Abstract

Efficient enrichment of nuclease-induced mutant cells by using surrogate reporters

Hyojin Kim

Department of Chemistry

The Graduate School

Seoul National University

Programmable nucleases such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are powerful tools for genome engineering. They can be used to induce double-strand breaks (DSBs) at specific target sites in the genome, which are repaired by intrinsic cellular mechanisms such as

non-homologous end-joining (NHEJ) or homologous recombination (HR). Although this technology has been successfully used to achieve targeted genome engineering in many animals, plants, and stem cells, a robust method to isolate cells carrying genetic modifications induced by programmable nucleases from morphologically indistinguishable unmodified cells would enable broader applications of this technology in biomedical and bio-industrial research. Here, I describe efficient methods for enrichment of endogenous gene-modified cells by using the surrogate reporter system. Surrogate reporters are designed to express marker proteins following the introduction of frameshift mutations by programmable nucleases. When programmable nucleases and the surrogate reporter were introduced into cells, the population of cells that expressed the marker gene from surrogate reporters showed a high rate of mutation on the endogenous nuclease target site. Using three different marker genes, I developed three types of surrogate reporters that enable different selection methods such as flow cytometry-based cell sorting, antibody-assisted separation, and drug selection, each with advantages in different cell types and suited for differed research purposes. The three reporter systems can support the enrichment of mutant cell populations with 6- to 92-fold higher mutation rates than unsorted cells. In conclusion, the surrogate reporter system described in this dissertation provides a simple and quick method to separate nuclease-induced mutant cells, thereby enabling efficient application of the programmable nuclease technology in a broad spectrum of biomedical and bio-industrial research.

Keywords: Zinc-finger nuclease (ZFN), TAL-effector nuclease (TALEN), DNA double-strand breaks (DSB), Non-homologous end-joining (NHEJ), Surrogate reporter, Gene mutagenesis

Student Number: 2007-20304

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CNV	Copy number variation
DSB	Double-strand break
FACS	Fluorescence-activated cell sorting
FAM	Carboxyl-fluo-rescine
HCV	Hepatitis C virus
HEK	Human embryonic kidney
HIV	Human immunodeficiency virus
HR	Homologous recombination
KO	Knockout
NHEJ	Non-homologous end-joining
PCR	Polymerase chain reaction
SSA	Single-strand annealing
SSB	Single-strand break
T7E1	T7 Endonuclease 1
ZF	Zinc-finger
ZFN	Zinc-finger nuclease
TALEN	Transcription activator-like effector nuclease

I. Introduction

Programmable nucleases such as zinc-finger nucleases (ZFNs) and TAL-effector nucleases (TALENs) (Miller et al., 2010) are powerful and versatile tools for genome engineering (Urnov et al., 2005; Beumer et al., 2006; Doyon et al., 2008; Cermak et al., 2011). They allow endogenous gene disruption (Perez et al., 2008), targeted gene addition (Moehle et al., 2007), and chromosomal rearrangements (Lee et al., 2010) in cells and organisms. With the advent of the Human Genome Project, genetic information at the nucleotide sequence level can be used in permanent therapy for genetic diseases caused by specific mutations in specific genes.

Gene targeting is a process that manipulates the genome by homologous recombination (HR)-mediated exchange between sequences at targeted sites in the genome and those in the donor DNA template. However, the limitation of gene targeting is its low efficiency (typically lower than 10^{-6} (Capecchi et al., 1989). The efficiency of HR events is not sufficient to modify plant and mammalian cell genomes a number of methods have been investigated to increase the HR efficiency. When a donor DNA that is homologous to the target on both sides of the double-strand break (DSB) is provided, the genomic site can be repaired by homology-directed repair, allowing the incorporation of exogenous sequences between the homologous regions (Jasin et al., 1996; Urnov et al., 2005)

ZFNs and TALENs are artificial restriction enzymes that can be

designed to induce DSBs at specific target sites (Mani et al., 1999; Smith et al., 2005). Efficient gene targeting can be achieved by using these nucleases to increase HR efficiency (Bibikova et al., 2001; Porteus et al., 2003). Engineered nucleases induce site-specific DNA DSBs in the genome, and genome integrity is restored via endogenous DNA repair systems known as non-homologous end-joining (NHEJ) and HR, resulting in targeted mutagenesis and gene modifications. In the absence of homologous donor DNA, whose preparation is cumbersome and laborious, nuclease-induced DSBs are efficiently repaired by NHEJ, which is dominant over HR in higher eukaryotic cells and organisms (Lieber, 2010). Since NHEJ is intrinsically error-prone, small insertions and deletions (indels) are generated at the DSB site, often inducing frameshift mutations and gene disruptions. Thus, NHEJ can facilitate efficient and rapid establishment of knockout (KO) cells.

Gene KO is a technique in which a gene is made inoperative. KO organisms have been used to learn more about a gene that has been sequenced, but whose function is unknown or incompletely known. KOs are primarily used to understand the role of a specific gene or DNA region by comparing the KO organism with a wild type having a similar genetic background. KOs are also used as screening tools in drug development to target specific biological processes or deficiencies by using a specific KO or to understand the mechanism of action of a drug by using a library of KO organisms spanning the entire genome.

Recent work has shown that the programmable nuclease-based gene KO method is highly effective for creating animal models using not

only cell lines but also embryos. Such developments enable targeted genetic engineering of organisms such as rats (Geurts et al., 2009; Mashimo et al., 2010) and zebrafish (Doyon et al., 2008; Meng et al., 2008; Foley et al., 2009), where conventional methods of genetic manipulation are unsuccessful. Engineering-based gene targeting has been proven to work in a wide variety of organisms, including rats, mice, rabbits, zebrafish, *Drosophila*, and *Caenorhabditis elegans*, with testing in additional model systems underway0.0. Especially, the high efficiency of gene targeting by ZFNs or TALENs in human embryonic stem (ES) cells or induced pluripotent stem (iPS) cells raises the possibility of patient-specific gene therapy as a clinical application in regenerative medicine (Lombaro et al., 2007; Zou et al., 2009; Zou et al., 2011; Sebastiano et al., 2011; Yao et al., 2012; Lei et al., 2011; Hockemeyer et al., 2012; Yusa et al., 2012). Although, the generation of patient-specific iPS cells may be possible, the iPS cells still harbor the disease-causing mutations. Using designer nucleases, mutations can be corrected by specific gene targeting and a permanent therapeutic regimen (Schambach et al., 2010).

Furthermore, by introducing two ZFN pairs at the same time, a broad range of genome rearrangements are possible (Lee et al., 2010). According to a recent report on ZFN-induced chromosomal deletion, two concurrent DSBs can induce the deletion of a fragment between two sites. Using two ZFNs targeting different sites on the same chromosome (Chr. 3), They generated a targeted deletion of up to 15 Mbp at frequencies of 10^{-1} to 10^{-3} (Lee et al., 2010) In addition, the

use of two ZFNs targeting an endogenous locus on different chromosomes would lead to concurrent formation of DSBs and translocation by misjoining of the break points. The frequencies of such translocations were 10^{-4} to 10^{-5} in HEK293-derived cells and 10^{-5} to 10^{-6} in human ES cells (Brunet et al., 2009). These findings indicate that copy number variations (CNVs) can occur when two or more DSBs occur in the genome at the same time. Thus, using the engineered-nuclease technology, researchers may be able to induce specific genome rearrangements, such as large deletions, insertions, replacements, or translocations, at their favorite loci.

Millions of SNPs have been reported, and some of these are known to be associated with diseases (International HapMap Consortium, 2003). Genetic disorders caused by genome rearrangements were revealed relatively recently. CNVs caused by deletions, insertions, inversions, and duplications can also appear in benign polymorphic variants. Furthermore, the CNV-based de novo mutation rate is much higher than SNPs (Lupski, 2007). CNVs have been discovered in ~30% of human genomes and represent a significant source of genetic variation (Zhang et al., 2009). Chromosomal rearrangements have been reported to be associated with human diseases, particularly neurodegenerative diseases (Zhang et al., 2009). At least 13 loci, including triplication of SNCA on 4q21, are associated with Parkinson's disease (Polymeropoulos et al., 1996). An increase in the copy number of APP, the gene encoding amyloid precursor protein, is hypothesized to be one of the reasons for Alzheimer disease, and

duplications of APP have been reported (Goate, 1991). CNVs have been observed at high frequency in autism spectrum disorder (ASD) patients. The deletion and its reciprocal duplication on chromosome 16 were detected in about 1% of autism cases (Sebat et al., 2007). CNV-related disease models can be generated and used to elucidate the causes of complex diseases. The engineered-nuclease technology will be important as a next generation tool for genome engineering, including genome rearrangements.

One of the biggest roadblocks in the application of programmable nucleases in gene therapy and basic research is the lack of methods to enrich or select gene-modified cells. For example, the therapeutic efficacy of ZFNs that induce targeted disruption of the human chemokine receptor 5 (*CCR5*) gene, which encodes a co-receptor for the human immunodeficiency virus (HIV), depends on the number of *CCR5*-KO cells induced by ZFNs (Perez etal., 2008). However, only a limited fraction of cells are mutated by these ZFNs, and the remaining cells, which contain at least single copy of the intact *CCR5* gene, will serve as hosts for HIV replication. Although *CCR5*-KO cells can be selected *in vivo* because of their immunity to HIV infection, enrichment of mutant cells before transplantation could enhance their potential therapeutic efficacy. In addition, laborious screening of numerous clones is often required to obtain gene-disrupted cells because only a minor fraction of cells are modified by nucleases (Santiago et al., 2008; Kim et al., 2009). A generalized method to enrich or isolate cells in which a target gene has been disrupted by

programmable nucleases would greatly facilitate their widespread application.

In this report, I describe a novel surrogate reporter system that enables efficient enrichment of gene-modified cells (Kim et al., 2011). I previously noted that ZFNs often induce bi-allelic modifications in clonal populations of cells. Thus, when a target sequence on a chromosome is mutated by a pair of ZFNs, which function as dimers, the same target sequence on the homologous chromosome in the same cell is much more likely to be mutated as well than the target sequence in a different cell (Perez et al., 2008; Lee et al., 2010; Kim et al., 2009). Up to 33% of mutant cells contain bi-allelic mutations. Apparently, only a fraction of cells is susceptible to ZFN-driven mutations in these cells mutations at one allele often coincide with those at the other allele. This observation suggested that a surrogate reporter containing the target sequence of a programmable nuclease may faithfully reflect the activity of the nuclease in a given cell, and therefore maybe used for selective enrichment of cells in which the endogenous target sequence is also modified by the nuclease.

I. Materials and Methods

1. Construction of ZFN by modular assembly

Plasmids that encode ZFNs used in this study were constructed by 2-finger modular assembly (Figure 1) (Kim et al., 2011). This modularly-assembled ZFNs contain three or four zinc-fingers and the genome editing activity of ZFNs largely depends on the choice of zinc-finger module. Total 33 zinc-fingers with different DNA-binding specificity were chosen, which collectively recognize 35 out of 64 3-base pair sub-sites (Table 1). A combinatorial library of two-finger modules were prepared, this library consists of 1,089 (= 33 x 33) 2-finger arrays, each linked to the FokI nuclease domain. DNA segments, each encoding one of 33 zinc-fingers (Table 1), were individually cloned into the p3 vector, a derivative of pcDNA3 (Invitrogen). In each plasmid, the zinc-finger-encoding fragment was flanked by an XmaI site and an AgeI site. These two restriction enzymes recognize different 6-base pair DNA sequences but produce compatible cohesive ends, enabling the construction of zinc-finger arrays by repetition of a simple subcloning procedure. Additionally, to expedite the process of ZFN assembly, a library of single-finger ZFNs was established by cloning a DNA segment encoding the FokI endonuclease domain into a region downstream of each zinc-finger in the p3 vector using AgeI and XhoI sites. To prepare a combinatorial library of 2-finger ZFNs, we digested 33 single-finger ZFN-encoding plasmids with AgeI and XhoI to prepare vector fragments (p3-ZF) and

with XmaI and XhoI to prepare insert fragments (ZF-FokI). Ligation of each of the vector and insert fragments in all possible combinations (33 x 33) produced a semi-assembled ZFN library encoding 1,089 different 2-finger arrays. This library allows us to construct 3-finger or 4-finger ZFNs in a single sub-cloning step (Figure. 1), a process that could save time and labor significantly. With this pre-assembled two-finger library, it now takes only a few days to prepare several ZFNs to target a gene of interest. All the nuclease domains of ZFNs used in this study were obligatory heterodimers (KK/EL or sharkey DAS/RR) (Miller et al., 2007; Guo et al., 2010).

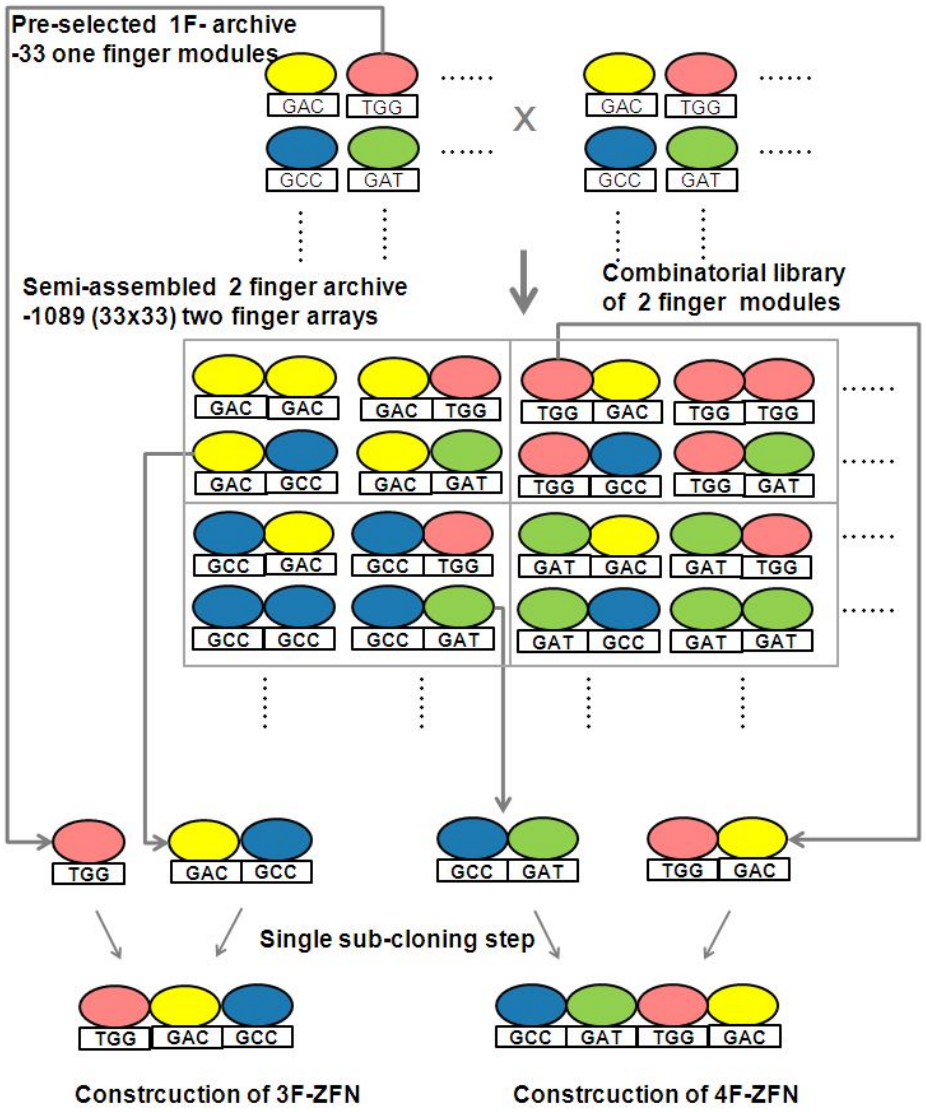


Figure 1. Schematic of ZFN modular assembly using semi-assembled zinc-finger arrays.

Table 1. List of zinc-finger used for the construction of semi-assembled zinc-finger arrays

ZF No.[1]	ZF name	Source	Amino acid sequence	Target subsite
	DSAR2	ToolGen	YSCGICGKSFSDSSAKRRHCILH	GTC
ZF108	DSCR	ToolGen	YTCSDCGKAFRDKSCLNRHRRTH	GCC
	DSNR	ToolGen	YRCKYCDRSFSDSSNLQRHVRNIH	GAC
ZF112	ISNR	ToolGen	YRCKYCDRSFSISSNLQRHVRNIH	GAT
ZF113	KSNR	ToolGen	YGCHLCGKAFSKSSNLRRHEMIH	GAG
	QNTQ	ToolGen	YTCSYCGKSFTQSNTLKQHTRIH	ATA
ZF117	QSHR2	ToolGen	YKCGQCGKFYSQVSHLTRHQKIH	GGA
ZF120	QSHV	ToolGen	YECDHCGKSFSQSSHLNVHKRTH	YGA
	QSNT	ToolGen	YECVQCGKGFTQSSNLITHQRVH	AAA
ZF123	QSNR1	ToolGen	FECKDCGKAFIQKSNLIRHQRTH	GAA
	QSNV3	ToolGen	YKCDECGKNFTQSSNLIVHKRIH	CAA
ZF126	QSSR1	ToolGen	YKCPDCGKSFSQSSSLIRHQRTH	GYA
ZF128	QTHQ	ToolGen	YECHDCGKSFRQSTHLTQHRRIH	AGA
	RDER2	ToolGen	YHCDWDGCGWKFARSDELTRHYRKH	GYG
ZF132	RDHT	ToolGen	FQCKTCQRKFSRSDHLKTHTRTH	HGG
ZF134	RSHR	ToolGen	YKCMECGKAFNRRSHLTRHQRIH	GGG
	VDYK	ToolGen	FHCGYCEKSFSVKDYLTKHIRTH	TAT
ZF137	VSNV	ToolGen	YECDHCGKAFSVSSNLNVHRRIH	MAT
ZF139	VSTR	ToolGen	YECNYCGKTFSVSSTLIRHQRIH	GCT
ZF140	WSNR	ToolGen	YRCEECGKAFRWPSNLTRHKRIH	GGT
ZF52	rdnq	Sangamo	FACPECPKRFMRSDNLTQHIKTH	AAG
ZF75	dgnv	Barbas	FQCRICMRNFSDSGNLRVHIRTH	AAC
ZF91	rdne	Barbas	YKCPECGKSFSRADNLTEHQRTH	CAG
ZF93	thse	Barbas	YKCPECGKSFSTSHSLTEHQRTH	CCA
ZF96	tnse	Barbas	YKCPECGKSFSTKNSLTEHQRTH	CCT
ZF98	hghe	Barbas	YKCPECGKSFSHTGHLLEHQRTH	CGC
ZF100	srta	Barbas	YKCPECGKSFSSRRTCRAHQRTH	CGT
ZF101	qste	Barbas	YKCPECGKSFSQNSTLTEHQRTH	CTA
ZF104	rdnt	Barbas	YKCPECGKSFSREDNLHTHQRTH	TAG
ZF61	dghr	Barbas	YKCPECGKSFSDPGHLVRHQRTH	GGC
ZF78	sadr	Barbas	YKCPECGKSFSSPADLTRHQRTH	ACA
ZF81	tldr	Barbas	YKCPECGKSFSTHLDLIRHQRTH	ACT
ZF90	skae	Barbas	YKCPECGKSFSSKKALTEHQRTH	CAC

[1] ZF No. is based on the numbering scheme of the zinc-finger Consortium Modular Assembly Kit 1.0 available from Addgene.

