec.2016.01.009

Digital Object Identifier (DOI):

[10.1016/j.jbiotec.2016.01.009](https://doi.org/10.1016/j.jbiotec.2016.01.009)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Journal of Biotechnology

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Manuscript Number: JBIOTEC-D-15-00897R2

Title: Comparison of surrogate reporter systems for enrichment of cells with mutations induced by genome editors

Article Type: Short Communication

Section/Category: Nucleic Acids / Molecular Biology

Keywords: ZFN; CRISPR/Cas9; Surrogate reporter; Siglec-1

Corresponding Author: Dr. Zuyong He, Ph.D.

Corresponding Author's Institution: Sun Yat-Sen University

First Author: Zuyong He, Ph.D.

Order of Authors: Zuyong He, Ph.D.; Xuan Shi; Meirui Liu; Guangjie Sun; Chris Proudfoot, Dr.; Bruce Whitelaw; Simon Lillico; Yaosheng Chen, Dr.

Abstract: Genome editors are powerful tools that allow modification of the nuclear DNA in eukaryotic cells both in vitro and in vivo. In vitro modified cells are often phenotypically indistinguishable from unmodified cells, hampering their isolation for analysis. Episomal reporters encoding fluorescent proteins can be used for enrichment of modified cells by flow cytometry. Here we compare two surrogate reporters, RGS and SSA, for the enrichment of porcine embryonic fibroblasts containing mutations induced by ZFNs or CRISPR/Cas9. Both systems were effective for enrichment of edited porcine cells with the RGS reporter proving more effective than the SSA reporter. We noted a higher-fold enrichment when editing events were induced by Cas9 compared to those induced by ZFNs, allowing selection at frequencies as high as 70%.

**Comparison of surrogate reporter systems for enrichment of cells with mutations
induced by genome editors**

Zuyong He^{1§}, Xuan Shi^{1§}, Meirui Liu¹, Guangjie Sun¹, Chris Proudfoot², C. Bruce A. Whitelaw², Simon G. Lillico^{2,*} and Yaosheng Chen^{1,*}

1. State Key Laboratory of Biocontrol, School of Life Sciences, SunYat-senUniversity, Guangzhou 510006, P. R. China.
2. The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Edinburgh, EH25 9RG, UK.

[§]These authors contribute equally to this work

* For correspondence: Yaosheng Chen [chyaosh@mail.sysu.edu.cn];

Simon Lillico [simon.lillico@roslin.ed.ac.uk]

Abstract: Genome editors are powerful tools that allow modification of the nuclear DNA in eukaryotic cells both *in vitro* and *in vivo*. *In vitro* modified cells are often phenotypically indistinguishable from unmodified cells, hampering their isolation for analysis. Episomal reporters encoding fluorescent proteins can be used for enrichment of modified cells by flow cytometry. Here we compare two surrogate reporters, RGS and SSA, for the enrichment of porcine embryonic fibroblasts containing mutations induced by ZFNs or CRISPR/Cas9. Both systems were effective for enrichment of edited porcine cells with the RGS reporter proving more effective than the SSA reporter. We noted a higher-fold enrichment when editing events were induced by Cas9 compared to those induced by ZFNs, allowing selection at frequencies as high as 70%.

Keywords: ZFN; CRISPR/Cas9; Surrogate reporter; *SIGLEC-1*

Programmable nuclease-based technologies such as zinc-finger nucleases (ZFNs) (Urnov et al., 2010) and the bacterial clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) system (Cong et al., 2013) are powerful and versatile tools for genome editing. They can both induce site-specific DNA double-strand breaks (DSBs) in the genome of eukaryotic cells by utilising either the non-specific nuclease *FokI* or Cas9 respectively. While ZFNs direct localisation of the *FokI* dimer to the target sequence by specific DNA binding by zinc-finger domains, the CRISPR/Cas9 system relies on Watson/Crick base pairing between a small guide RNA and a 20-nucleotide target site flanked by a protospacer-adjacent motif (PAM) to direct the Cas9 nuclease (Sternberg et al., 2014). The error-prone NHEJ pathway repairs many nuclease-mediated DSBs, resulting in the inclusion of small insertions or deletions (indels) at the repair site. These may result in frame shift if the target site is situated in protein coding sequence, and thereby functional gene disruption. Although application of designer nucleases have enabled targeted genome editing in numerous eukaryotic species and cell types (Hai et al., 2014; He et al., 2014; Ma et al., 2013), there is often difficulty in the isolation of mutant cells post-modification due to grossly similar phenotypes between modified and wild-type counterparts, hampering their application to a certain extent. Recently, two different episomal surrogate reporter systems have been successfully used to enrich cells containing mutations induced by programmable nucleases. One surrogate reporter, called RGS (Kim et al., 2011), encodes a monomeric mRFP-eGFP fusion protein with a cloning site for introduction of the proposed target site located between the two reading frames. In its native conformation the mRFP is expressed from the constitutive promoter while the eGFP sits out of frame downstream (Fig. 1B). Consequently, upon transfection cells express mRFP but not eGFP. DSB creation by engineered nucleases at the introduced target site