2. Construction of TALEN

Plasmids that encode TALENs that target *CCR5* locus were constructed by modular assembly and the *BRCA1*-targeting TALENs, which were constructed by golden gate cloning, were obtained from ToolGen (Seoul, South Korea). The FokI domian of *CCR5*-TALEN were wild type and *BRCA1*-TALEN were obligatory heterodimers (sharkey DAS/RR). The *CCR5*-TALENs were constructed by following methods (Figure 2., Kim et al., 2012). Oligonucleotides that encode each TALE repeat module were synthesized and subcloned into the XbaI/NheI site in p3. The DNA sequence of a module termed HD is as follows:

5'-tctagagaccgtgcagcgcctgctgcccgtgctgtgccaggcccacggcctgacccccgagcaggtggt ggccatcgccagccacgacggcggcaagcaggcgctagc-3'.

Underlined sequences were changed to "aatggc", "aatatt", or "aataac" to encode NG, NI, or NN, respectively. One plasmid was digested with XbaI and XhoI to yield a vector backbone and the other with NheI and XhoI to yield an insert segment. To create a plasmid encoding a two-repeat array, the insert segment was ligated with the vector backbone. The resulting plasmids were subjected to the next round of subcloning using the same sets of restriction enzymes. Finally, modularly-assembled repeat arrays were subcloned into the expression vector p3 that encodes the Δ153 N-terminal domain of AvrBs3 at the N terminus and the FokI nuclease domain at the C terminus to create TALEN expression vectors.

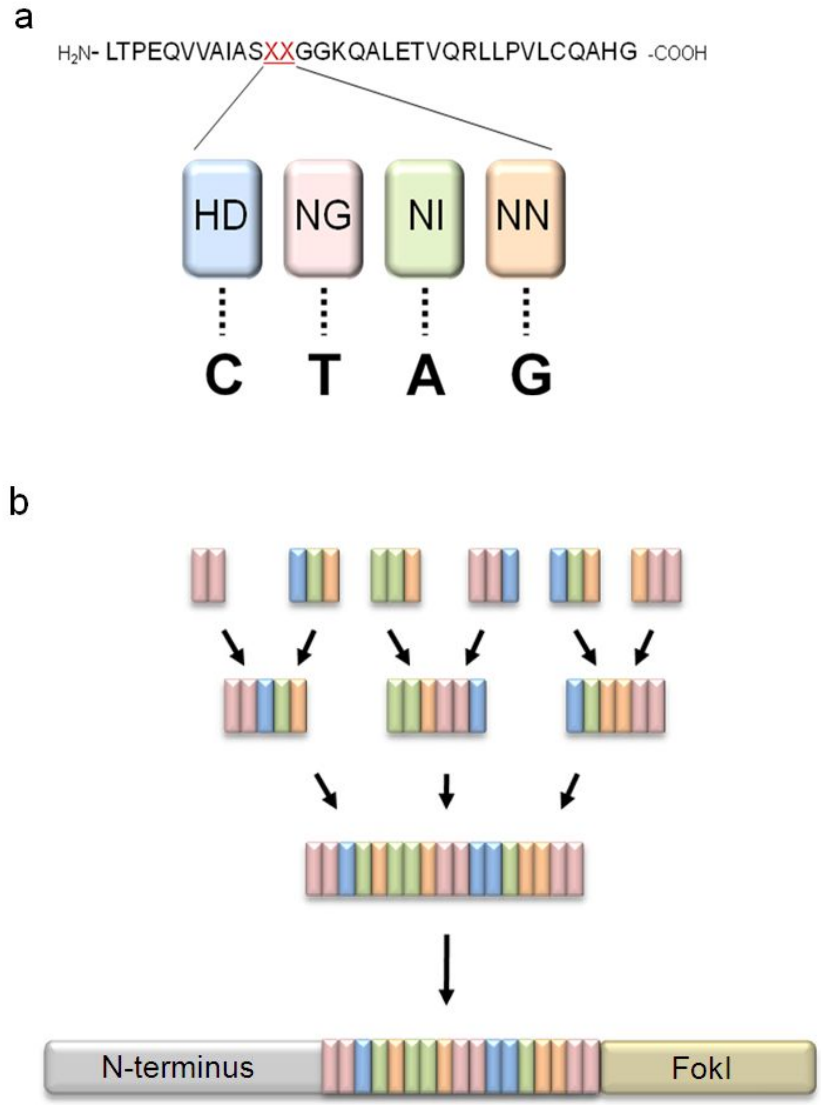


Figure 2. Schematic of the construction of dTALEs (Kim et al., 2012)**.** (a) The four TALE-repeat modules used for the construction of dTALEs. The amino acid sequence of a repeat module is shown. XX

denotes hyper-variable amino-acids at positions 12 and 13 which determine the specificity of base recognition. These two residues are shown in the boxes that represent repeat modules. (b) Step-wise construction of dTALEs. Plasmids encoding each repeat module are digested with either XbaI/XhoI or NheI/XhoI, and the insert segment from the NheI/XhoI-digested plasmid was ligated with the vector backbone from the XbaI/XhoI-digested plasmid to yield a two-repeat module. The resulting plasmids were subjected to the next round of subcloning using the same sets of restriction enzymes. Finally, modularly-assembled repeat arrays were subcloned into the expression vector p3.

3. Reporter plasmid contruction

The mRFP gene was amplified from pcDNA3-mRFP using the primers 1 and 2 (Table 2), and the amplified product was cloned into the NheI site of pEGFP-N1 (Clontech). The eGFP gene was amplified using the primers 3 and 4 and was then sequentially cloned into the BamHI and NotI site of the resulting plasmid. This vector was named pRGS. Oligonucleotides that contained target sites (Table 3) were synthesized (Bioneer, Daejon, South Korea) and annealed *in vitro*. The annealed oligonucleotides were ligated into the vector (pRGS) digested with EcoR1 and BamH1. Note that an in-frame stop codon should be avoided at the ZFN target site either by altering the frame or by changing the orientation of the target site. The sequence of reporter that contained.

The 2A sequence was inserted into the pRGS reporter[2] using synthesized oligonucleotides (Bioneer, Daejon, South Korea). The mouse $H2K^k$ gene was amplified from pMACS Kk (Miltenyi Biotech, Germany) using primers 5 and 6 (Table 2), and the amplified product was cloned into the modified pRGS vector by isothermal cloning using primers 7, 8, 9 and 10 (Table 2) (Gibson et al., 2009). The hygomycinB gene was amplified from pBABE-hygro-hTERT (Addgene, plasmid #1773) using using primers 11 and 12 (Table 2), and the amplified product was cloned into the NheI site of the modified pRGS vector.

4. Cell culture and transfection

Human embryonic kidney 293T/17 (ATCC, CRL-11268^{TM}) cells and Hur7.5 cells were maintained in Dulbecco's modified Eagle medium (DMEM, Invitrogen) supplemented with 100 units/ml penicillin, 100 mg/ml streptomycin, and 10% fetal bovine serum. A mouse induced pluripotent stem (iPS) cell line established by Andras Nagy and Knut Woltjen was obtained from Andras Nagy (Mount Sinai Hospital, Toronto, Canada) and was cultured in Glasgow modified Eagle medium (Sigma) supplemented with 10% fetal bovine serum, 0.1 mM nonessential amino acids (Invitrogen), 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, 2000 U/mL of leukemia inhibitory factor (LIF), 100 units/ml penicillin, and 100 ug/ml streptomycin on gelatinized culture dishes without feeder cells. For derivation of fibroblasts from the mouse iPS cells, cells were cultured in the absence of LIF for 3 weeks.

Human embryonic kidney 293T/17 (ATCC, CRL-11268^{TM}) cells and Hur7.5 cells were transfected using lipofectamine 2000 (Invitrogen) or polyethyleneimine (linear, MW ~25,000, Polysciences); transfection of fibroblasts derived from mouse iPS cells was performed using Magnetofection (Chemicell) at the weight ratio of 1:1:2 (plasmid encoding a ZFN: plasmid encoding the other ZFN: reporter) or 10:10:1 (hygromycin reporter).

5. Cell sorting by flow cytometry (FACS)

Adherent cells were trypsinized and resuspended in 2% FBS in PBS. Single cell suspensions were analyzed and sorted using the FACS AriaII (BD Biosciences, San Jose, CA) or FACS VantageSE (BD Biosciences). To collect cells that contain nuclease-induced mutations, cells with strong GFP signals were sorted. Untransfected cells and cells transfected with reporters alone were used as controls.

6. Magnetic-activated cell separation (MACS)

The transfected cells were cultured for two days and subjected to magnetic separation. Trypsinized cell suspensions were mixed with magnetic bead-conjugated antibody against H-2k^k (MACSelect Kk microbeads; Miltenyi Biotech, Germany) and incubated for 15 min at 4℃. Labeled cells were separated using a column (MACS LS or MS column; Miltenyi Biotech) according to the manufacturer's instructions.

7. Hygromycin selection

Two days after transfection, hygromycin selection was performed by culturing the cells in the presence of 2 mg/ml of hygromycin B for two days at 37°C.

8. T7E1 assay for mutation detection

Genomic DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. The region of DNA containing the nuclease target site was PCR-amplified using the primers described in Table 4. The amplicons were denatured by heating and annealed to form heteroduplex DNA, which was treated with 5 units of T7 endonuclease 1 (New England Biolabs) for 20 min at 37°C and then analyzed by agarose gel electrophoresis.

9. Fluorescent PCR analysis

Genomic DNA (100 ng per reaction) was subjected to PCR analysis using phusion polymerase (Fynnzymes) and 5'-FAM labeled primers in Table 5. The PCR amplicons were studied using an ABI 3730xl DNA analyzer. The positions and sizes of the peaks indicate the lengths and relative amounts of PCR products.

10. Estimation of mutation frequency by sequencing

PCR amplicons that contained ZFN-induced small indel mutations were purified using the Gel Extraction Kit (MACHERRY-NALGEN) and cloned into the T-Blunt vector using the T-Blunt PCR Cloning Kit (SolGent). Cloned amplicons were sequenced with primers used for PCR amplification.

11. Clonal analysis of single cells and colonies

Before and after cell sorting, single cells were isolated using a mouth pipette under a microscope and transferred to PCR tubes. PCR products were cloned and sequenced. In order to obtain clonal populations of cells, sorted and unsorted cells were plated at a density of 1,000 cells/100mm dish or 0.25 cells/well in 96well plates and colonies were manually picked two weeks later.

Table 2. The primer sequences for reporter construction

Primer Name	Sequence (5' to 3')
1	GCGGCTAGCCACCATGGCCTCCTCCGAGGACGTCATC
2	GCGGCTAGCGAATTCGGCGCCGGTGGAGTGGCGGCCC
3	GCGGGATCCAGTGAGCAAGGGCGAGGAGCTG
4	GTCGCGGCCGCTTTACTTGTAC
5	GGCGCTAGCATGGCACCCTGCATGCTGCTT
6	GCCGCGGCCGCTTACCCTCCTTTTCCACCTGTGTT
7	GAGTTTCAACAAAGCGTAGTTAGTACATTGCTTGTACA GCTCGTCCATGC
8	GAAGGAGAAACACAGGTGGAAAAGGAGGGTAAAGCG GCCGCGACTCTAG
9	ATCACTCTCGGCATGGACGAGCTGTACAAGCAATGTA CTAACTACGCTTT
10	GATTATGATCTAGAGTCGCGGCCGCTTTACCCTCCTTTT
11	GGCGCTAGCATGAAAAAGCCTGAACTCACCG
12	GCGTCTAGAGTCGACTTCCTTTGCCCTCGGACGAGTG

Table 3. The target sequences of programmable nucleases

Programmable nuclease	Target sequence (5' to 3')
TP53-ZFN_L	GGCGCGGACGCG
TP53-ZFN_R	CATCTACAAGCA
CCR5-ZFN224_L	GATGAGGATGAC
CCR5-ZFN224_R	AAACTGCAAAAG
CCR5-Z891_L	ATAGATGATGGG
CCR5-Z891_R	GTCGGGGAGAAG
CCR5-TALEN_L	TGCATCAACCCCATCATC
CCR5-TALEN_R	TAGTTTCTGAACTTCTCCCC
Thumpd3-ZFN_L	CGAGCACGCCGC
Thumpd3-ZFN_R	GGAGACCGGAAG
CD81-ZFN_L	GGCATCTACATC
CD81-ZFN_R	AGCGCCCACAGC
BRCA1-TALEN_L	TTTGCAGAATACATTCAAGG
BRCA1-TALEN_R	TGAAAACGGAGCAAATGACT

Table 4. The primer sequences for T7E1 analysis

Primer Name	Sequence (5' to 3')
TP53-ZFN-F	GCAGGAGGTGCTTACGCATGTTTGT
TP53-ZFN-R	GCTGCTCACCATCGCTATCTGAGC
CCR5-ZFN224-F	GAGCCAAGCTCTCCATCTAGT
CCR5-ZFN224-R	CTGTATGGAAAATGAGAGCTGC
CCR5-Z891-NF	GAGCCAAGCTCTCCATCTAGT
CCR5-Z891-F	TTAAAGATAGTCATCTTGGGGC
CCR5-Z891-R	TCACAAGCCCACAGATATTT
CCR5-TALEN-NF	GAGCCAAGCTCTCCATCTAGT
CCR5-TALEN-F	TTAAAGATAGTCATCTTGGGGC
CCR5-TALEN-R	TCACAAGCCCACAGATATTT
Thumpd3-ZFN-F	CAACCGAGCATCCGCTCGCTAGG
Thumpd3-ZFN-R	GAAGGGGCTGGAGTGGTGTTACCG
BRCA1-TALEN-NF	TGATGGGAAAAAGTGGTGGT
BRCA1-TALEN-F	CAAAACCTAGAGCCTCCTTTGA
BRCA1-TALEN-NR	GCAACTGGAGCCAAGAAGAG
BRCA1-TALEN-R	GGCACTCAGGAAAGTATCTCG

Table 5. The primer sequences for fluorescent PCR analysis

Primer Name	Sequence (5' to 3')
TP53-ZFN-fF	GCAGGAGGTGCTTACGCATGTTTGT
TP53-ZFN-fR	*FAM*-GCTGCTCACCATCGCTATCTGAGC
CCR5-ZFN224-fF	TGCACAGGGTGGAACAAGATGG
CCR5-ZFN224-fR	*FAM*-GAGCCCAGAAGGGGACAGTAAGAAGG
CCR5-Z891-fF	*FAM*-GAATAATTGCAGTAGCTCTAACAGG
CCR5-Z891-fR	CTCTTGCTGGAAAATAGAACAG

III. Results

A. Enrichment of nuclease-induced mutant cells by FACS-based surrogate reporter system

1. Establishment of surrogate reporter system

I investigated whether it is possible to selected mutant cells by reporter system which have been used by measuring of programmable nuclease's activity indirectly. To this end, I prepared a reporter plasmid that encodes a monomeric red fluorescent protein (mRFP)-enhanced green fluorescent protein (eGFP) fusion protein (Figure 3a). A programmable nuclease target site is inserted between the DNA sequences encoding mRFP and eGFP such that the eGFP sequence is fused to the mRFP sequence out of frame. Thus, cells transfected with this reporter plasmid express mRFP (RFP^+) but not eGFP (GFP^-) (Figure 3b). In contrast, a fraction of cells co-transfected with the reporter and the nuclease-encoding plasmids would become RFP^+ GFP^+ if the nuclease cleaves DNA at the target site in the reporter and generates a DSB, which would then be repaired by error-prone NHEJ, often causing indels at the target site. This indel formation can give rise to frame shifts, leading to the expression of the functional mRFP-eGFP fusion protein. Because NHEJ-mediated indel formation is a random process, only one third of the reporter plasmids could

generate in-frame fusions. In practice, this is rarely a problem because multiple copies of the reporters are transfected into a single cell.

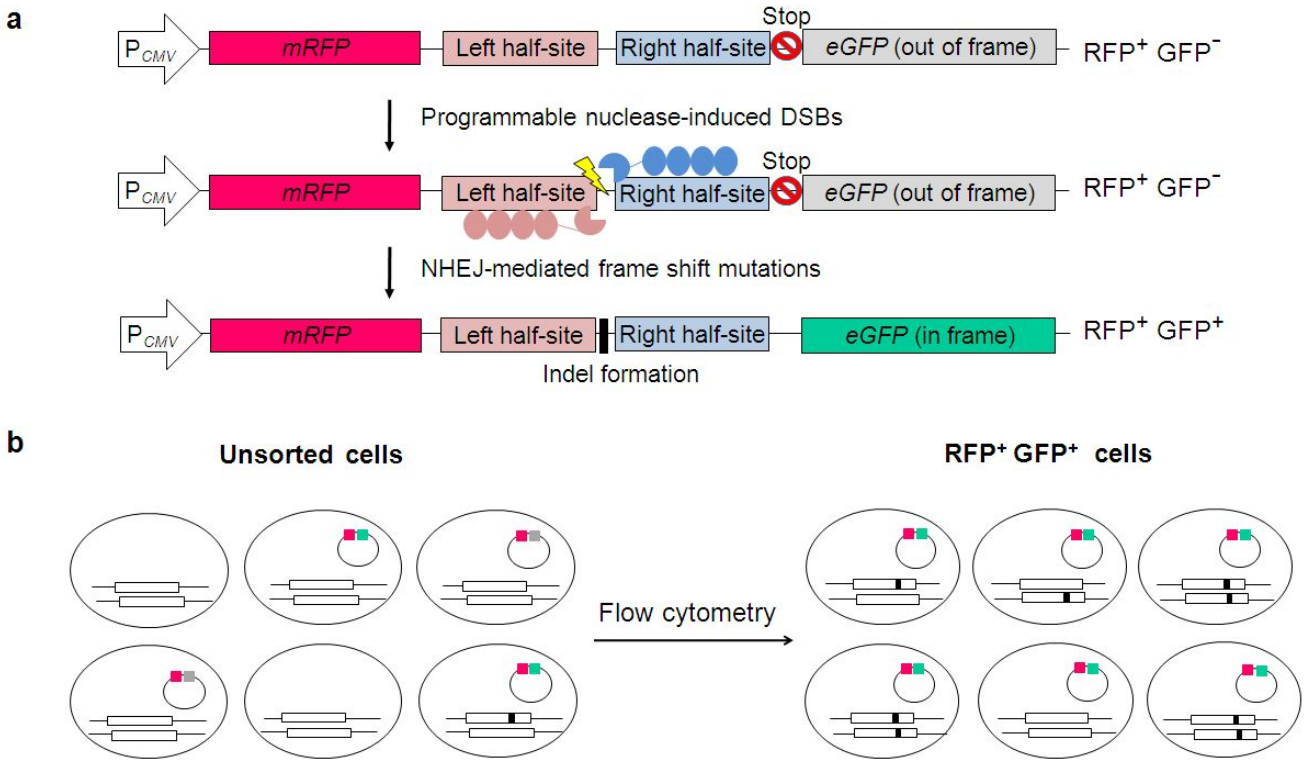

a
Stop
P CMV
mRFP
Left half-site
Right half-site
eGFP (out of frame)
RFP+ GFP-
Programmable nuclease-induced DSBs
Stop
P CMV
mRFP
Left half-site
Right half-site
eGFP (out of frame)
RFP+ GFP-
NHEJ-mediated frame shift mutations
P CMV
mRFP
Left half-site
Right half-site
eGFP (in frame)
RFP+ GFP+
Indel formation
b
Unsorted cells
RFP+ GFP+ cells
Flow cytometry

Figure 3. The structure of reporters and flow cytometric enrichment of gene-modified cells using the reporters. (a) The reporter consists of the mRFP gene, the programmable nuclease's target sequence and the eGFP gene. mRFP is constitutively expressed by the CMV promoter, whereas functional eGFP is not expressed because it is out of frame in the absence of a programmable nuclease activity. When a DSB is introduced into the target sequence by programmable nucleases, the break is repaired by NHEJ, which often causes frame shift mutations. Such mutations can render eGFP in frame with mRFP, inducing the expression of the fully functional mRFP-eGFP fusion protein. (b) Three or four days after co-transfection of reporter plasmids and plasmids encoding programmable nucleases, cells were sorted and analyzed using flow cytometry.