between the two FPs results in NHEJ and frame shift mutations within the construct, resulting in a proportion of cells expressing both mRFP and eGFP (Fig.1C) which can be detected and/or selected by FACS. The second reporter system also utilises a fluorescent switch. Named SSA (for single-strand annealing) this construct contains a non-functional GFP with a duplicated central domain containing the nuclease target site (Doyon et al., 2013). In this instance generation of a DSB can restore GFP function when repair is via the SSA pathway (Fig. 1E).

To our knowledge, there is no report of a direct comparison between these two reporters for the enrichment of genome-modified primary cells. Competition between different DNA repair pathways has been documented, and which pathway predominates may be influenced by cell type and the nature of the broken ends (*e.g.*, blunt end vs. overhangs) (Preston et al., 2002). As a designer-nuclease target gene we chose *SIGLEC-1*, as the encoded protein, sialoadhesin, is expressed in a tissue-specific manner on cell surface of macrophages and is a known receptor for porcine reproductive and respiratory syndrome virus (PRRSV) (Van Breedam et al., 2013). We utilized either ZFN or CRISPR/Cas9 targeting *SIGLEC-1* to compare the RGS and SSA reporters for enrichment of edited porcine embryonic fibroblasts (PEFs). The PEFs used throughout this study were primarily cultured from fetuses of Large White pigs at 30 days after insemination.

We designed a ZFN pair (ordered from Sigma-Aldrich) and one gRNA to target the same sequence within exon 6 of porcine *SIGLEC-1* (Fig. 1A & Supplementary table 1). We co-transfected the plasmids px330 (Addgene, #42230; 3 µg) encoding cas9 and sgRNA together with the RGS reporter (2.5 µg) containing the sgRNA/ZFN target sequence (Fig.1 B & Supplementary table 2) into 1×10^6 PEF cells through electroporation at 1400 V, 40 ms, 1 pulse using the Neon transfection system (Invitrogen). Transfected cells were subjected to flow

cytometry after 2 days in culture (Fig.1 C). Negative control cells transfected with the RGS reporter in the absence of px330 expressed mRFP (49%) but low levels of eGFP (3.3%) while co-transfection of both plasmids resulted in 26.5% red cells of which 15.3% were green. All green cells expressed mRFP (Figs. 2A & B). Next, we isolated the brightest 3.6% eGFP positive cells from the co-transfected population in order to analyse the *SIGLEC-1* sequence at the genomic locus. We then repeated the above experiment but with the SSA construct (3 µg for transfection) in place of the RGS reporter. Flow cytometric analysis showed that 15.8% cells were eGFP positive, compared to only 4.5% cells when transfected with the SSA reporter alone (Figs. 2C & D). FACS-sorted eGFP cells from each experiment were analysed for targeted mutation frequencies at the *SIGLEC-1* genomic target site by T7 endonuclease I assay (T7E1) (Supplementary table 3), and compared with their unsorted counterparts. The RGS reporter system enabled a 2.8-fold enrichment in targeted mutation (36% vs. 13%) when compared to the unsorted pool (Fig. 2E). The SSA reporter system was comparable, with a 2-fold enrichment (38% vs. 19%).