2. Enrichment of nuclease-induced mutant cells by surrogate reporter system

I co-transfected a set of plasmids encoding a ZFN pair that are designed to target the human *TP53* gene together with a reporter plasmid that contains the nuclease's target site, into HEK293 cells. As controls, HEK293 cells were transfected with either the reporter plasmid alone or the ZFN plasmids alone. At 24 hrs after transfection, a significant fraction of cells became $RFP^{+,}$ whereas GFP^{+}cells were barely detected. The number of GFP^{+}cells gradually increased over 3 days all of these GFP^{+} cells were also RFP^{+} as expected (Figure 4). At 72 hrs after transfection, flow cytometric analysis showed that about 16% and 5% of cells became RFP^{+} GFP^{-} and RFP^{+} GFP^{+}, respectively (Figure 5a). These cell populations were sorted by flow cytometry, and the genomic DNA was isolated and analyzed to evaluate the extent of nuclease-driven mutations. A T7 endonuclease I (T7E1) assay revealed that the mutation frequency at the *TP53* gene in the RFP^{+} GFP^{+} cells was 37%, which is 13-fold higher than that in the unsorted cells (Figure 5b), demonstrating efficient enrichment of gene-disrupted cells. Mutation frequencies in unsorted cells, RFP^{-} GFP^{-} cells, RFP^{+} GFP^{-} cells, and cells transfected with ZFN plasmids alonewere all comparable, ranging from 2.8 to 4.8%.

I also performed fluorescent polymerase chain reaction (fPCR) analysis to quantify ZFN-driven indels and found that indels were enriched in RFP^{+} GFP^{+} populations by up to 39-fold as compared to

unsorted cells, RFP^- GFP^- cells, or RFP^+ GFP^- cells (Figure 5c), corroborating the T7E1 assay result. The T7E1 assay tends to underestimate fold enrichment. At high mutation frequencies, mutant sequences can form homoduplexes, which are insensitive to T7E1 digestion.

I next cloned PCR products, determined the DNA sequences around the target site, and found that the mutation frequency in the RFP^+ GFP^+ fraction and in unsorted cells was 20% and 1%, respectively, indicating a 20-fold enrichment of mutations by cell sorting (Figure 5d). Taken together, these results demonstrate that the surrogate reporter is a reliable system to monitor the activity of ZFNs in live cells, enabling efficient enrichment of target gene-modified cells.

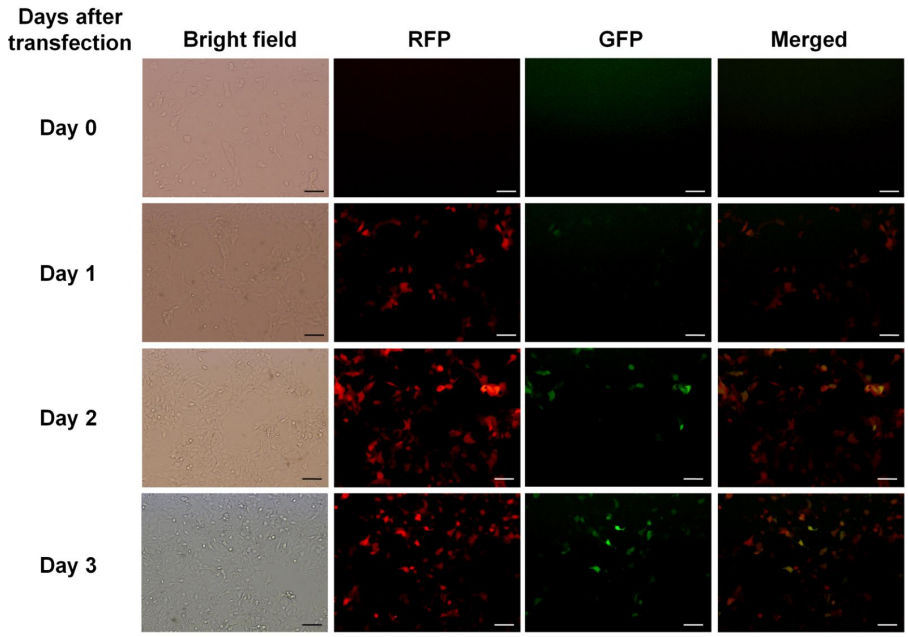
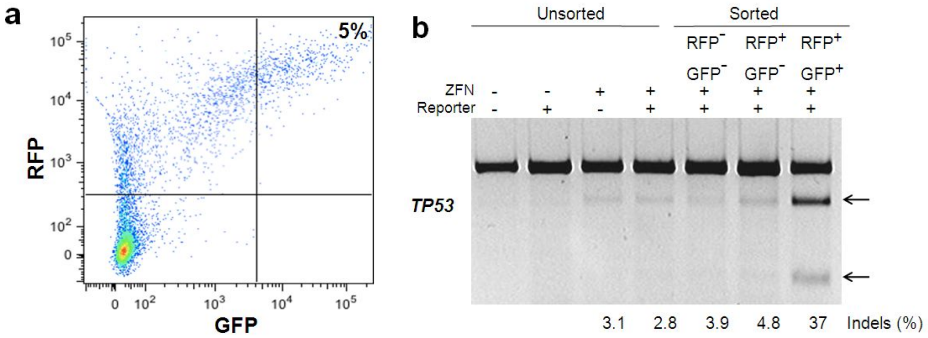
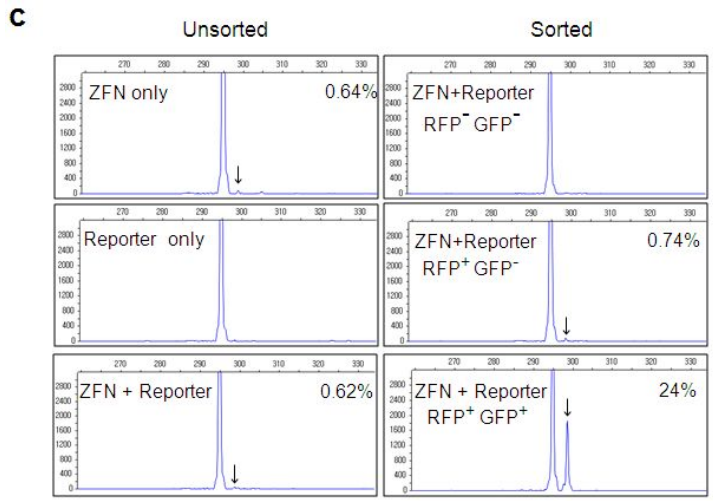


Figure 4. Expression of RFP and GFP in HEK293 cells.

HEK293 cells were cotransfected with a reporter plasmid and plasmids encoding ZFNs that target the *TP53* gene and observed daily using fluorescent microscopy. Scale bar = 100 mm.





d

Unsorted, 1.0% (= 3/288) mutated

```
GGCACCCGCGTCCGCGCC-----ATGGCCATCTACAAGCAGTCAC  (WT)
GGCACCCGCGTCCGCGCCatggtATGGCCATCTACAAGCAGTCAC  (X1)
GGCACCCGCGTCCGCGCCatgg-ATGGCCATCTACAAGCAGTCAC  (X1)
GGCACCCGCGTCCGCGCCat---ATGGCCATCTACAAGCAGTCAC  (X1)
```

Sorted (RFP⁺ GFP⁺), 20% (= 8/40) mutated

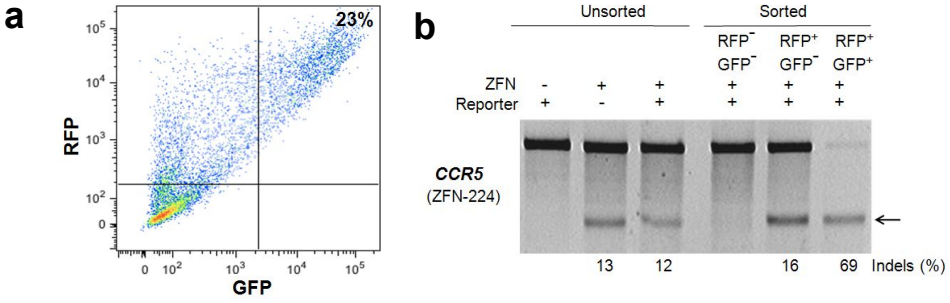
```
GGCACCCGCGTCCGCGCC-----ATGGCCATCTACAAGCAGTCAC  (WT)
GGCACCCGCGTCCGCGCCcatggATGGCCATCTACAAGCAGTCAC  (X1)
GGCACCCGCGTCCGCGCC-atggATGGCCATCTACAAGCAGTCAC  (X5)
GGCACCCGCGTCCGCGCC-at--ATGGCCATCTACAAGCAGTCAC  (X1)
GGCACCCGCGTCCGCGCCc----ATGGCCATCTACAAGCAGTCAC  (X1)
```

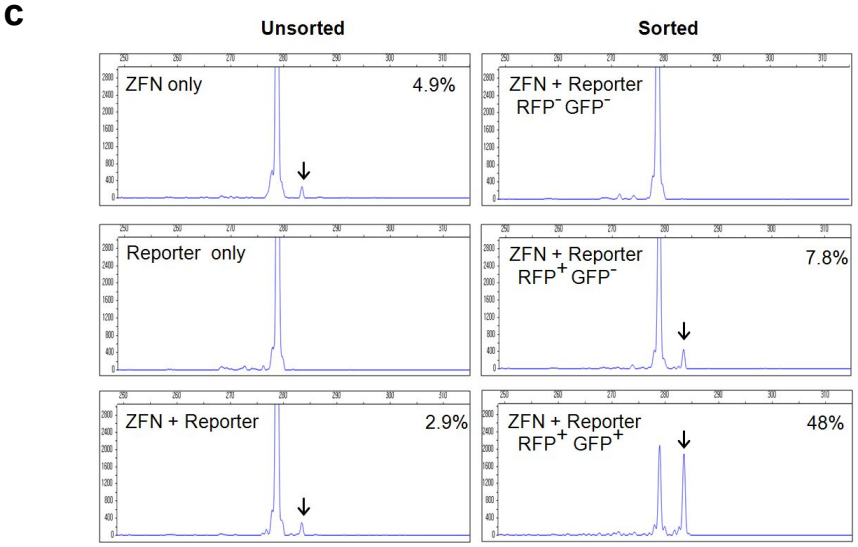
Figure 5. A surrogate reporter drastically enriches *TP53* gene-disrupted cells. (a) Flow cytometry of HEK293 cells three days after co-transfection of the *TP53*-targeting ZFN and the reporter. (b) ZFN-driven mutations detected by the T7E1 assay. Arrows indicate the expected positions of DNA bands cleaved by mismatch-sensitive T7E1. The numbers at the bottom of the gel indicate mutation frequencies measured by band intensities. (c) ZFN-driven mutation rates measured by fPCR. Arrows indicate amplified DNA peaks that correspond to small insertions. The tallest peaks correspond to wild-type amplicons. Mutation rates are calculated by measuring the peak area. (d) DNA sequences of the *TP53* gene targeted by the ZFN. ZFN recognition sites are underlined. Dashes indicate deleted bases, and small bold letters indicate inserted bases. In cases in which a mutation was detected more than once, the number of occurrences is shown in parentheses. Mutation frequencies are calculated by dividing the number of mutant clones by the number of total clones. WT, wild-type sequence.

I next determined whether this reporter system is useful for the enrichment of gene-modified cells by other ZFNs. For this analysis, I used two ZFNs, Z891 (Kim et al., 2009) and ZFN-224 (Perez et al., 2008), which target two different sites in the human *CCR5* gene. At 72 hours after the transfection of ZFN-224, 23% of cells became RFP^+ GFP^+ (Figure 6a). The T7E1 assay showed that the mutation rate in the RFP^+ GFP^+ cells was 69%, 5.8-fold higher than that in the unsorted cells (Figure 6b). Mutation frequencies in unsorted cells, RFP^- GFP^- cells, and RFP^+ GFP^- cells were all comparable, ranging from 12% to 16%. fPCR analysis also showed that mutations were enriched by 17-fold in RFP^+ GFP^+ cells as compared to unsorted cells (Figure 6c). I cloned PCR products, determined the DNA sequences around the target site, and found that the mutation frequency in the RFP^+ GFP^+ fraction and in unsorted cells was 50% and 8.3%, respectively. (Figure 6d). A derivative of ZFN-224 is now under clinical investigation in the U.S. for the treatment of AIDS. This surrogate reporter system is likely to enhance the efficacy of this ZFN significantly by enriching *CCR5*-knockout cells if used in ex vivo gene therapy.

I also found that Z891-driven mutations were enriched in RFP^+ GFP^+ cells by 11-fold (T7E1 assay, Figure 7b) or 38-fold (fPCR, Figure 7c) as compared to unsorted cells. Futhermore, this surrogate system is noninvasive, allowing additional cycles of transfection and cell sorting to further enrich the populations of mutant cells (Figrue 8). Repeated cycles of cotransfection and sorting collectively enabled a total

60-fold (= 48/0.8) enrichment of mutant cells, leading to the isolation of a cell population in which almost the half of the *CCR5* alleles were disrupted.





d

Unsorted, 8.3% (=2/24) mutated

```
ATGCTGGTCATCCTCATCCTGAT-----AAACTGCAAAAGGCTGAA (WT)
ATGCTGGTCATCCTCATCCTGATctgatAAACTGCAAAAGGCTGAA (x2)
```

Sorted (RFP⁺ GFP⁺), 50% (=6/12) mutated

```
ATGCTGGTCATCCTCATCCTGAT-----AAACTGCAAAAGGCTGAA (WT)
ATGCTGGTCATCCTCATCCTGATctgatAAACTGCAAAAGGCTGAA (x2)
ATGCTGGTCATCCTCATCCTGATgat--AAACTGCAAAAGGCTGAA (x1)
ATGCTGGTCATCCTCATCCTG--------------------CTGAA (x1)
ATGCTGGTCATCCT------------------------------GAA (x1)
ATGCTGGTCATCCTCATCCTGAT--------------------GAA (x1)
```

Figure 6. Enrichment of *CCR5* gene-modified cells induced by ZFN-224. HEK293 cells were cotransfected with the reporter plasmid and plasmids encoding ZFN-224, a ZFN that targets the *CCR5* gene. Three days after transfection, cells were sorted using flow cytometry and genomic DNA was isolated. (a) Flow cytometry of the transfected cells. (b) T7E1 assay using genomic DNA isolated from the flow-cytometrically sorted cells. (c) fPCR was performed to determine the efficiency of indel formation in the *CCR5* gene. Arrows indicate amplified DNA peaks that correspond to small insertions. (d) DNA sequences of the *TP53* gene targeted by the ZFN. ZFN recognition sites are underlined. Dashes indicate deleted bases, and small bold letters indicate inserted bases. In cases in which a mutation was detected more than once, the number of occurrences is shown in parentheses. Mutation frequencies are calculated by dividing the number of mutant clones by the number of total clones. WT, wild-type sequence.

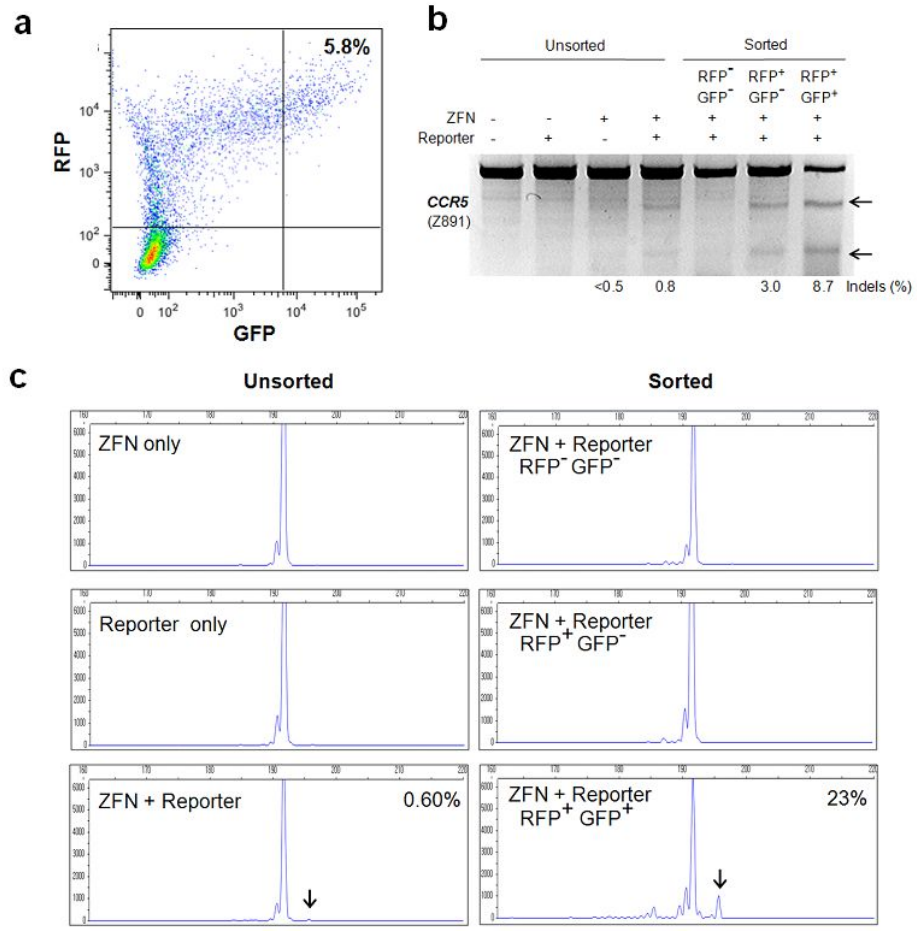


Figure 7. Enrichment of *CCR5* gene-modified cells induced by Z891. HEK293 cells were cotransfected with the reporter plasmid and plasmids encoding Z891, a ZFN that targets the *CCR5* gene. Three days after transfection, cells were sorted using flow cytometry and genomic DNA was isolated. (a) Flow cytometry of the transfected cells. (b) T7E1 assay using genomic DNA isolated from the flow-cytometrically sorted cells. (c) fPCR was performed to determine the rate of indel formation in the *CCR5* gene.

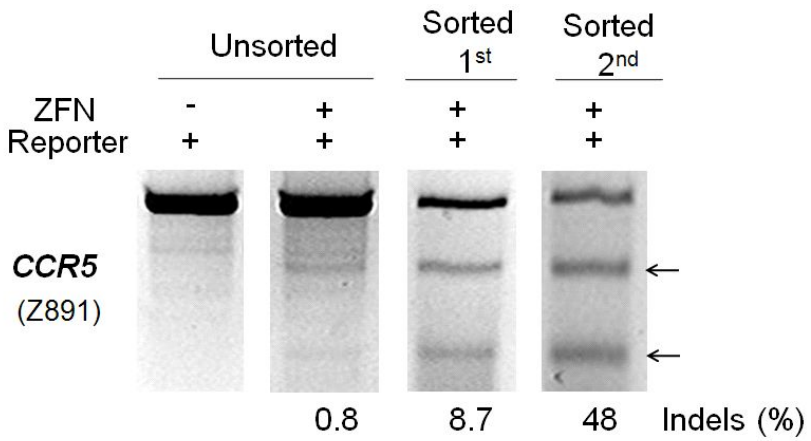


Figure 8. Repeated cycles further enriches *CCR5* gene-modified cells. HEK293 cells were transfected with a reporter plasmid and plasmids encoding Z891. Three days after transfection, RFP^+ GFP^+ cells were sorted using flow cytometry (sorted 1^{st}) and analyzed with the T7E1 assay. These cells were cultured for 24 hours and subjected to an additional cycle of cotransfection using the reporter plasmid and the Z891 plasmids and cell sorting. The isolated RFP+ GFP+ cells were also analyzed with the T7E1 assay (sorted 2^{nd}).

I also investigated whether this reporter system is compatible with TALENs, a new class of programmable nucleases. Although TALEN activity was barely detected in unsorted cells, the gene disruption was clearly observed in RFP^+ GFP^+ cells, revealing at least 8.6-fold enrichment of mutated cells as compared to unsorted cells (Figure 9).