The T7E1 assay tends to underestimate fold-enrichment at high mutation frequencies as mutant sequences can form homoduplexes, which are insensitive to T7E1 digestion (Kim et al., 2011). To assess this, PCR amplicons containing the nuclease target sequences were cloned into the pMD18-T simple vector for Sanger sequencing to quantify the NHEJ events. Mutation frequency in the mRFP/eGFP positive fraction and in unsorted cells was 71.4% (35/49) and 9.26% (5/54) respectively, indicating a nearly 8-fold enrichment of mutations by using RGS reporter (Fig. 2F). In contrast, only 3-fold (52% (26/50) vs. 17.3% (9/52)) enrichment of mutations was found by using SSA reporter (Fig. 2G). Both the T7E1 assay and sequencing results demonstrate that both surrogate reporter systems are reliable for the efficient enrichment of gene-disrupted cells,

with the RGS reporter proving more effective than the SSA reporter. By using the px458 vector (Addgene, #48138), which has a 2A-EGFP following Cas9, as a control (Fig.S1A) we demonstrated that the transfection efficiency of the RGS reporter (5.37kb) and SSA reporter (5.15kb) are comparable (Figs. S1B & D), indicating that transfection efficiency is unlikely to be a contributory factor influencing the difference in enrichment efficiency between these two reporter systems. A repair hierarchy has been shown to exist in mammalian cells in which NHEJ is promoted to limit the access of the more mutagenic SSA mechanism to the break site (Mansour et al., 2008). Therefore, we speculated that after nuclease cleavage the SSA pathway mediated DNA repair may be either less frequent, or more time consuming than the NHEJ. Thus, the SSA reporter maybe a less direct indicator than the RGS reporter of a concurrent break repair at the genomic target site.

We next compared enrichment of cells cotransfected with a ZFN pair (1.5 µg DNA of each ZFN plasmid) in place of the CRISPR/Cas9 plasmid px330 (Figs.1 B & D, and supplementary table 2) and either the RGS or SSA reporters. Cotransfection of ZFN/RGS plasmids produced considerably more eGFP positive cells than RGS plasmid alone (4.8% vs 0.7%, Figs. 3A & C); eGFP positive cells were selected by FACS for subsequent mutation characterization (Fig. 3B). By comparison, cotransfection of ZFN/SSA plasmids resulted in 9.46% cells that were eGFP positive as opposed to 4.57% cells transfected with the SSA reporter alone; eGFP positive cells were selected by FACS for subsequent mutation characterization (Fig. 3D).

To evaluate the mutation enrichment efficiency of RGS and SSA reporter, the T7E1 assay was performed on isolated gDNA, demonstrating a mutation frequency of 17% and 18% respectively. This was 1.7- and 2-fold higher than in unsorted cells (10% and 9%, respectively; Fig. 3E). The

enrichment efficiencies were confirmed by sequencing the target region, which revealed the mutation frequency after RGS or SSA reporter selection was 1.98-fold (21.95% (9/41) vs. 11.11% (6/54)) or 1.84-fold (17.31% (9/52) vs. 9.4% (5/53)) higher in sorted cells respectively (Figs. 3F & G). Surrogate reporters have previously been used for enriching cells with mutations induced by programmable nucleases in tumor cell lines (Cradick et al., 2014; Kim et al., 2011; Ramakrishna et al., 2014). However, the targeted tumor cells identified by this enrichment cannot be used as donors for either biomedical applications or the production of genome edited animals. Although genome editing using engineered nucleases, especially the CRISPR-Cas9 system, is highly efficient in a broad range of immortal cell lines, genome editing in primary mammalian cells is generally more challenging. The exact reasons for this remain elusive, but differences in transfection rates, promoter activity, exonuclease activity, and DNA repair fidelity may contribute. In this study, both ZFN and CRISPR/Cas9 induced targeted mutations in primary PEFs at a frequencies of around 10%, which is lower than previously reported in immortal cell lines (Segal and Meckler, 2013). Therefore, using a reliable selection strategy to improve the targeting efficiency in PEF is of importance. To our knowledge, this is the first report comparing the RGS and SSA surrogate reporter systems for enrichment of gene targeted porcine primary cells. The RGS reporter slightly outperformed the SSA reporter when used in conjunction with both CRISPR/Cas9 or a pair of ZFNs to enrich gene-disrupted PEFs at the identical *SIGLEC-1* locus, making it an attractive tool to improve the efficiency of generating genome edited pigs.

Acknowledgements

This work was supported by the joint funds of NSFC-Guangdong (U1201213); National Transgenic Breeding Program (2014ZX08006005-005) and Chinese Scholarship Scheme [ZH: grant number 201208440261].