Two different target sites in the human genome were disrupted by using ZFNs or TALENs and the reporter system allows to enable enrichment of nuclease-induced mutant cells (Figure 10, Table 6).

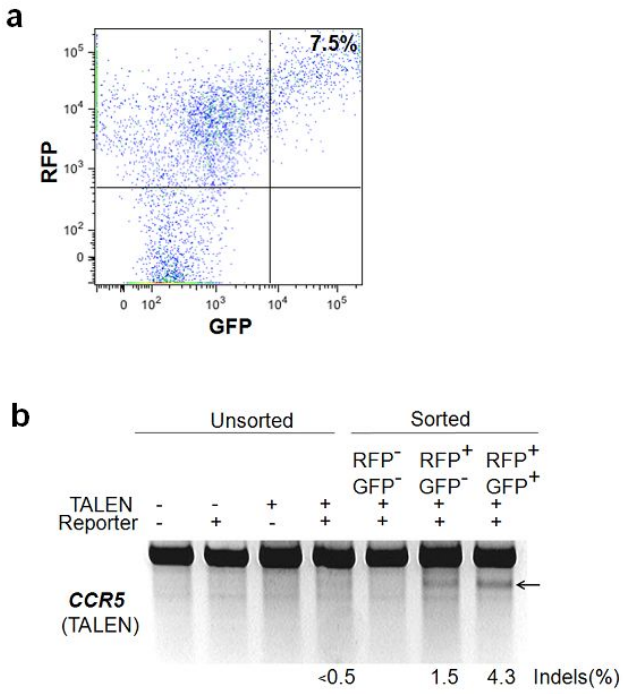
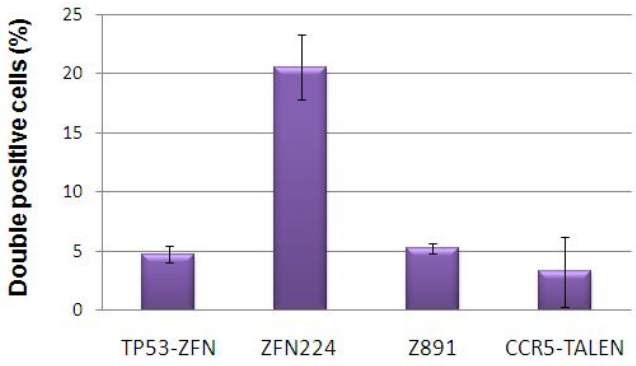
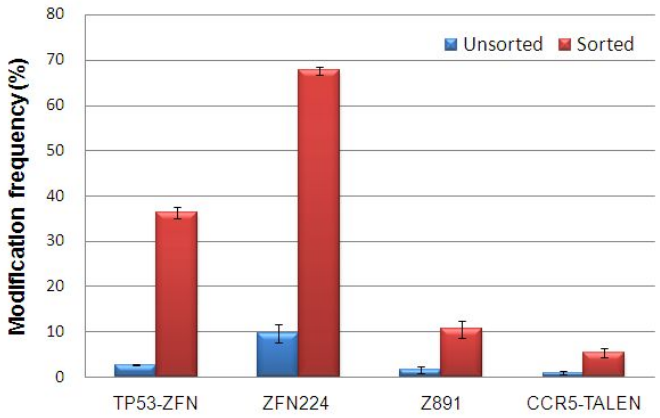


Figure 9. Enrichment of cells with TALEN-mediated disruption of the *CCR5* gene. HEK293 cells were cotransfected with the reporter plasmid and plasmids encoding a TALEN pair that targets the *CCR5* gene. The cells were then cultured at 37°C for one day and 30°C for the following three days, and subjected to analysis. (a) Flow cytometry of the transfected cells. (b) The genomic DNA was isolated from the flow-cytometrically sorted cells and subjected to the T7E1 assay. *CCR5* gene-modified cells were enriched in the RFP^+ GFP+ population by > 8.6 fold (= 4.3/0.5) as compared to unsorted cells. The bands indicated by the arrow represent the amplicon cut by T7E1; relative band density indicates TALEN activity.

a


Double positive cells (%)
25
20
15
10
5
0
TP53-ZFN
ZFN224
Z891
CCR5-TALEN

b


Modification frequency (%)
80
70
60
50
40
30
20
10
0
Unsorted
Sorted
TP53-ZFN
ZFN224
Z891
CCR5-TALEN

c

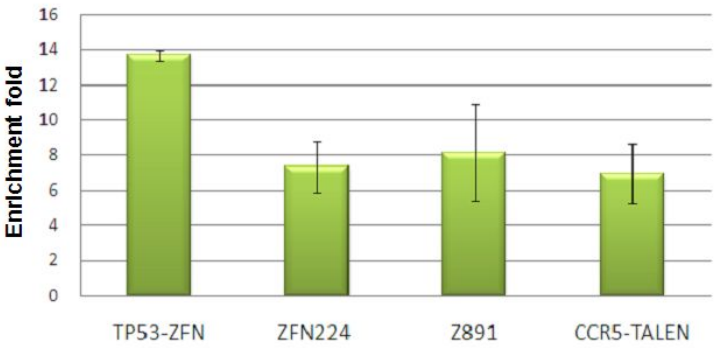

Enrichment fold
16
14
12
10
8
6
4
2
0
TP53-ZFN
ZFN224
Z891
CCR5-TALEN

Figure 10. Enrichment of ZFN/TALEN-mediated mutation frequencies after sorting. (a) RFP^+ GFP^+ cells were analyzed by using a flow cytomer. (b) Modification frequencies at the target locus of each ZFNs or TALENs. Blue bars indicate the mutation rates of unsorted cells and red bars indicate sorted cells. (c) Enrichment fold of mutation frequencies in sorted cells as compared to unsorted cells. This data represent an average of the three independent experiments and the standard error of the mean is shown.

Table 6. List of mutation frequencies after sorting

Target locus		Mutation rates (%)		Fold enrichment
		Unsorted	Sorted	
TP53	ZFN	2.7 ± 0.15	36 ± 1.2	14 ± 0.32
CCR5	ZFN-224	9.6 ± 2.0	68 ± 0.90	7.3 ± 0.44
	Z891	1.0 ± 0.75	11 ± 1.8	8.1 ± 2.8
	TALEN	0.9 ± 0.35	5.3 ± 1.0	6.9 ± 1.7

3. Isolation of mutant clones in mouse fibroblasts derived from induced pluripotent stem cells.

I next tested this reporters in a cell line derived from a different species. Plasmids encoding a ZFN pair that target a mouse gene, *Thumpd3* (Kim et al., 2011), and the reporter were co-transfected into mouse fibroblasts derived from induced pluripotent stem cells. Almost half of the *Thumpd3* alleles (46%) were disrupted in RFP^+ GFP^+ cells, whereas ZFN activity was barely detectable in unsorted cells, suggesting at least 92-fold enrichment of gene-disrupted cells (Figure 11).

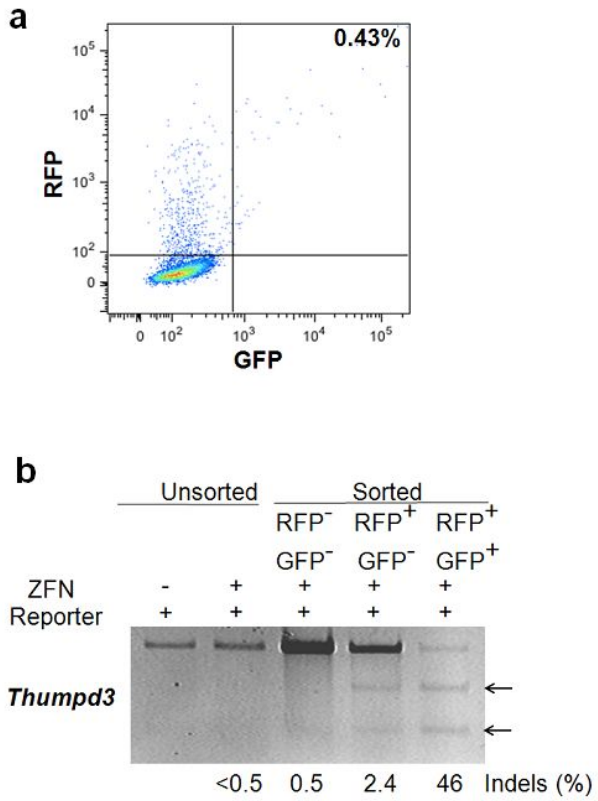


Figure 11. Enrichment of mouse cells with a disrupted *Thumpd3* gene. Mouse fibroblasts derived from induced pluripotent stem cells were analyzed three days after cotransfection with the reporter plasmid and plasmids encoding a ZFN pair that targets the *Thumpd3* gene. (a) Flow cytometry of the transfected cells. (b) The genomic DNA was isolated from the flow-cytometrically sorted cells and subjected to the T7E1 assay. The bands indicated by the arrows represent the amplicon cut by T7E1; relative band density indicates ZFN activity.

I then analyzed single cells and clonal populations of cells before and after cell sorting. None of 21 single mouse fibroblast cells contained ZFN-induced mutations at the *Thumpd3* site before cell sorting (Figure 12). In contrast, 9 single cells out of 10 RFP^{+} GFP+ cells contained mutations and, among these mutant cells, 4 cells contained different biallelic mutations. Furthermore, I was able to isolate two independent clones of *Thumpd3*-disrupted fibroblasts only after cell sorting, in which ZFN-induced mutagenesis is very inefficient (Figure 12). Taken together, these clonal analyses corroborate that the reporter system enables efficient enrichment of gene-disrupted cells with both monoallelic and biallelic mutations.

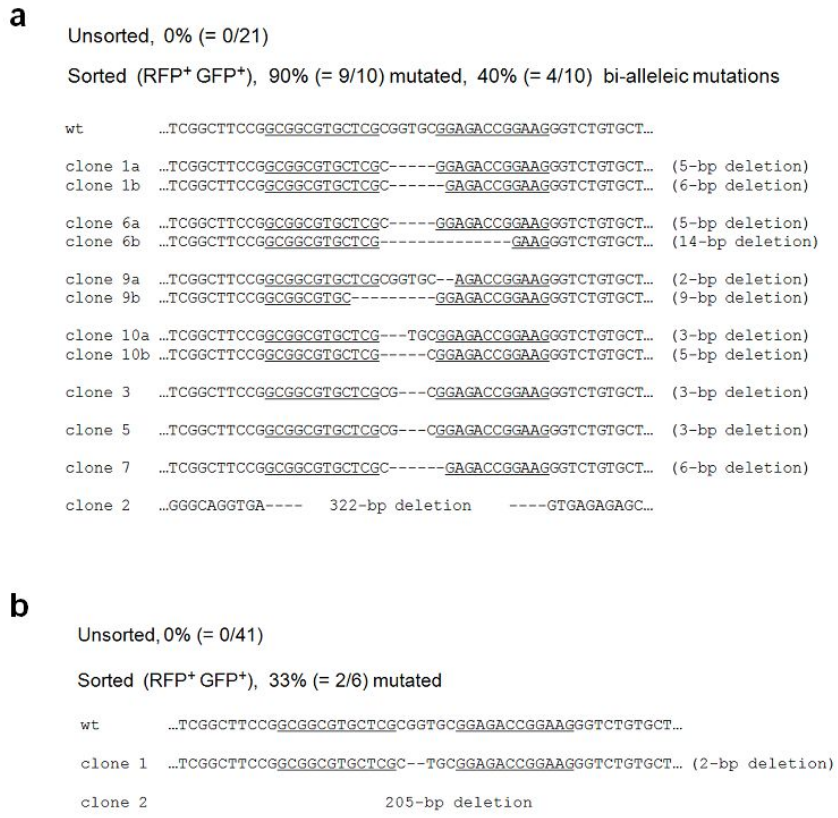


Figure 12. Clonal analysis of single cells and colonies. (a) Single cell analysis. Mouse fibroblasts derived from induced pluripotent stem cells were cotransfected with the reporter plasmid and plasmids encoding a ZFN pair that targets the *Thumpd3* gene. Three days after transfection, cells were sorted using flow cytometry, and single cells were isolated using a mouth pipette under a microscope and transferred to PCR tubes. PCR products from 21 unsorted cells and 10 sorted cells were cloned and sequenced. No mutations were detected in unsorted cells. In contrast, 9 of 10 sorted cells contained mutations, and four cells

contained biallelic mutations. "1a and 1b" indicate DNA sequences that resulted from bialleleic mutations in a single clone. (b) Clonal populations of cells. Sorted and unsorted cells were plated at a density of 1,000 cells/100 mm dish, and colonies were manually picked two weeks later. All 41 clones from unsorted cells contained the wild-type sequence. In contrast, two of 6 sorted clones contained ZFN-induced mutations.

4. Off-target effects analysis

Bioinformatic analysis was performed to search for sites that are most similar to the *TP53* target site (Cradick et al., 2011). All potential half-sites for the two ZFN monomers were identified in the human genome, allowing up to 2-base mismatches from the *TP53* target site. Two-half sites separated by a 5- or 6-bp spacer were identified and ranked based on the similarity score, which was calculated as the product of the percent identify at the two half-sites. The top 5 potential off-target sites are listed in Table 7. The T7E1 assay results are shown in Figure 13. Importantly, this enrichment process did not exacerbate off-target effects of the *TP53* ZFN.

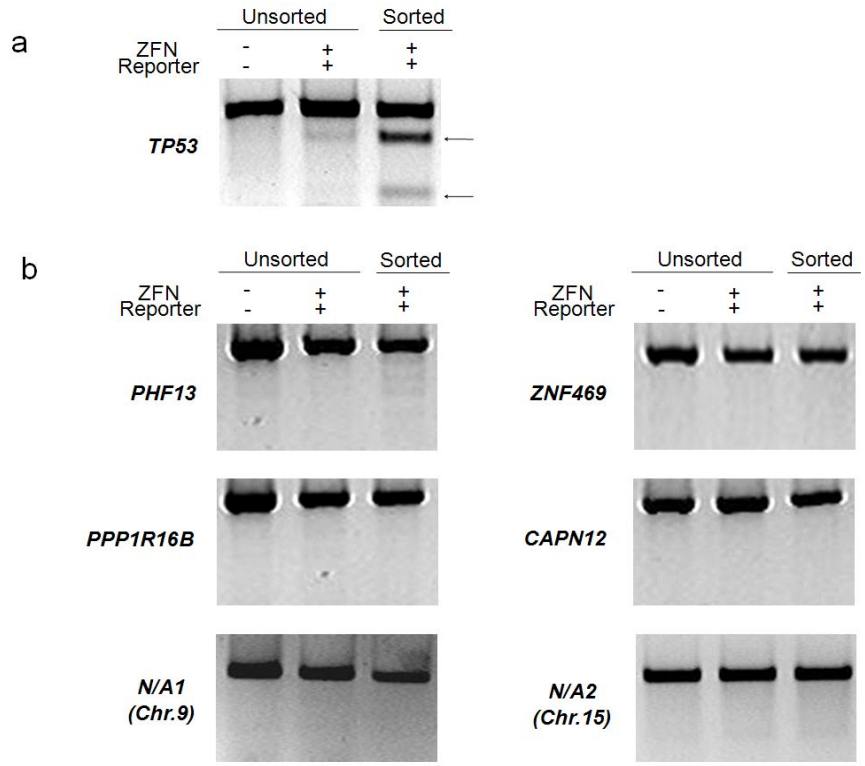


Figure 13. No off-target mutations in sorted and unsorted cells.

HEK293 cells were cotransfected with a reporter plasmid and plasmids encoding the ZFN pair that targets the *TP53* gene. Three days after transfection, cells were sorted using flow cytometry, and genomic DNA was isolated from both unsorted cells and sorted cells (RFP^+ GFP^+) and subjected to the T7E1 assay. No mutations were detected (assay sensitivity, ~0.5%) at potential off-target sites (listed in Table 7). An arrow indicates the expected position of DNA bands cleaved by T7E1. NA (not applicable) indicates an intergenic site. The primers used in this assay described in Table 8.

Table 7. Potential off-target sites of the *TP53*-targeting ZFN in the human genome

Rank	Score	Chromosome number	Gene name	Left half-site (5' to 3')	Mismatch	Right half-site (5' to 3')	Mismatch	Spacer (bp)	Homodimer or Heterodimer
Intended	1	17	*TP53*	*GGCGCGGACGCG*	*0*	*CATCTACAAGCA*	*0*	*5*	*Heterodimer*
1	*0.86*	*16*	*ZNF469*	*GGCGaGtACcCG*	*3*	*CATCaACAAGCA*	*1*	*5*	*Heterodimer*
2	*0.78*	*1*	*CAPN2*	*GcCGCGGACtCG*	*2*	*gATCTgCAAaCA*	*3*	*6*	*Heterodimer*
2	*0.78*	*20*	*PPP1R16B*	*GGCGgGGtCGCG*	*3*	*CAgCTACAAGgA*	*2*	*6*	*Heterodimer*
4	*0.69*	*9*	*NA*	*CATgTACAAGCA*	*1*	*CAcCcACAAGCA*	*2*	*6*	*Homodimer*
5	*0.68*	*15*	*NA*	*GGCGgGGACcCG*	*2*	*GGCGCGGgCGgG*	*2*	*6*	*Homodimer*

The top 5 potential off-target sites of *TP53*-targeting ZFN. Mismatching bases are shown in lowercase letters.

NA (not applicable) indicates an intergenic site.

Table 8. The primer sequences for *TP53*-ZFN off target analysis

Primer Name	Sequence (5' to 3')
ZNF469-F	CCTCGGAATCAGACCGCGATTC
ZNF469-R	CAACTGATAGCTTTGGGGTCAGG
CAPN2-F	GGTTCTCCAAGTGCAAGAAAGCG
CAPN2-R	CCACAAACTGTGCCTGTGGCC
PPP1R16B-F	CCAAGATCCCACGAGGTGAACTGG
PPP1R16B-R	CCCAGGTGTGTGCTAAGGAAGG
NA(Chr.9)-F	GGACCGTGCAGCAAGATGGAC
NA(Chr.9)-R	GAGGATATCGAGGGGGCCACGG
NA(Chr.15)-F	CCTCTTGGCTCCCGGGGTGGTAG
NA(Chr.15)-R	CGACTCCCTCCCCCCTCGCGTG

5. Determination of contributing factors in surrogate reporter system

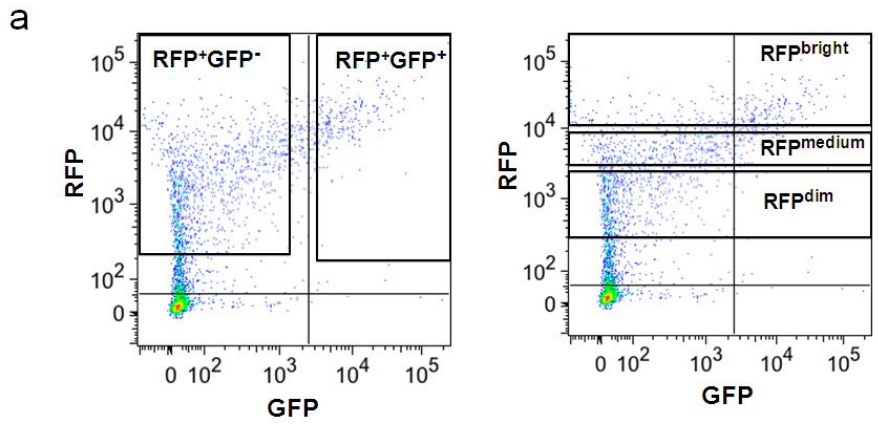
Two different factors may additively contribute to the enrichment of mutant cells by this reporter system: One is a co-transfection effect and the other is a surrogate reporter effect. These two factors can be distinguished by comparing mutation rates (or indels (%)) in RFP^{+} GFP^{-} cells (co-transfection effect alone) with those in RFP^{+} GFP^{+} cells (both effects). I observed only a few percent increase in mutation rates in RFP^{+} GFP^{-} cells as compared to those in unsorted cells (Table 9). In contrast, mutation rates in RFP^{+} GFP^{+} cells were much higher, indicating that the surrogate reporter effect is the major contributing factor for the enrichment of mutant cells. To investigate the potential mechanisms underlying the enrichment by the reporter further, we determined the mutation rates and nuclease levels in RFP^{dim}, RFP^{medium}, and RFP^{bright} cells as well as in RFP^{-} GFP^{+} and RFP^{+} GFP^{+} cells. The T7E1 assay revealed that genome-modified cells were enriched in the following order: RFP^{bright} (10% mutation), RFP^{medium} (6.3%), and RFP^{dim} (1.2%) (Figure 14), suggesting that high transfection efficiencies can lead to high mutation frequencies. However, the fold enrichment of gene-disrupted cells in the RFP^{bright} populations (7.7-fold as compared to unsorted cells) failed to reach that in RFP^{+} GFP^{+} cells (44% mutation, 34-fold), indicatingthat our reporter system is more useful than simple sorting based on transfection efficiency alone.

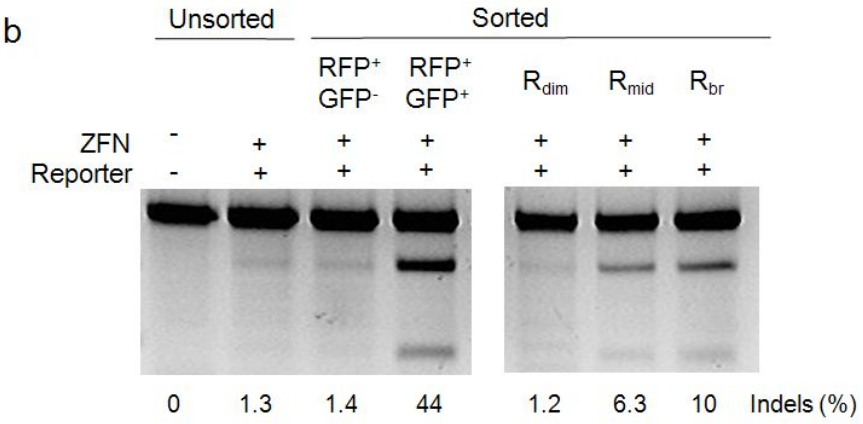
Furthermore, Western blotting showed that the nuclease levels correlated with mutation frequencies, suggesting that the high mutation rates in RFP^+ GFP^+ cells were, at least in part, mediated by the high nuclease concentrations.

Table 9. Comparison of mutation frequencies in sorted and unsorted cells

	Mutation rates (%)			Fold enrichment	
Target gene	Unsorted	Sorted (RFP^+ GFP^-)	Sorted (RFP^+ GFP^+)	Sorted (RFP^+ GFP^-)	Sorted (RFP^+ GFP^+)
TP53	2.8	4.8	37	1.7	13
CCR5 **(ZFN-224)**	12	16	69	1.3	5.8
CCR5 **(Z891)**	0.8	3.0	8.7	3.8	11
CCR5 **(TALEN)**	0.5	1.5	4.3	3.0	8.6
Thumpd3	0.5	2.4	46	4.8	92

Mutation frequencies are obtained from data shown in Figure 4b, 5b, 6b, 8b, and 9b.


a
RFP+GFP-
RFP+GFP+
RFP
GFP
RFPbright
RFPmedium
RFPdim


b
Unsorted
Sorted
RFP+ GFP-
RFP+ GFP+
Rdim
Rmid
Rbr
ZFN
Reporter
0 1.3 1.4 44 1.2 6.3 10 Indels (%)

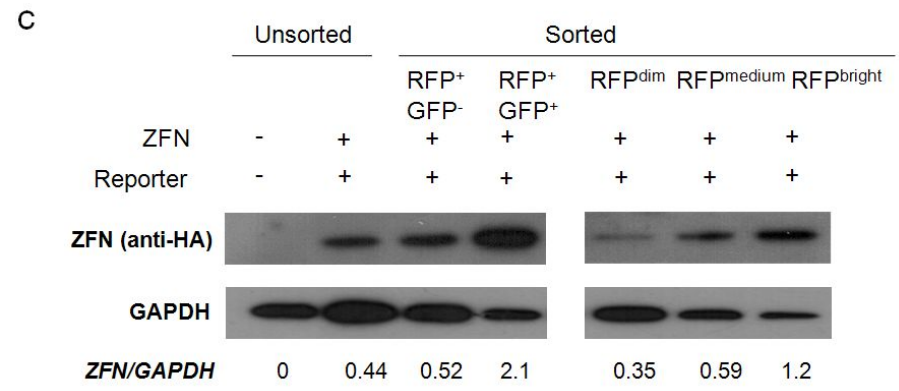

c
Unsorted
Sorted
RFP+ GFP-
RFP+ GFP+
RFPdim
RFPmedium
RFPbright
ZFN
Reporter
ZFN (anti-HA)
GAPDH
ZFN/GAPDH
0 0.44 0.52 2.1 0.35 0.59 1.2

Figure 14. Mutation frequencies and nuclease levels in sorted cells.

HEK293 cells were transfected with a reporter plasmid and plasmids encoding *TP53*-targeting ZFNs. Three days after transfection, RFP^{+} GFP^{-}, RFP^{+} GFP^{+}, RFP^{dim}, RFP^{medium}, and RFP^{bright} cells were sorted using flow cytometry and subjected to the T7E1 assay (b) and Western blotting (c). (a) Flow cytometry of the cells. Areas marked with red boxes indicate the gates used for sorting. (b) Mutation frequencies wereevaluated by T7E1 assay using the isolated genomic DNA. The bands indicated by the arrows represent the amplicon cut by T7E1; relative band density indicates ZFN activity. (c) The protein levels of ZFN tagged with HA were determined by Western blot using cell extracts. GAPDH was used as an internal control.

B. Enrichment of cells with nuclease-induced mutations by antibody-assisted separation.

1. Construction of reporter for magnetic-activated cell separation.

I previously described surrogate reporters that enable the efficient enrichment of cells containing nuclease-induced mutations via flow cytometry. This method is, however, limited by the availability of flow cytometers. Furthermore, sorted cells occasionally fail to form colonies after exposure to a strong laser and hydrostatic pressure

I devised reporters that express a truncated H-2Kk surface marker and eGFP only when target sequences in the reporter plasmids are cleaved and small insertions or deletions (indels) are generated via mutagenic non-homologous end-joining (Figure 15). This reporter system allows mutant cell enrichment without a flow cytometer, the cells which expressed surface marker by engineered nucleases can be magnetically separated using anti-H-2kk antibody conjugated to a magnetic bead. After magnetic separation, the selected cells which also expressed eGFP makers can be confirmed by using a flow cytometry or a fluorescent microscopy.

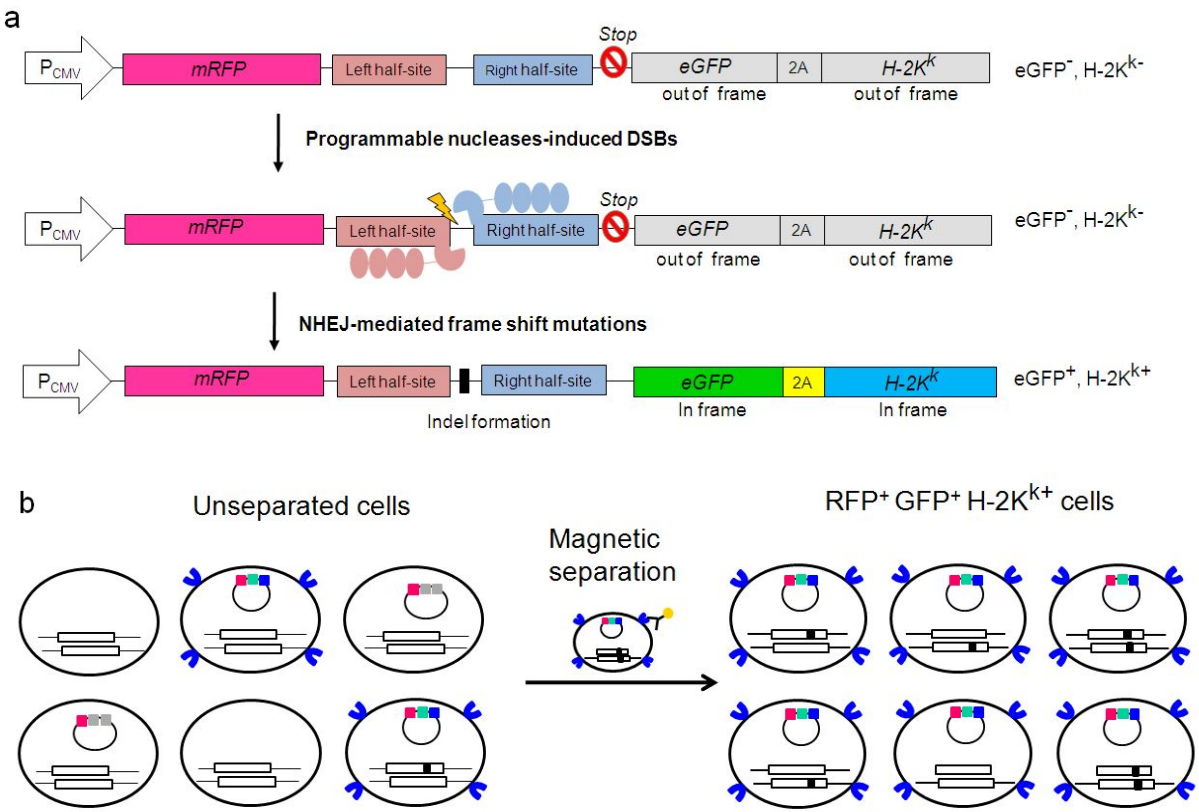

a
P CMV
mRFP
Left half-site
Right half-site
Stop
eGFP
2A
H-2K^k
out of frame
out of frame
eGFP-, H-2K^k-
Programmable nucleases-induced DSBs
NHEJ-mediated frame shift mutations
Indel formation
In frame
In frame
eGFP+, H-2K^k+
b
Unseparated cells
Magnetic separation
RFP+ GFP+ H-2K^k+ cells

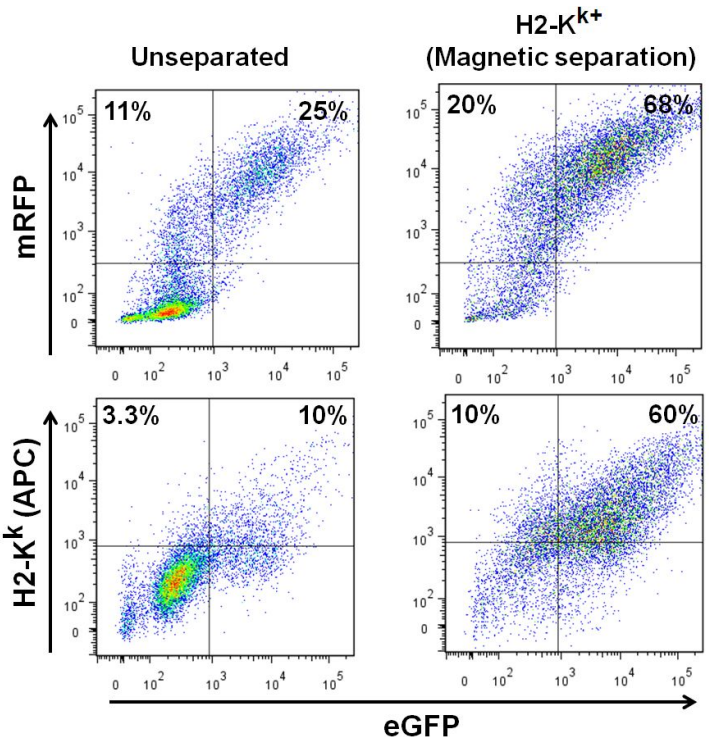
Figure 15. Enrichment of gene-edited cells using magnetic separation.

(a) The working mechanism of the H-2Kk reporter. In reporters, mRFP is constitutively expressed by the CMV promoter (P_{CMV}), whereas eGFP and H-2Kk are not expressed without the activity of engineered nucleases because their sequences are out of frame. If a double-strand break is introduced into the target sequence by engineered nucleases, the break is repaired by non-homologous end-joining (NHEJ), which often makes indels. Indel generation can cause frame shifts, making eGFP and H-2Kk in frame and leading to the expression of eGFP and H-2Kk. (b) A schematic depicting the enrichment of mutant cells using the H-2Kk reporter. H-2Kk-expressing cells can be magnetically separated using anti-H-2kk antibody conjugated to a magnetic bead. Mutant cells were enriched in this population of H-2kk-expressing cells. Reporter plasmids and chromosomal target loci are shown. Black spots represent mutations.

2. Enrichment of nuclease-induced mutant cells by magnetic separation

To determine whether this reporter can be used for the enrichment of mutant cells, I cotransfected plasmids encoding the *CCR5*-specific ZFN (Z891) and its reporter into HEK293 cells. Magnetic separation of cells that express H-2Kk (designated H-2Kk$^+$ cells) after labeling with anti-H-2Kk antibody led to the enrichment of eGFP$^+$ cells (figure 16). A T7 endonuclease I (T7E1) assay showed that the mutation frequency at the *CCR5* gene in H-2Kk$^+$ cells was 46%, 12-fold higher than that in unsorted cells (Figure 17a), demonstrating efficient enrichment of *CCR5*-disrupted cells. DNA sequencing analysis also confirmed the enrichment of mutant cells in the H-2Kk$^+$ cell population (Figure 17b).

a


Unseparated
H2-K^k+
(Magnetic separation)
11%
25%
20%
68%
3.3%
10%
10%
60%
mRFP
H2-K^k (APC)
eGFP

b

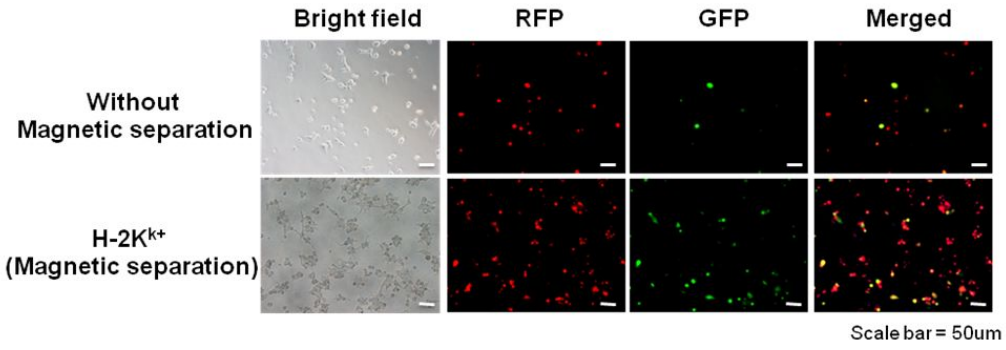
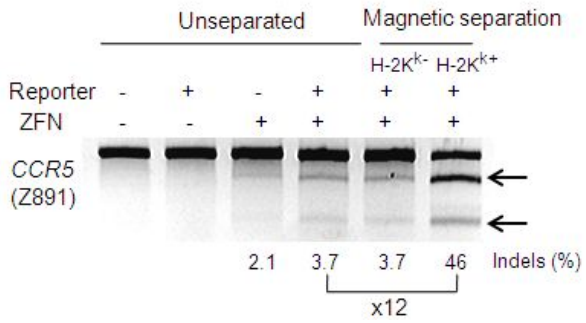

Bright field
RFP
GFP
Merged
Without
Magnetic separation
H-2K^k+
(Magnetic separation)
Scale bar = 50um

Figure 16. Enrichment of $RFP^{+}GFP+$ cells after magnetic separation.

Three days after a reporter plasmid and plasmids encoding *CCR5*-targeting ZFNs (Z891) were cotransfected into HEK293 cells, cells were labeled with magnetic bead-conjugated antibody against H-2Kk and magnetically separated. Eight hours after the separated cells were plated into culture dishes, cells were analyzed using flow cytometry (a) or observed using fluorescent microscopy (b). Scale bar = 50 mm.

a



b

Unseparated, 2.8% (2/71) mutated

```
TCAACCCCATCATCTATGCCTTTGTCGGGGAGAAGTCAG  (WT)
TCAACCCCATCATC-----CTTTGTCGGGGAGAAGTCAG  (x1)
TCAACCCCATCATCT--------GTCGGGGAGAAGTCAG  (x1)
```

Magnetically separated, 60% (6/10) mutated

```
TCAACCCCATCATCTATGCCTT----TGTCGGGGAGAAGTCAG  (WT)
TCAACCCCATCATCTATGCCTTcctTGTCGGGGAGAAGTCAG  (X1)
TCAACCCCATCATCTATGCCTT-------------GAAGTCAG  (X1)
TCAACCCCAT-----------------TGTCGGGGAGAAGTCAG  (X1)
TCAACCCCATC-----------------------GAAGTCAG  (X1)

TCAACCCCATCATCTATGCCTT 131bp insertion TGTCGGGGAGAAGTCAG (X1)
---------------------- 201bp deletion  TGTCGGGGAGAAGTCAG (X1)
```

Figure 17. Enrichment of mutant clones after magnetic separation.

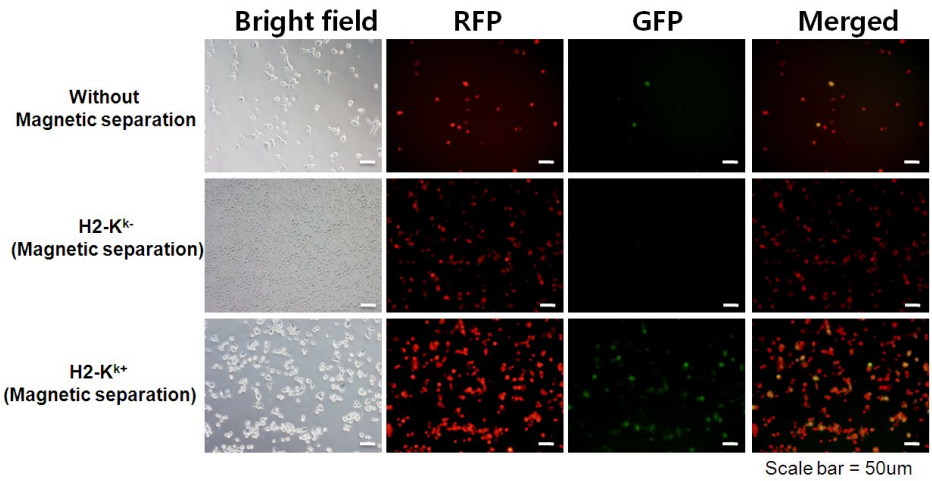
(a) ZFN-driven mutations detected by the T7E1 assay. Arrows indicate the expected positions of DNA bands cleaved by mismatch-sensitive T7E1. The numbers at the bottom of the gel indicate mutation frequencies measured by band intensities. (b) DNA sequences of the *CCR5* gene targeted by the ZFN. ZFN recognition sites are underlined. Dashes indicate deleted bases, and small bold letters indicate inserted bases. In cases in which a mutation was detected more than once, the number of occurrences is shown in parentheses.

Next, I confirmed whether this reporter system is portable to other ZFNs and TALENs. Three days after a reporter plasmid and plasmids encoding ZFNs or TALENs that target *TP53* or *BRCA1* were cotransfected into HEK293 cells, cells were labeled with magnetic bead-conjugated antibody against H-2Kk and magnetically separated. The eGFP-expressed cells were enriched in the separated cells (Figure 18a, Figure 19a). This system enabled at least 17-fold enrichment of cells containing mutations induced by *TP53*-ZFN (Figure 18b) and *BRCA1*-TALEN (Figure 19b).