References

- Cong, L., Ran, F.A., Cox, D., Lin, S.L., Barretto, R., Habib, N., Hsu, P.D., Wu, X.B., Jiang, W.Y., Marraffini, L.A., Zhang, F., (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 819-823.
- Cradick, T.J., Antico, C.J., Bao, G., (2014) High-throughput cellular screening of engineered nuclease activity using the single-strand annealing assay and luciferase reporter. *Methods Mol Biol* 1114, 339-352.
- Doyon, Y., Holmes, M.C., Li, T., (2013) Rapid screening of biologically active nucleases and isolation of nuclease-modified cells. Patent Application Publication. Sangamo BioSciences, Inc., Richmond, CA (US), United States, p. 11.
- Hai, T., Teng, F., Guo, R., Li, W., Zhou, Q., (2014) One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell Res* 24, 372-375.
- He, Z., Proudfoot, C., Mileham, A.J., McLaren, D.G., Whitelaw, C.B., Lillico, S.G., (2014) Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol Bioeng*.
- Kim, H., Um, E., Cho, S.R., Jung, C., Kim, J.S., (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods* 8, 941-943.
- Ma, Y., Zhang, X., Shen, B., Lu, Y., Chen, W., Ma, J., Bai, L., Huang, X., Zhang, L., (2013) Generating rats with conditional alleles using CRISPR/Cas9. *Cell Res*.
- Mansour, W.Y., Schumacher, S., Roskopf, R., Rhein, T., Schmidt-Petersen, F., Gatzemeier, F., Haag, F., Borgmann, K., Willers, H., Dahm-Daphi, J., (2008) Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. *Nucleic Acids Res* 36, 4088-4098.
- Preston, C.R., Engels, W., Flores, C., (2002) Efficient repair of DNA breaks in *Drosophila*: evidence for single-strand annealing and competition with other repair pathways. *Genetics* 161, 711-720.
- Ramakrishna, S., Cho, S.W., Kim, S., Song, M., Gopalappa, R., Kim, J.S., Kim, H., (2014) Surrogate reporter-based enrichment of cells containing RNA-guided Cas9 nuclease-induced mutations. *Nat Commun* 5, 3378.
- Segal, D.J., Meckler, J.F., (2013) Genome engineering at the dawn of the golden age. *Annu Rev Genomics Hum Genet* 14, 135-158.
- Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A., (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62-67.
- Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., Gregory, P.D., (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11, 636-646.
- Van Breedam, W., Verbeeck, M., Christiaens, I., Van Gorp, H., Nauwynck, H.J., (2013) Porcine, murine and human sialoadhesin (Sn/Siglec-1/CD169): portals for porcine reproductive and respiratory

syndrome virus entry into target cells. J Gen Virol 94, 1955-1960.

Figure Legends

Figure 1. ZFN and gRNA design, and the principle of the use of surrogate reporters for enrichment of nuclease-modified cells. (A) The schematic diagram of target sites of ZFN and gRNA on the exon 6 of porcine *SIGLEC-1* gene. Coloured arrow lines indicate the sequence used for the guide segment of gRNA. The NGG nucleotide protospacer adjacent motif (PAM) sequences in red. Colour boxed sequences denote the DNA binding regions of the ZFN proteins. (B) The working mechanism of the RGS reporter. mRFP is constitutively expressed by the CMV promoter, whereas eGFP is not expressed without ZFN or CRISPR/Cas9 activity because the eGFP sequence is out of frame and there is a stop codon before eGFP. If a double-strand break (DSB) is introduced into the target sequence by the programmable nucleases, the break is repaired by error-prone nonhomologous end joining (NHEJ), which often results in indels. This indel formation can cause frame shifts, making either of the eGFP genes in frame and expressed. (C) A schematic illustrating the enrichment of nuclease-induced mutations in mRFP⁺eGFP⁺ cell population sorted by flow cytometry. Reporter plasmids and chromosomal target loci are illustrated. Mutations are shown as yellow spots. (D) The working mechanism of the SSA reporter. The SSA reporter consists of a sequence encoding GFP, and ZFN or CRISPR/Cas9 target site which disrupts the expression of GFP. The unique 5' GFP sequence, middle repeated GFP sequence, and unique 3' GFP sequence are designated as G, F, and P respectively. Following cleavage with the programmable nucleases and SSA-mediated repair, the functional GFP open reading frame is reconstituted by loss of sequences between the two identical 5' and 3' F sequences.

(E) Schematic illustrates enrichment of nuclease-induced mutations in eGFP⁺ cells sorted by flow cytometry. Reporter plasmids and chromosomal target loci are illustrated. Mutations are shown as yellow spots.