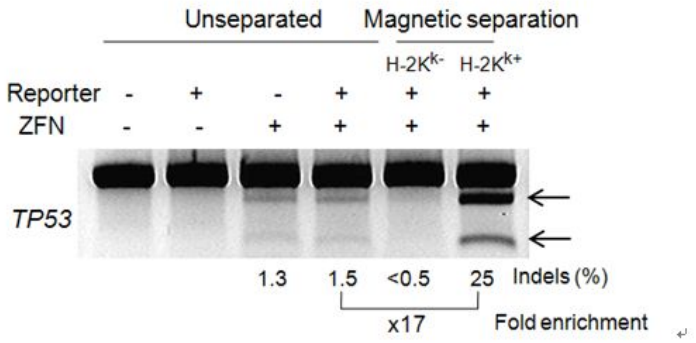
I next tested whether this reporter system enables to enrich ZFN-mediated *CD81*-disrupted cells in a Huh7.5 cell line. After magnetical separation, *CD81*-disrupted cells were enriched at least 17-fold (Figure 20, Table 10). A *CD81* gene is one of the most important receptors which related to Hepatitis C Virus (HCV) cell entry, interfering with the interaction between HCV *E2* protein and *CD81* receptor could be used as an anti-HCV therapy. This reporter system is able to enhance the efficacy of the *CD-81* targeting ZFN by enriching *CD81*-knockout cell if used in gene therapy.

Four different target sites in the human genome were disrupted by using ZFNs or TALENs and the reporter system allows to enable enrichment of nuclease-induced mutant cells (Figure 21, Table 10).

a



b



c

MACS Sorted 21 % (= 8/39) mutated

```
GGCACCCGCGTCCGCGCC-----ATGGCCATCTACAAGCAGTCAC (WT)
GGCACCCGCGTCCGCGCCatgg-ATGGCCATCTACAAGCAGTCAC (x3)
GGCACCCGCGTCCGCGCC-----ATGGC--TCTACAAGCAGTCAC (x1)
GGCACCCGCGTCCGCGCCatggcATGGCCATCTACAAGCAGTCAC (x1)
GGCACCCGCGTCCGCGCCatggaATGGCCATCTACAAGCAGTCAC (x1)
GGCACCCGCGTCCGCGCCat---ATGGCCATCTACAAGCAGTCAC (x1)

         ---      210bp deletion   ---         (x1)
```

F**igure 18. Enrichment of *TP53* gene disrupted cells after magnetic separation.** (a) Eight hours after the separated cells were plated into culture dishes, cells were observed using fluorescent microscopy. Scale bar = 50 mm. (b) T7E1 assays were performed using genomic DNA isolated from the magnetically-separated cells. Arrows indicate the expected positions of DNA bands cleaved by mismatch-sensitive T7E1. The numbers at the bottom of the gel indicate mutation frequencies measured by band intensities. (c) DNA sequences of the *TP53* gene targeted by the ZFN. ZFN recognition sites are underlined. Dashes indicate deleted bases, and small bold letters indicate inserted bases. In cases in which a mutation was detected more than once, the number of occurrences is shown in parentheses.

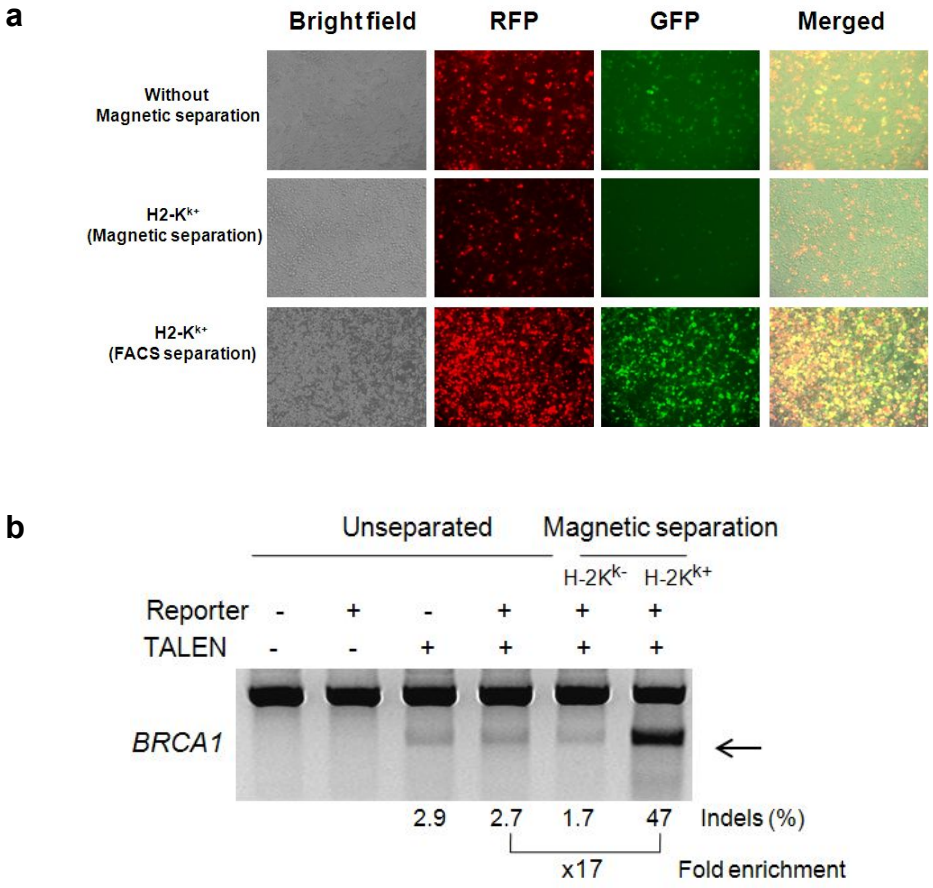


Figure 19. Enrichment of cells with TALEN-driven mutations after magnetic separation. (a) Eight hours after the separated cells were plated into culture dishes, cells were observed using fluorescent microscopy. Scale bar = 50 mm. (b) T7E1 assays were performed using genomic DNA isolated from the magnetically-separated cells. Arrows denote the expected position of DNA bands cleaved by T7E1.

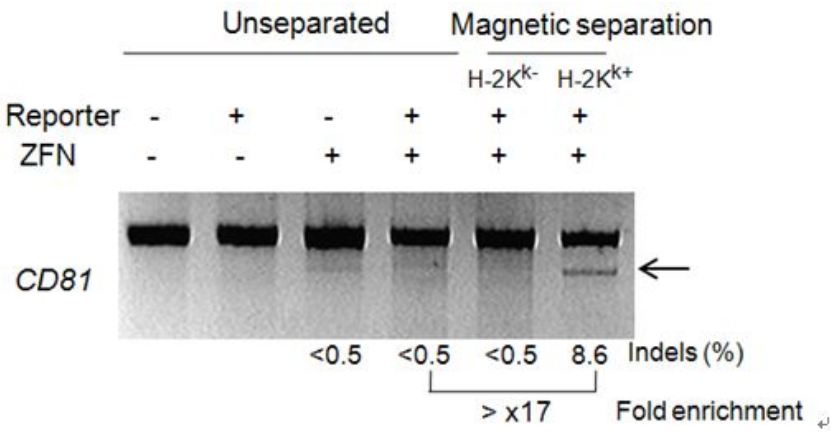


Figure 20. Enrichment of *CD81* gene disrupted cells after magnetic separation. Three days after a reporter plasmid and plasmids encoding *CD81* ZFNs were cotransfected into Huh7.5 cell line, cells were labeled with magnetic bead-conjugated antibody against H-2Kk and magnetically separated. T7E1 assays were performed using genomic DNA isolated from the magnetically-separated cells. Arrows denote the expected position of DNA bands cleaved by T7E1.

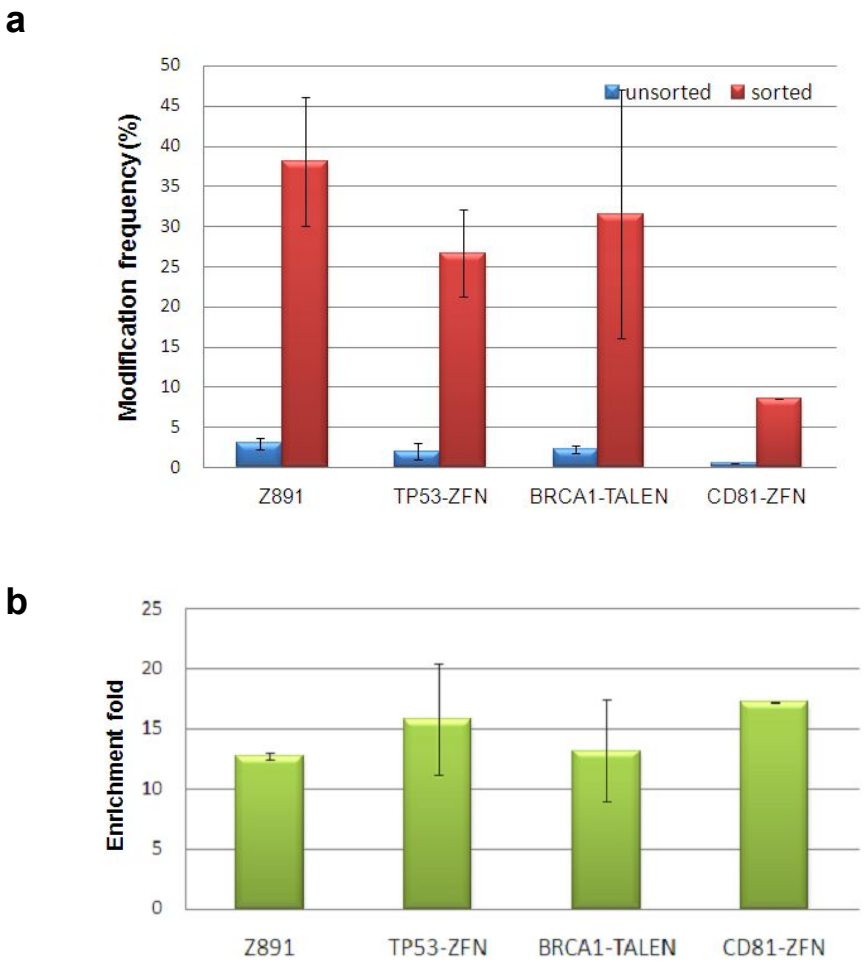


Figure 21. Enrichment of ZFN/TALEN-mediated mutation frequencies after magnetic separation. (a) Modification frequencies at the target locus of each ZFNs or TALENs. Blue bars indicate the mutation rates of unseparated cells and red bars indicate separated cells. (b) Enrichment fold of mutation frequencies in separated cells as compared to unseparated cells. This data represent an average of the duplicate independent experiments and the standard error of the mean is shown.

Table 10. List of mutation frequencies after magnetic separation

Target locus		Mutation rates (%)		Fold enrichment
		Unsorted	Sorted	
TP53	ZFN	2.0 ± 1.1	27 ± 5.4	16 ± 4.6
CCR5	Z891	3.0 ± 0.70	38 ± 8.0	13 ± 0.31
BRCA1	TALEN	2.3 ± 0.45	32 ± 16	13 ± 4.3
CD81	ZFN	0.5 ± n.d	8.6 ± n.d	17 ± n.d

3. Isolation of mutant clones after magnetic separation

I then analyzed clonal populations of cells before and after magnetic cell separation. One of 88 clonal populations of cells contained ZFN-induced mutations at the target site of Z891 before cell sorting. In contrast, three clones out of 10 H-2Kk postive clonal populations contained mutations, demonstrating 14-fold enrichment of mutant cells (Figure 22).

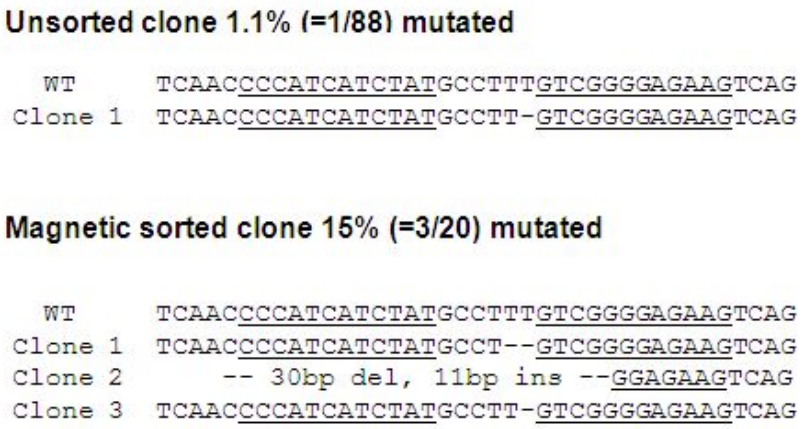


Figure 22. Clonal analysis after magnetic separation. Three days after a reporter plasmid and Z891 plasmids cotransfected into HEK293 cells, cells were labeled with magnetic bead-conjugated antibody against H-2Kk and magnetically separated. The selected or unselected (control) cells were plated at a density of 0.25cells/well in 96well plates, and the clonal colonies were manually picked 10 days after plating.

C. Enrichment of cells with nuclease-induced mutations by hygromycin selection based surrogate reporter system

1. Construction of repoter for drug selection

I also developed reporters that express a hygromycin-resistance protein (HygroR)-GFP fusion only when the target sequences are cleaved by nucleases (Figure 23).

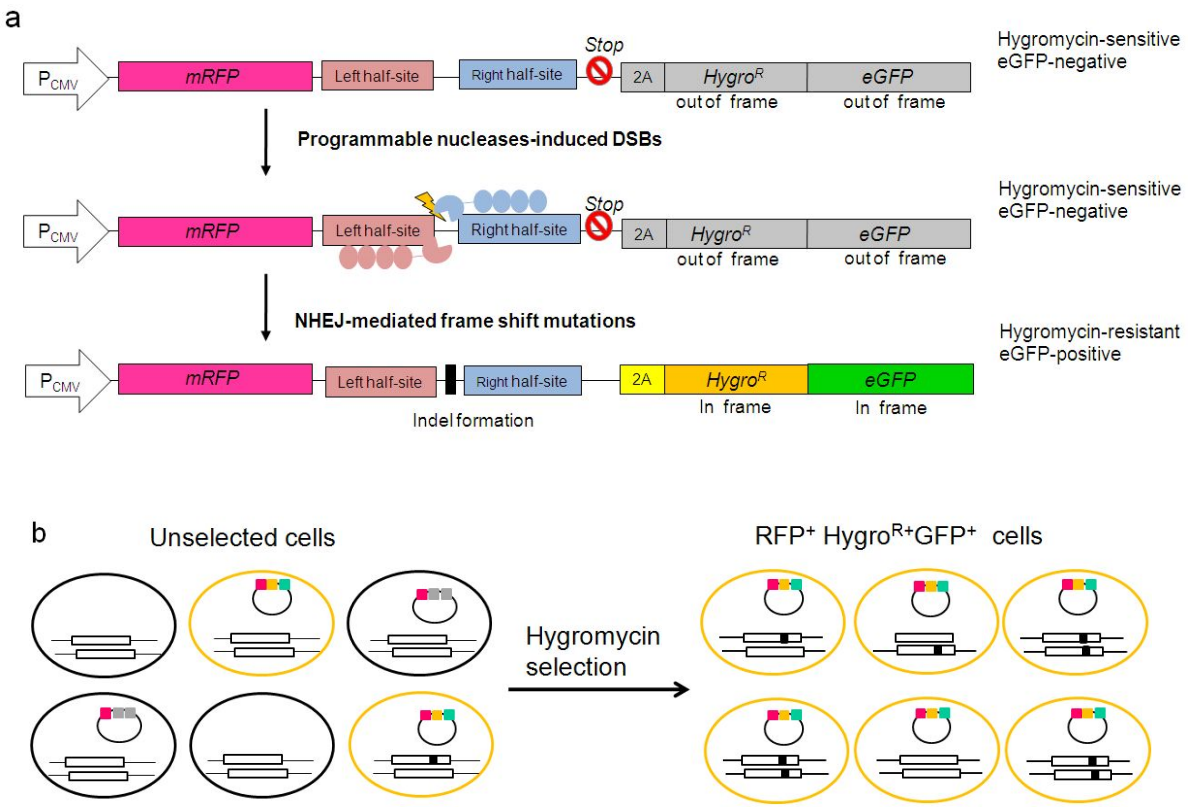
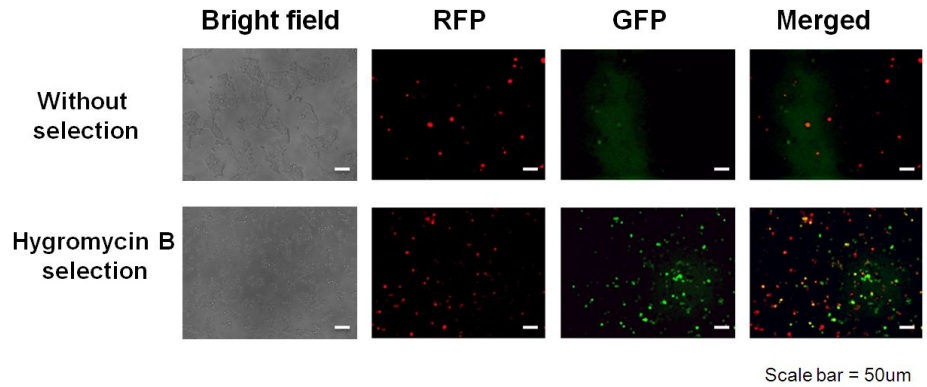

a
Stop
Hygromycin-sensitive
eGFP-negative
P CMV
mRFP
Left half-site
Right half-site
2A
Hygro R
eGFP
out of frame
out of frame
Programmable nucleases-induced DSBs
Hygromycin-sensitive
eGFP-negative
NHEJ-mediated frame shift mutations
Hygromycin-resistant
eGFP-positive
Indel formation
In frame
In frame
b
Unselected cells
RFP+ Hygro R+GFP+ cells
Hygromycin
selection

Figure 23. Enrichment of gene-editedcells using hygromycin selection.
(a) The working mechanism of the hygromycin reporter. mRFP is constitutively expressed by the CMV promoter (P_{CMV}), whereas the hygroR-GFP fusion protein is not expressed in the absence of engineered nucleases because their sequences are out of frame. If a double-strand break is introduced into the target sequence by engineered nucleases, the break is repaired by NHEJ, which often makes indels. Indel generation can cause frame shifts, rendering hygroR-GFP in frame and expressed. (b) A schematic depicting the enrichment of mutant cells using the hygromycin reporters. Hygro-GFP fusion protein-expressing cells can be selected using hygromycin treatment. Mutant cells were enriched in this population of hygro-GFP-expressing cells. Reporter plasmids and chromosomal target loci are shown. Black spots represent mutations.