Figure 2. Comparison of the enrichment of *SIGLEC-1*-disrupted PEF cells induced by CRISPR/Cas9 between RGS reporter and SSA reporter. (A) The fluorescent microscopy of PEF 1d after cotransfection of the RGS reporter and the px330 plasmid encoding Cas9 and the gRNA targets *SIGLEC-1* gene; this is representative of three independent experiments. (B) Flow cytometry of PEF cells 2 d after cotransfection of RGS reporter and the px330 plasmid encoding Cas9 and the gRNA targets *SIGLEC-1* gene; this is representative of three independent experiments. Percentage of cells that express both mRFP and eGFP is indicated, and the percentage of the sorted population with strongest eGFP signal also is denoted. (C) The fluorescent microscopy of PEF 1d after cotransfection of the SSA reporter and the px330 plasmid encoding Cas9 and the gRNA targets *SIGLEC-1* gene; this is representative of three independent experiments. (D) Flow cytometry of PEF cells 2d after cotransfection of SSA reporter and the px330 plasmid encoding Cas9 and the gRNA targets *SIGLEC-1* gene; this is representative of three independent experiments. Percentage of cells that express eGFP is indicated, and the percentage of the sorted population with strongest eGFP signal also is denoted. (E) The frequency of CRISPR/Cas9-induced mutations as determined by the T7E1 assay. Genomic DNA of samples from three independent experiments was mixed for assay. Arrows indicate the expected positions of DNA bands cleaved by mismatch-sensitive T7E1. The numbers at the bottom of the gel indicate mutation percentages estimated by band

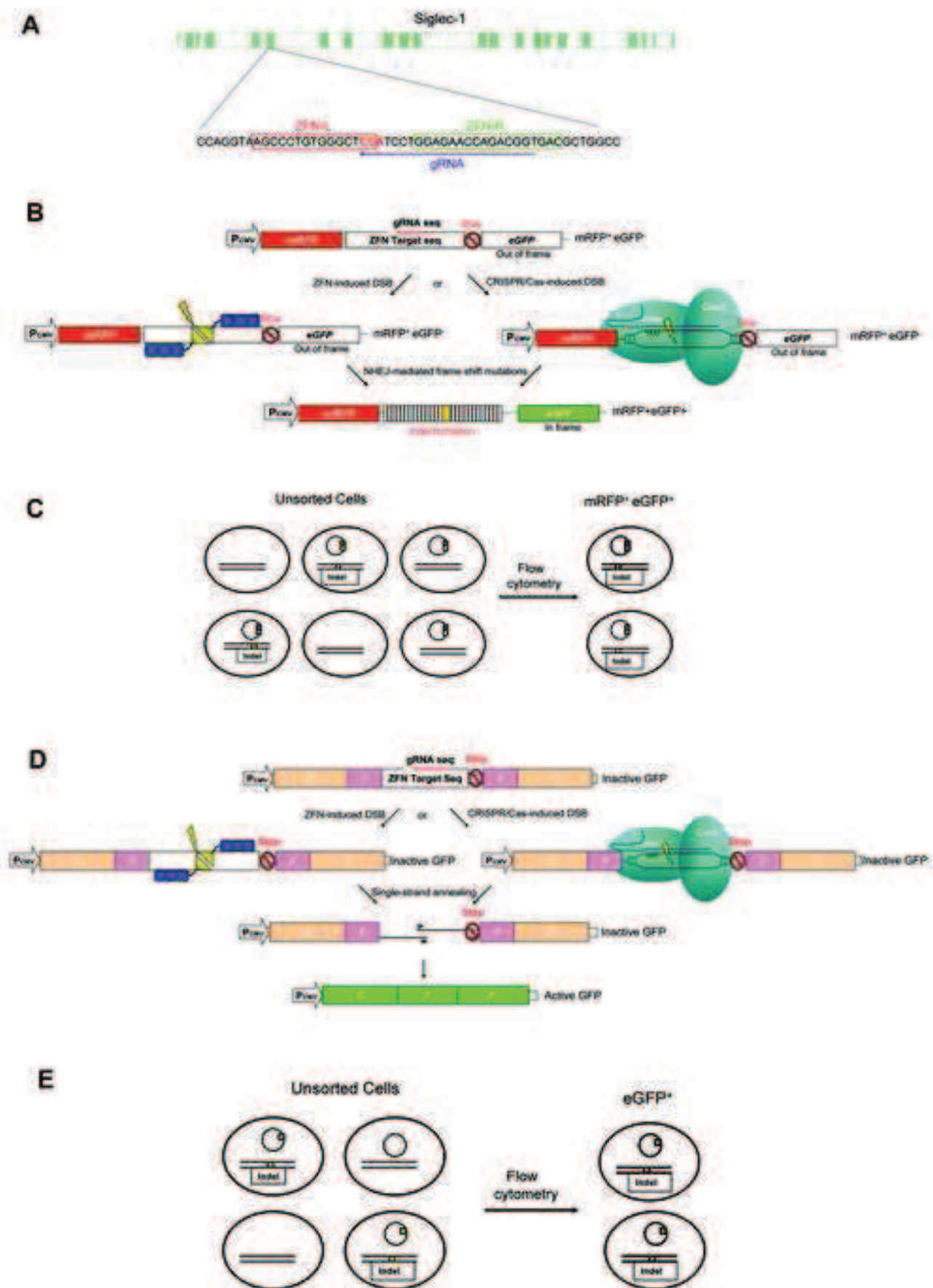
intensities with ImageJ. (F, G) DNA sequences of the wild-type (WT) and mutant clones, with CRISPR/Cas9 recognition sites shown in red and the protospacer adjacent motif (PAM) sequence in bold characters. Genomic DNA of samples from three independent experiments was mixed for sequencing. Dashes and lower case letters indicate deleted and inserted bases, respectively (the number of inserted or deleted bases are described in the parentheses). The number of occurrences is shown in parentheses; ×1 and ×2 indicate the number of each clone. Mutation frequencies were obtained by dividing the number of mutant clones by the number of total clones.