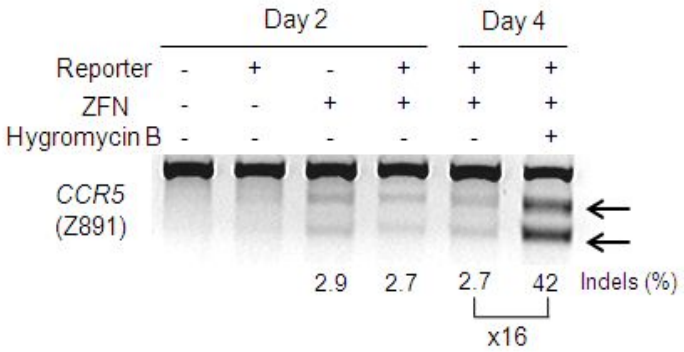
2. Enrichment of nuclease-induced mutant cells by drug selection

Hygromycin treatment after transfection of Z891-encoding plasmids and its reporter into HEK293 cells led to the enrichment of GFP^+ cells (Figure 24a). The mutation frequency at the *CCR5* gene in the hygromycin-resistant cells was 42%, 16 fold higher than that in unselected cells (Figure 24b). DNA sequencing analysis corroborated this efficient enrichment of mutant cells (Figure 24c). Furthermore, this reporter system allowed 15-fold enrichment of mutant cells induced by a *BRCA1*targeting TALEN (Figure 25, Table 11).

a



b



c

Unselected, 4.6% (4/87) mutated

```
TCAACCCCATCATCTATGCCTTTGTCGGGGAGAAGTCAG  (WT)
TCAACCCCAT----------GTCGGGGAGAAGTCAG  (x1)

---------- 221bp deletion CTTTGTCGGGGAGAAGTCAG (x3)
```

Hygromycin selected cells, 39% (9/23) mutated

```
TCAACCCCATCATCTATGCCTTTGTCGGGGAGAAGTCAG  (WT)
TCAACCCCATCATCTATGCCTT----GGGGAGAAGTCAG  (x1)
TCAACCCCATCATCTAT------------------CAG  (x1)
TCAACCCCATCA------CCTTTGTCGGGGAGAAGTCAG  (x1)
TCAACCCCATC-------CCTTTGTCGGGGAGAAGTCAG  (x1)
TCAACCCCATC---------------GGGGAGAAGTCAG  (x1)

TCAACCCCATCATC----cctttgtcCCTTTGTCGGGGAGAAGTCAG  (x1)
---------- 221bp deletion CTTTGTCGGGGAGAAGTCAG (x2)
---------- 106bp deletion CTTTGTCGGGGAGAAGTCAG (x1)
```

Figure 24. Enrichment of *CCR5*-disrupted cells after hygromycin selection. (a) Two days after a reporter plasmid and plasmids encoding *CCR5*-targeting ZFN plasmids (Z891) were cotransfected into HEK293 cells, cells were cultured either in the absenceor presence of 2 mg/ml hygromycin for an additional two days and observed using fluorescent microscopy. Scale bar = 50 mm. (b) ZFN-driven mutations detected by the T7E1 assay. Arrows indicate the expected positions of DNA bands cleaved by mismatch-sensitive T7E1. The numbers at the bottom of the gelindicate mutation frequencies measured by band intensities. (c) PCR-amplified DNA from cells transfected with Z891-encoding plasmids and its reporter plasmid was cloned and sequenced. The mutation frequency in the hygromycin-selected populations was 39%, 8.4-fold higher than that in unselected cells. ZFN recognition sites are underlined. Dashes indicate deleted bases, and small bold letters indicate inserted bases. In cases in which a mutation was detected more than once, the number of occurrences is shown in parentheses.

Two different target sites in the human genome were disrupted by using ZFNs or TALENs and the reporter system allows to enable enrichment of nuclease-induced mutant cells (Figure 26, Table 11).

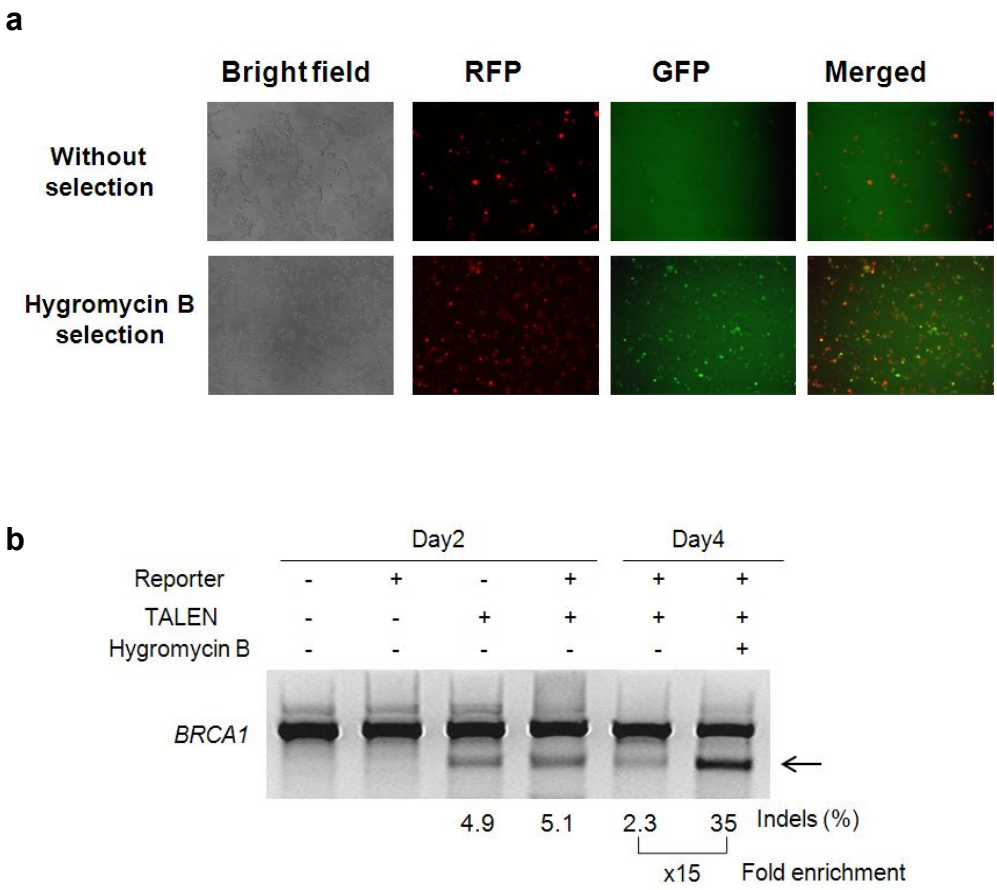


Figure 25. Enrichment of cells with TALEN-driven mutations after hygromycin selection. (a) Two days after a reporter plasmid and plasmids encoding *BRCA1*-targeting TALEN plasmids were cotransfected into HEK293 cells, cells were cultured either in the absence or presence of 2 mg/ml hygromycin for an additional two days and observed using fluorescent microscopy. Scale bar = 50 mm. (b) T7E1 assays were performed using genomic DNA isolated from the magnetically-separated cells. Arrows denote the expected position of DNA bands cleaved by T7E1.

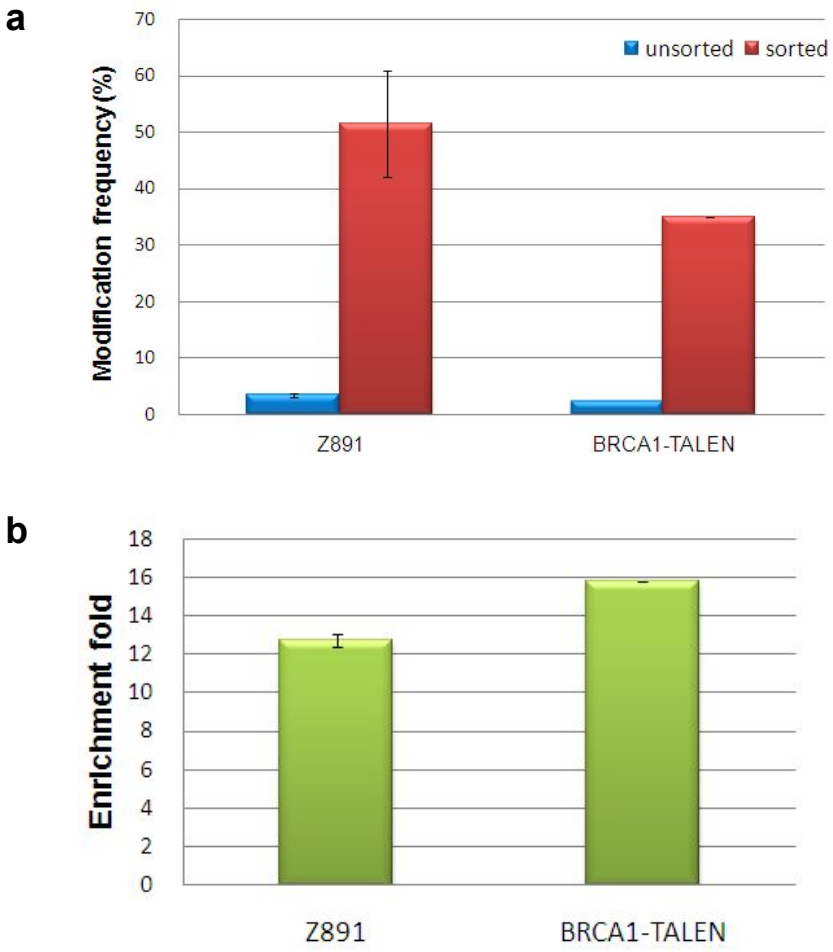


Figure 26. Enrichment of ZFN/TALEN-mediated mutation frequencies after hygromycin selection. (a) Modification frequencies at the target locus of each ZFNs or TALENs. Blue bars indicate the mutation rates of unselected cells and red bars indicate selected cells. (b) Enrichment fold of mutation frequencies in selected cells as compared to unselected cells. This data represent an average of the duplicate independent experiments and the standard error of the mean is shown.

Table 11. List of mutation frequencies after hygromycin selection

Target locus		Mutation rates (%)		Fold enrichment
		Unselected	Selected	
CCR5	Z891	3.4 ± 0.7	52 ± 9.5	15 ± 0.34
BRCA1	TALEN	2.3 ± n.d	35 ± n.d	16 ± n.d

3. Isolation of mutant clones by hygromycin selection

When I analyzed single cell-derived colonies, the frequency of mutant colonies was 39% (11/28) in the drug-treated group and 1.8% (1/56) in the untreated group, demonstrating 26-fold enrichment of mutant cells (Figure 27, Figure 28). Importantly, 5 of 11 clones in the hygromycin-treated group but none in the untreated group contained biallelic or triallelic modifications (Figure 28).

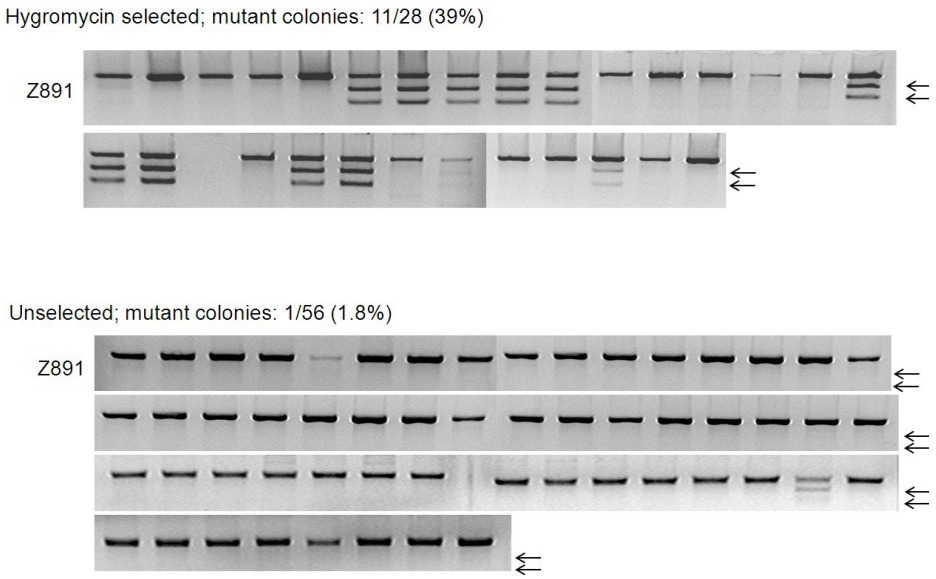


Figure 27. Enrichment of clonal populations of cells with ZFN-driven mutations using the hygromycin reporter. Two days after a reporter plasmid and plasmids encoding ZFN (Z891) were cotransfected into HEK293 cells, hygromycin selection was performedby culturing the cells in the presence of 2 mg/ml hygromycin B for two days. The selected or unselected (control) cells were plated at a density of 3,000 cells/100mm dish, and the clonal colonies were manually picked 10 days after plating. T7E1 assays were performed using genomic DNA isolated from the colonies. Arrows indicate the expected position of DNA bands cleaved by T7E1.

Unselected, 1.8% (1/56) mutated

```
   WT      TCAACCCCATCATCTATGCC----TTTGTCGGGGAGAAGTCAG
clone 1a   TCAACCCCATCATCTATGCCttccTTTGTCGGGGAGAAGTCAG
clone 1b   TCAACCCCATCATCTATG----------TCGGGGAGAAGTCAG
```

Hygromycine selected, 39% (11/28) mutated

```
   WT      TCAACCCCATCATCTATGCCTT------TGTCGGGGAGAAGTCAG
```

Singe allele mutation

```
clone 1    TCAACCCCATCATCTATGCCTT-----tTGTCGGGGAGAAGTCAG
clone 2    TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 4    TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 9    TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 10   TCAACCCCATCATCTATGCCTTgtccttTGTCGGGGAGAAGTCAG
clone 11   TCAACCCCATCATCTA----TT------TGTCGGGGAGAAGTCAG
```

BI- or tri-allele mutation

```
clone 3a   TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 3b   TCAACCCCATCATC-----CTT------TGTCGGGGAGAAGTCAG
clone 5a   TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 5b   TCAACCCCATCATC-----CTT------TGTCGGGGAGAAGTCAG
clone 6a   TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 6b   TCAACCCCATCATC-----CTT------TGTCGGGGAGAAGTCAG
clone 7a   TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 7b   TCAACCCCATCATC-----CTT------TGTCGGGGAGAAGTCAG
clone 8a   TCAACCCCATCATCTATGCCTT--ccttTGTCGGGGAGAAGTCAG
clone 8b   TCAACCCCATCATC-----CTT------TGTCGGGGAGAAGTCAG
```

Figure 28. Sequence analysis of clonal populations of cells with ZFN-driven mutations. PCR products from 56 unsorted cells and 28 sorted cells were cloned and sequenced. Single mutation was detected in unsorted cells. In contrast, 11 of 28 mutations were detected in sorted cells. "1a and 1b" indicate DNA sequences that resulted from bialleleic mutations in a single clone.

IV. Discussion

Continual efforts have brought the ZFN and TALEN technology into focus to provide a faster and more efficient method of genome engineering at selected target sites in diverse eukaryotic species. Thus far, the most successful applications of the engineered nuclease technology have been reported for the introduction of targeted modifications intothe genome to create transgenic species, including *C. elegans* (Morton et al., 2006; Wood et al., 2012), *Drosophila melanogaster* (Bibikova et al., 2002), *Xenopus tropicalis* (Young et al., 2011), silkworms (Takasu et al., 2010), zebrafish (Doyon et al., 2008; Meng et al., 2008; Moore et al., 2012), sea urchins (Ochiai et al., 2010), *Arabidopsis thaliana* (Osakabe et al., 2010; Zhang et al., 2010), tobacco (Cai et al., 2009; Maeder et al., 2008; Townsend et al., 2009), corn (Shukla et al., 2009), soybean (Curtin et al., 2011), rice (Li et al., 2012), mice (Carbery et al., 2010), rats (Geurts et al., 2009), pigs (Hauschild et al., 2011), rabbits (Flisikowska et al., 2011), and cows (Yu et al., 2011).

The ability to target mutations and other genetic modifications using ZFN/TALEN-mediated gene editing in these organisms represents a significant technological breakthrough. Engineered nucleases generate mutations by the well-characterized mechanism of NHEJ. When a DSB is introduced, most cell types will attempt to repair the damage by rejoining the broken DNA strands. This can result in the perfect repair of the junction. However, frequently, the ends of the DNA are

modified before ligation, resulting in small-to-moderately sized insertions or deletions (indels) at the breakpoint. The frequency and variety of such indels depends on a number of factors, such as the cell type, various pathways competing to resolve the break, cell cycle status, and experimental conditions. The factors that regulate DNA repair, which have been well characterized, could therefore potentially be targeted to drive specific repair pathways.

The ZFN/TALEN-stimulated NHEJ approach to making gene knock-outs for the study of gene function in animals and plants could also be used to knock out genes for therapeutic purposes in humans. Potential therapeutic knock-out targets include the genomes of infectious agents, such as tuberculosis or HIV cellular genes required by infectious agents, such as the *CCR5* co-receptor for HIV dominant negative mutant alleles that cause disease, such as in Huntington's disease or genes required for disease progression, such as the multi drug resistance gene (MDR1) in cancer. Competing methodologies in this area of gene therapy typically target RNA molecules, such as miRNAs. Since many RNA transcripts are made from DNA of the corresponding gene, targeting DNA genes using engineered nucleases should be more efficient. In addition, the engineered nuclease-directed modifications are permanent and heritable, and therefore should not require periodic re-administration or life-long expression of the therapeutic transgene.

More recently, a series of key publications emphasized the significant therapeutic potential of ZFNs, placing this technology into

the spotlight in the field of gene-based therapy. Gene-corrected, induced pluripotent stem cells derived from patients hold great potential for the treatment of human genetic diseases, the isolation of which requires *ex vivo* expansion. Potential therapeutic gene correction and insertion targets include recessive monogenic genetic disorders, among which fragile-X, cystic fibrosis, and Duchenne muscular dystrophy are the most common. Early-stage investigations of potential ZFN-based therapies have been reported for X-linked SCID, sickle cell anemia, cystic fibrosis, and myotonic dystrophy (Porteus, 2006).

In the past, the construction of the ZFNs to disrupt the target site may have been time-consuming, when complex selection strategies were necessary to produce ZFNs (Urnov et al., 2005). The availability of alternative platforms based on pre-assembled two-finger archives (Kim et al., 2010; Sander et al., 2010) reduces the time-frame for generating functional ZFNs to a few weeks, with little hands-on time. Moreover, alternative designer platforms based on TALENs (Miller et al., 2010) will further reduce the production time needed to create a functional nuclease to about a week (Cermak et al., 2011; Zhang et al., 2011; Morbitzer et al., 2011; Reyon et al., 2012). Interestingly, a side-by-side comparison of the ZFNs and TALENs suggests that the latter may be more specific and less cytotoxic (Mussolino et al., 2011).

ZFN-associated cytotoxicity was correlated to the ZFN dose and the type of nuclease variant employed. The use of obligate heterodimeric ZFN variants has been shown to greatly reduce ZFN-associated toxicity in cultured cells (Szczepek et al., 2007; Miller

et al., 2007; Sollu et al., 2010; Ramalingam et al., 2011; Doyan et al. 2011) and in zebrafish embryos (Gupta et al., 2011). Moreover, it has been shown that obligate heterodimeric ZFNs are more specific than their counterparts, which contain a wild-type FokI nuclease domain (Gabriel et al., 2011). Although the use of obligate heterodimeric ZFNs reduces off-target cleavage, it cannot be fully prevented (Gabriel et al., 2011; Pattanayak et al., 2011). Recent reports demonstrate that the zinc-finger nickase, which produces single-strand breaks (SSBs) instead of double-strand breaks (DSBs), allowed site-specific genome modifications to take place only at the on-target site, without the induction of unwanted indels (Kim et al., 2012; Ramirez et al., 2012; Wang et al., 2012).

Many of the novel approaches were found to increase the gene targeting frequency considerably, all of which were still dependent on either positive selection markers to enrich the targeted cells, or on the screening of large numbers of clones. At present, the methods available for gene targeting rely on positive selection to isolate rare clones that have undergone homologous recombination. To remove the unwanted selection cassettes, Cre/*loxP* or Flp/*FRT* recombination systems are used, which leave behind single *loxP* or *FRT* sites. These small ectopic sequences have the potential to interfere with transcriptional regulatory elements of surrounding genes, most of which have not been fully characterized in the human genome.

The surrogate reporter system has several advantages. First, the reporter reflects the activity of programmable nucleases without affecting

their activity and is compatible with other methods of enhancing the activity of the ZFNs (Doyon et al., 2010; Doyon et al., 2011). Second, the surrogate system is noninvasive, allowing additional cycles of transfection and cell sorting to further enrich the populations of mutant cells. Bi-allelic knockout cells can readily be obtained by repeating this procedure. In addition, gene-modified cells sorted by flow cytometry or isolated under a fluorescent microscope are alive and suitable for subsequent experiments, such as somatic cell nuclear transfer or the derivation of induced pluripotent stem cells. Third, because the surrogate reporter is delivered as an episomal plasmid, which disappears together with the ZFN plasmids after one or two weeks of cell division in culture, the genome as a whole, other than the ZFN target site, remains intact. The disappearance of reporter plasmids can easily be verified using fluorescent microscopy or flow cytometry. Thus, gene-modified cells enriched using the proposed episomal reporter system might be appropriate for use in gene or cell therapy. Enrichment of mutant cells through these new reporter systems was as efficient as using flow cytometry (Table 12). The characteristics of the three reporter systems have been summarized (Table 13), so that researchers will be able to choose appropriate reporters, depending on their experimental conditions. Given that the ZFNs and TALENs are used in various research environments, the proposed three reporters will facilitate the use of engineered nucleases in a wider range of biomedical research.