Figure 3. Comparison of the enrichment of *SIGLEC-1*-disrupted PEF cells induced by ZFN between RGS reporter and SSA reporter. (A) The fluorescent microscopy of PEF 1d after cotransfection of the RGS reporter and the plasmids encoding ZFN pair; this is representative of three independent experiments. (B) Flow cytometry of PEF cells 2d after cotransfection of RGS reporter and the plasmids encoding ZFN pair; this is representative of three independent experiments. Percentage of cells that express both mRFP and eGFP is indicated, and the percentage of the sorted population with strongest eGFP signal also is presented. (C) The fluorescent microscopy of PEF 1d after cotransfection of the SSA reporter and the plasmids encoding ZFN pair; this is representative of three independent experiments. (D) Flow cytometry of PEF cells 2d after cotransfection of SSA reporter and the plasmids encoding ZFN pair; this is representative of three independent experiments. Percentage of cells that express eGFP is indicated, and the percentage of the sorted population with strongest eGFP signal also is denoted. (E) The frequency of ZFN-induced mutations as determined by the

T7E1 assay is shown. Genomic DNA of samples from three independent experiments was mixed for assay. Arrows indicate the expected positions of DNA bands cleaved by mismatch-sensitive T7E1. The numbers at the bottom of the gel indicate mutation percentages calculated by band intensities with ImageJ. (F, G) DNA sequences of the wild-type (WT) and mutant clones, with ZFN recognition sites shown in red and the spacer sequence in lower case. Genomic DNA of samples from three independent experiments was mixed for sequencing. Dashes and lower case letters indicate deleted and inserted bases, respectively (the number of inserted or deleted bases are described in the parentheses). The number of occurrences is shown in parentheses; ×1 and ×2 indicate the number of each clone. Mutation frequencies were obtained by dividing the number of mutant clones by the number of total clones.

Figure S1. Comparison of transfection efficiency between CRISPR/Cas9 cotransfected with RGS reporter and SSA reporter. (A) The schematic diagram of Cas9 expression px458 which has an 2A-EGFP tag in comparison to px330. (B) FACS analysis of transfection efficiency of PEF cells 2d after cotransfection of px458 plasmid and RGS reporter by measuring the percentage of GFP positive cells; this is representative of three independent experiments. (C) FACS analysis of transfection efficiency of PEF cells 2d after cotransfection of px458 plasmid and RGS reporter by measuring the percentage of RFP positive cells. (D) FACS analysis of transfection efficiency of PEF cells 2d after cotransfection of px458 plasmid and SSA reporter by measuring the percentage of GFP positive cells; this is representative of three independent experiments.

Figure1
[Click here to download high resolution image](#)



Supplementary table 1

[Click here to download Supplementary File: Supplementary Table 1.doc](#)

Supplementary table 2

[Click here to download Supplementary File: Supplementary Table 2.doc](#)

Supplementary table 3

[Click here to download Supplementary File: Supplementary Table 3.doc](#)

Supplementary figure 1

[Click here to download Supplementary File: Fig. S1.jpg](#)