In summary, the novel surrogate reporter system reported herein

allows simple and non-invasive indirect determination of the activity of programmable nucleases in any cell type, enabling the efficient enrichment of genome-modified cells. I propose that this surrogate reporter system will greatly facilitate the application of programmable nucleases in gene therapy and biotechnology, as well as in basic research.

Table 12. Efficiencies of mutant cell enrichment via different reporter systems

Target gene	Mutation frequency (%)		Fold enrichment	Enrichment method
	Before enrichment	After enrichment		
CCR5 (Z891 ZFN)	0.8	8.7	11	Flow cytometry
	3.7	46	12	Magnetic separation
	2.7	42	16	Hygromycin selection
TP53 (ZFN)	2.8	37	13	Flow cytometry
	1.5	25	17	Magnetic separation
BRCA1 (TALEN)	2.7	47	17	Magnetic separation
	2.3	35	15	Hygromycin selection

Table 13. Comparison of enrichment methods

Enrichment method	Flow cytometry	Magnetic separation	Hygromycin selection
Required machines or instruments	Flow cytometers	Magnetic instruments	None
Times required for the enrichment process	Several hours	Several hours	Several days
Optimization of the enrichment process	Usually not necessary	Usually not necessary	Hygromycin concentration and exposure time need to be optimized for each cell type

V. References

Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J.K. and Carroll, D. 2006. Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics,* 172**:** 2391-2403.

Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol,* 21: 289-291.

Bibikova, M., Golic, M., Golic, K.G. and Carroll, D. 2002, Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics,* 161: 1169-1175.

Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V.M., Gregory, P., Urnov, F., Weinstock, D.M. and Jasin, M. 2009. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci USA,* 106: 10620-10625.

Cai, C.Q., Doyon, Y., Ainley, W.M., Miller, J.C., Dekelver, R.C., Moehle. E.A., Rock, J.M., Lee, Y.L., Garrison, R., Schulenberg, L., Blue, R., Worden, A., Baker, L., Faraji, F., Zhang, L., Holmes, M.C., Rebar, E.J., Collingwood, T.N., Rubin-Wilson, B., Gregory, P.D., Urnov, F.D. and Petolino, J.F. 2009. Targeted transgene integration in plant cells using designed zinc-finger nucleases. *Plant Mol Biol,* 69: 699-709.

Capecchi, M.R. 1989. Altering the genome by homologous recombination. *Science,* 244: 1288-1292.

Carbery, I.D., J,i, D., Harrington, A., Brown, V., Weinstein E.J., Liaw L. and Cui, X. 2010. Targeted genome modification in mice using zinc-finger nucleases. *Genetics,* 186: 451-459.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res,* 39: e82.

Cradick, T.J., Ambrosini,G., Iseli, C., Bucher, P. and McCaffrey, A.P. 2011. ZFN-site searches genomes for zinc-finger nuclease target sites and off-target sites. *BMC Bioinformatics*, 13: 152.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., Reyon, D., Dahlborg, E.J., Goodwin, M.J., Coffman, A.P., Dobbs, D., Joung, J.K., Voytas, D.F. and Stupar, R.M. 2011. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol,* 156: 466-473.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D. and Amacher, S.L. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol,* 26: 702-708.

Doyon, Y., Choi, V.M., Xia, D.F., Vo, T.D., Gregory, P.D. and Holmes, M.C. 2010. Transient cold shock enhances zinc-finger nuclease-mediated gene disruption. *Nat Methods,* 7: 459-460.

Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia,

D.F., Miller, J.C., Urnov, F.D., Gregory, P.D. and Holmes, M.C. 2011. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods*, 8: 74-79.

Flisikowska, T., Thorey, I.S., Offner, S., Ros, F., Lifke, V., Zeitler, B., Rottmann, O., Vincent, A., Zhang, L., Jenkins, S., Niersbach, H., Kind, A.J., Gregory, P.D., Schnieke, A.E. and Platzer, J. 2011. Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc-finger nucleases. *PLoS ONE*, 6: e21045.

Foley, J.E., Yeh, J.R., Maeder, M.L., Reyon, D., Sander, J.D., Peterson R.T. and Joung J.K. 2009. Rapid mutation of endogenous zebrafish genes using zinc-finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS ONE,* 4: e4348.

Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., Holmes, M.C., Gregory, P.D., Glimm, H., Schmidt, M., Naldini, L. and von Kalle, C. 2011. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol*, 29: 816-823.

Geurts, A.M., Cost, G.J., Freyvert, Y., Zeitler, B., Miller, J.C., Choi, V.M., Jenkins, S.S., Wood, A., Cui, X., Meng, X., Vincent, A., Lam, S., Michalkiewicz, M., Schilling, R., Foeckler, J., Kalloway, S., Weiler, H., Ménoret, S., Anegon, I., Davis, G.D., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., Jacob, H.J.

and Buelow, R. 2009. Knockout rats via embryo microinjection of zinc-fingernucleases. *Science,* 325: 433.

Goate, A. 2006. Segregation of a missense mutation in the amyloid beta-protein precursor gene with familial Alzheimer's disease. *J Alzheimers Dis,* 9: 341-347.

Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. and Wolfe, S.A. 2011. zinc-finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc-finger nucleases. *Nucleic Acids Res,* 39: 381-392.

Hauschild, J., Petersen, B., Santiago, Y., Queisser, A.L., Carnwath, J.W., Lucas-Hahn, A., Zhang, L., Meng, X., Gregory, P.D., Schwinzer, R., Cost, G.J. and Niemann, H. 2011. Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc Natl Acad Sci USA,* 108: 12013-12017.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller J.C., Zeitler, B., Cherone, J.M. Meng, X., Hinkley, S.J., Rebar, E.J., Gregory, P.D., Urnov, F.D. and Jaenisch, R. 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol,* 29: 731-734.

Jasin, M. 1996. Genetic manipulation of genomes with rare cutting endonucleases. *Trends Genet,* 12: 224-228.

Kim, E., Kim, S., Kim, D.H., Choi, B.S., Choi, I.Y. and Kim, J.S. 2012. Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res,* doi:10.1101/gr.138792.112.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. and Kim, J.S. 2009. Targeted genome editing with zinc-finger nucleases constructed via modular assembly. *Genome Res,* 19: 1279-1288.

Kim, H.J. 2012. Targeted mutagenesis of the human *chemokine (C-C motif) receptor 5* gene using zinc-finger nucleases and TAL effector nucleases. *Ph.D. Dissertation*, Seoul National University, Seoul, Korea.

Kim, H., Um, E., Cho, S.R., Jung, C., Kim. H. and Kim, J.S. 2011. Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods,* 8: 941-943.

Kim, J.S., Lee, H.J. and Carroll D. 2010. Genome editing with modularly assembled zinc-finger nucleases. *Nat. Methods,* 7: 91.

Kim, S., Lee, M.J., Kim, H., Kang, M. and Kim, J.S. 2011. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat Methods,* 8: 7.

Kim, S., Kim, E.J. and Kim, J.S. 2010. Construction of combinatorial libraries that encode zinc-finger-based transcription factors. *Methods Mol Biol,* 649: 133-147.

Lee, H.J., Kim, E. and Kim, J.S. 2010. Targeted chromosomal deletions in human cells using zinc-finger nucleases. *Genome Res*, 20**:** 81-89.

Lei, Y., Lee, C.L., Joo, K.I., Zarzar, J., Liu, Y., Dai, B., Fox, V. and Wang, P. 2011. Gene editing of human embryonic stem cells via an engineered baculoviral vector carrying zinc-finger nucleases. *Mol Ther,* 9: 942-950.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P. and Yang, B. 2012. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol,* 30: 390-392.

Lieber, M.R. 2010. The mechanism of double-strand DNA break repair by the non-homologous DNA end-joining pathway. *Annu Rev Biochem*, 79: 181

Lombardo, A. Genovese, P., Beausejour, C.M., Colleoni, S., Lee, YL, Kim, KA, Ando, D., Urnov, F.D., Galli, C., Gregory, P.D., Holmes, M.C. and Naldini, L. 2007. Gene editing in human stem cells using zinc-finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol,* 25: 1298-1306.

Lupski, J.R. 2007. Genomic rearrangements and sporadic disease. *Nat Genet,* 39: S43-S47.

Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F. and Joung, J.K. 2009. Oligomerized pool engineering (OPEN): an ‘‘opensource’’protocol for making customized zinc-finger arrays. *Nat Protoc,* 4: 1471-1501.

Mani, M., Kandavelou, K., Dy, F. J., Durai, S. and Chandrasegaran, S. 2005. Design, engineering, and characterization of zinc-finger nucleases. *Biochem Biophys Res Commun,* 335: 447-457.

Mashimo, T., Takizawa, A., Voigt, B., Yoshimi, K., Hiai, H., Kuramoto, T. and Serikawa, T. 2010. Generation of knockout rats with X-linked severe combined immunodeficiency (X-SCID) using zinc-finger nucleases. *PLoS ONE,* 5: e8870.

Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D. and Wolfe, S.A.

2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol,* 26: 695-670.

Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., Gregory1, P.D., Pabo, C.O. and Rebar, E.J. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol,* 25**:** 778-785.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., Dulay, G.P., Hua, K.L., Ankoudinova, I., Cost, G.J., Urnov, F.D., Zhang, H.T., Holmes, M.C., Zhang, L., Gregory, P.D. and Rebar, E.J. 2010. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol,* 29**:** 143-148.

Moehle, E.A., Rock, J.M., Lee, Y.L., Jouvenot, Y., DeKelver, R.C., Gregory, P.D., Urnov, F.D. and Holmes, M.C. 2007. Targeted gene addition into a specified location in the human genome using designed zinc-finger nucleases. *Proc Natl Acad Sci USA*, 104: 3055-3060.

Moore, F.E., Reyon, D., Sander, J.D., Martinez, S.A., Blackburn, J.S., Khayter, C., Ramirez, C.L., Joung, J.K. and Langenau, D.M. 2012. Improved Somatic Mutagenesis in Zebrafish Using Transcription Activator-Like Effector Nucleases (TALENs). *PLoS ONE,* 7: e37877.

Morbitzer, R., Elsaesser, J., Hausner, J. and Lahaye, T. 2011. Assembly of custom TALE-type DNA binding domains by modular

cloning. *Nucleic Acids Res,* 39: 5790-5079.

Morton, J., Davis, M.W., Jorgensen, E.M. and Carroll, D. 2006. Induction and repair of zinc-finger nuclease-targeted double-strand breaks in *Caenorhabditis elegans* somatic cells. *Proc Natl Acad Sci USA,* 103: 16370-16375.

Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res*, 39: 9283-9293.

Ochiai, H., Fujita, K., Suzuki, K., Nishikawa, M., Shibata, T., Sakamoto, N. and Yamamoto, T. 2010. Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes Cells,* 15: 875-885.

Osakabe, K., Osakabe, Y. and Toki, S. 2010. Site-directed mutagenesis in *Arabidopsis* using custom-designed zinc-finger nucleases. *Proc Natl Acad Sci USA,* 107: 12034-12039.

Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods,* 8: 765-770.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L., Guschin, D.Y., Rupniewski, I., Waite, A.J., Carpenito, C., Carroll1, R.G., Orange, J.S., Urnov, F.D., Rebar, E.J., Ando, D., Gregory, P.D., Riley J.L., Holmes, M.C. and Juneet, C.H. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using

zinc-finger nucleases. *Nat Biotechnol,* 26**:** 808-816.

Polymeropoulos, M.H., Higgins, J.J., Golbe, L.I., Johnson, W.G., Ide, S.E., Di Iorio, G., Sanges, G., Stenroos, E.S., Pho, L.T., Schaffer, A.A., Lazzarini, A.M., Nussbaum, R.L. and Duvoisin, R.C. 1996. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science,* 274: 1197-1199.

Porteus, M.H. and Baltimore, D. 2003. Chimeric nucleases stimulate gene targeting in human cells. *Science,* 300: 763.

Porteus, M.H. 2006. Translating the lessons from gene therapy to the development of regenerative medicine. *Mol Ther,* 13: 349-341.

Ramalingam, S., Kandavelou, K., Rajenderan, R. and Chandrasegaran, S. 2011. Creating designed zinc-finger nucleases with minimal cytotoxicity. *J Mol Biol,* 405: 630-641.

Ramirez, C.L., Certo, M.T., Mussolino, C., Goodwin, M.J., Cradick, T.J.,McCaffrey, A.P., Cathomen, T., Scharenberg, A.M. and Joung, J.K. 2012. Engineered zinc-finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res,* doi:10.1093/nar/gks179.

Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. 2011. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol,* 30: 460-465.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y., Pierick, C.J., Hoffman, E., Maeder, M.L., Khayter, C., Reyon, D., Dobbs, D., Langenau, D.M., Stupar, R.M.,

Giraldez, A.J., Voytas, D.F., Peterson, R.T., Yeh, J.R. and Joung, J.K. 2011. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Method,* 8: 67-69.

Santiago, Y., Chan, E., Liu, P.Q., Orlando, S., Zhang, L., Urnov, F.D., Holmes, M.C., Guschin, D., Waite, A., Miller J.C., Rebar, E.J., Gregory, P.D., Klug, A. and Collingwoodet, T.N. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA,* 105**:** 5809-5814.

Schambach, A., Cantz, T., Baum, C. and Cathomen, T. 2010. Generation and genetic modification of induced pluripotent stem cells. *Expert Opin Biol Ther,* 10: 1089-103.

Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F., Artandi, S.E., Wernig, M. and Joung, J.K. 2011. In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells,* 29: 1717-1726.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimäki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K. and Wigler, M. 2007. Strong association

of de novo copy number mutations with autism. *Science,* 316: 445-449.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X., Choi, V.M., Rock, J.M., Wu, Y.Y., Katibah, G.E., Zhifang, G., McCaskill, D., Simpson, M.A., Blakeslee, B., Greenwalt, S.A., Butler, H.J., Hinkley, S.J., Zhang, L., Rebar, E.J., Gregory, P.D. and Urnov, F.D. 2009. Precise genome modification in the crop species Zea mays using zincfinger nucleases. *Nature,* 459: 437-441.

Smith, J., Berg, J. M. and Chandrasegaran, S. 1999. A detailed study of he substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Res,* 27: 674-681.

Sollu, C., Pars, K., Cornu, T.I., Thibodeau-Beganny, S., Maeder, M.L., Joung, J.K., Heilbronn, R. and Cathomen, T. 2010. Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res,* 38: 8269-8276.

Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol,* 25: 786-793.

The Int. HapMap Consort. 2003. The International HapMap Project. *Nature,* 426: 789-796.

Takasu, Y., Kobayashi, I., Beumer, K., Uchino, K., Sezutsu, H., Sajwan, S., Carroll, D., Tamura, T. and Zurovec, M. 2010.

Targeted mutagenesis in the silkworm Bombyx mori using zinc-finger nuclease mRNA injection. *Insect Biochem Mol Biol,* 40: 759-765.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung J.K. and Voytas, D.F. 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature,* 459: 442-445.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature,* 435**:** 646-651.

Wang, J., Friedman, G., Doyon, Y., Wang, N.S., Li, C.J., Miller, J.C., Hua, K.L., Yan, J.J., Babiarz, J.E., Gregory, P.D. and Holmes, M.C. 2012. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res,* doi: 10.1101/gr.122879.111.

Wood, A.J., Lo, T.W., Zeitler, B., Pickle, C.S., Ralston, E.J., Lee, A.H., Amora, R., Miller, J.C., Leung, E., Meng, X., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D. and Meyer, B.J. 2011. Targeted genome editing across species using ZFNs and TALENs. *Science,* 15: 307.

Yao, Y., Nashun, B., Zhou, T., Qin, L., Qin, L., Zhao, S., Xu, J., Esteban, M.A. and Chen, X. 2012. Generation of CD34+ cells from CCR5-disrupted human embryonic and induced pluripotent

stem cells. *Hum Gene Ther,* 23: 238-242.

Young, J.J., Cherone, J.M., Doyon, Y., Ankoudinova, I., Faraji, F.M., Lee, A.H., Ngo, C., Guschin, D.Y., Paschon, D.E., Miller, J.C., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., Harland, R.M. and Zeitler, B. 2011. Efficient targeted gene disruption in the soma and germ line of the frog *Xenopus tropicalis* using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA*, 108: 7052-7057.

Yu, S., Luo, J., Song, Z., Ding, F., Dai, Y. and Li, N. 2011. Highly efficient modification of beta-lactoglobulin (BLG) gene via zinc-finger nucleases in cattle. *Cell Res,* 21: 1638-1640.

Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordóñez, A., Hannan, N.R., Rouhani, F.J., Darche, S., Alexander, G., Marciniak, S.J., Fusaki, N., Hasegawa, M., Holmes, M.C., Di Santo, J.P., Lomas, D.A., Bradley, A. and Vallier, L. 2012. Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, 478: 391-394.

Zhang, F., Gu, W., Hurles, M.E. and Lupski, J.R. 2009. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet,* 10: 451-481.

Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T., Joung, J.K. and Voytas, D.F. 2010. High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc-finger nucleases.

Proc Natl Acad Sci USA, 107: 12028-12033.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M. and Arlotta, P. 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol*, 29: 149-153.

Zou, J., Maeder, M.L., Mali, P., Pruett-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q., Porteus, M.H., Joung, J.K. and Cheng, L. 2009. Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell,* 5: 97-110.

Zou, J., Mali, P., Huang, X., Dowey, S.N. and Cheng, L. 2011. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood,* 118: 4599-608.

국문 초록

ZFN 및 TALEN은 줄기세포를 비롯한 다양한 동·식물세포 안에 존재하는 유전자를 대상으로 특정 위치만을 인식하여 절단함으로써 유전자를 교정하거나 제거할 수 있는 새로운 방법으로 분자의학 및 생명공학 분야에서 크게 주목받고 있다. 그러나 유전자가위 기술을 통해 만든 돌연변이 세포와 정상 세포 간에 구별할 수 있는 방법이 없어 폭 넓은 사용에 제약이 있어왔다. 이번 연구에서는 세 종류의 리포터 시스템을 통하여 효과적으로 돌연변이 세포들을 선별할 수 있는 방법을 개발하였다. 돌연변이를 갖고 있는 대리 유전자와 유전자가위를 동시에 세포에 도입하면 일부 세포에서 유전자가위의 작용으로 돌연변이가 고쳐져서 대리 유전자가 발현되는데 이들 세포를 분리하면 세포 내 유전자에도 높은 효율로 돌연변이가 도입되어 있음을 확인하였다. Flow cytomer-based cell sorting, magnetically separation, drug selection의 방법 중 세포의 종류 및 목적에 맞게 선택하여 사용할 수 있도록 하였고 각 방법을 사용하였을 때 선별 전에 비해 선별 후 세포에서 돌연변이의 비율이 6배~92배 증가하는 것을 확인 하였다. 이 대리 유전자 리포터 시스템의 가장 큰 장점은 간단하면서도 효과적으로 돌연변이 세포들을 선별할 수 있다는 점이다. 이 방법은 유전자가위를 이용한 유전자치료의 효율을 획기적으로 개선할 수 있으며 에이즈와 같은 바이러스 질환과 여러 유전질환에 맞춤형 유전자가위 기술을 적용할 수 있는 가능성을 더 크게 만들어 줄 것이다.

학 번: 2007-20304