Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

Published Version
doi:10.1172/jci72992

Accessed
June 21, 2017 11:42:23 AM EDT

Citable Link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:22556419

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

(Article begins on next page)
The use of ZFNs, TALENs, and CRISPR-Cas9 for precision genome editing in mammals

Rajat M. Gupta MD$^{1,2}$ & Kiran Musunuru, MD, PhD, MPH$^{1,2}$

1 Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA
2 Division of Cardiovascular Medicine, Brigham and Women’s Hospital, Boston, MA 02115, USA

Correspondence to:
Kiran Musunuru, MD, PhD, MPH
Harvard University
Sherman Fairchild Biochemistry Bldg 160
7 Divinity Ave
Cambridge, MA 02138, USA
E-mail: kiranmusunuru@gmail.com / Phone: (617) 496-5361 / Fax (617) 496-8351
The past decade has been one of rapid innovation in genome-editing technology. The opportunity now exists for investigators to manipulate virtually any gene in a diverse range of cell types and organisms with targeted nucleases designed with sequence-specific DNA-binding domains. The rapid development of the field has allowed for highly efficient, precise, and now cost-effective means by which to generate human and animal models of disease using these technologies.

This review will outline the history and recent development of genome-editing technology, culminating with use of CRISPR-Cas9 to generate novel mammalian models of disease. While the road to using this same technology for treatment of human disease is long, the pace of innovation over the past 5 years and early successes in model systems builds anticipation for this prospect.

**The emergence of genome-editing technology**

The classical method for gene modification is homologous recombination. This approach has been widely used in mouse embryonic stem cells to generate germline knockout or knockin mice(1, 2). A disadvantage is that it typically takes more than a year to generate a genetically modified mouse using the standard approach. Furthermore, similar attempts at using homologous recombination in human cells have been proven to be far more challenging, and alternative approaches to knock down gene expression, such as antisense oligonucleotides and short interfering RNAs, have instead become standard. However, these approaches only transiently reduce gene expression, and the effect is usually incomplete and can often affect off-target
genes(3). These shortcomings have fueled the demand for more effective methods of gene modification.

A new wave of technology that is variously termed “gene editing,” “genome editing,” or “genome engineering” has emerged to address this demand by giving investigators the ability to precisely introduce a variety of genetic alterations, ranging from knockin of single nucleotide variants to insertion of genes to deletion of chromosomal regions, into mammalian cells far more efficiently than traditional homologous recombination. We describe the key advantages and disadvantages of the three most popular genome-editing tools (summarized in Table 1). This description is not meant to be a comprehensive review of the work leading to the development of the tools, but rather to give readers a working knowledge of the tools and be able to select among the tools for desired tasks.

**Zinc finger nucleases**

Zinc finger nucleases (ZFNs) are increasingly being used in academic and industry research for a variety of purposes ranging from the generation of animal models to human therapies(4). ZFNs are fusion proteins comprising an array of site-specific DNA-binding domains—adapted from zinc finger-containing transcription factors—attached to the endonuclease domain of the bacterial FokI restriction enzyme. Each zinc finger domain recognizes a 3- to 4-basepair (bp) DNA sequence, and tandem domains can potentially bind to an extended nucleotide sequence (typically with a length that is a multiple of 3, usually 9 bp or 12 bp) that is unique in a cell’s genome.
To cleave a specific site in the genome, ZFNs are designed as a pair that recognizes two sequences flanking the site, one on the forward strand, the other on the reverse strand. Upon binding of the ZFNs on either side of the site, the pair of FokI domains dimerize and cleave the DNA at the site, generating a double-strand break (DSB) with 5’ overhangs(4). Cells repair DSBs using either (1) non-homologous end-joining (NHEJ), which is more straightforward and can occur during any phase of the cell cycle, but occasionally results in erroneous repair, or (2) homology-directed repair (HDR), which typically occurs during late S phase or G2 phase when a sister chromatid is available to serve as a repair template (Fig. 1).

The error-prone nature of NHEJ can be exploited to introduce frame-shifts into the coding sequence of a gene, potentially knocking out the gene by either of two mechanisms: premature truncation of the protein and nonsense-mediated decay of the mRNA transcript (Fig. 2). Alternatively, HDR can be utilized in a fashion similar to homologous recombination, with the introduction of a repair template with a desired mutation flanked by homology arms (Fig. 2). Though mechanistically similar, the efficiency of genome editing with HDR is significantly improved over traditional homologous recombination, because the first step of the process (generation of a DSB) is induced rather than occurring spontaneously. The exogenous repair template can be either a double-strand DNA vector or a single-stranded DNA oligonucleotide (ssODN). For ssODNs, homology arms of as little as 20-nucleotide length can enable introduction of mutations into the genome(5-7). In many cases, the efficiency is sufficiently improved that antibiotic selection to identify correctly targeted clones is unnecessary. If antibiotic selection is not used, then extra steps to remove the cassette from the genome using
systems like Cre-lox and FLP-FRT are unnecessary, in contrast to traditional homologous recombination.

Despite the advantages of genome editing with ZFNs, there are several potential disadvantages. It has not proven to be straightforward to assemble zinc finger domains to bind an extended stretch of nucleotides with high affinity. This has made it difficult for non-specialists to routinely engineer ZFNs. To surmount this difficulty, an academic consortium has developed an “open-source” library of zinc finger components and protocols to perform screens to identify ZFNs that bind with high affinity to a desired sequence; nonetheless, it can still take months for non-specialists to obtain optimized ZFNs. A commercial option to obtain optimized ZFNs is available, but the expense may be prohibitive for some investigators.

Another potential disadvantage is that target site selection is limited—the “open-source” ZFN components can only be used to target binding sites every few hundred bp throughout the genome. While this may be a non-issue if an investigator seeks to knock out a gene, since a frameshift introduced anywhere in the early coding sequence of the gene can produce the desired result, it may present challenges if a particular site is required, e.g., to knock in a specific mutation into a gene. Since the introduction of the “open-source” platform, alternative platforms to engineer optimized ZFNs have since emerged, with varying degrees of speed, flexibility in site selection, and success rates.

Finally, a significant concern about the use of proteins designed to introduce DSBs into the genome is that they will do so not only at the desired site but also at off-target sites. In one study
in which ZFNs were used for genome editing in human pluripotent stem cells, the investigators identified 10 possible off-target genomic sites based on high sequence similarity to the on-target and found a single off-target mutation in 184 clones assessed(15). Two subsequent studies of ZFNs using unbiased genome-wide methods to identify potential off-target sites for several ZFN pairs revealed infrequent off-target events at numerous loci in a cultured human tumor cell line(16, 17). Thus, investigators should be cognizant of the possibility that ZFNs designed for a particular purpose may incur undesired off-target events at a low rate. One strategy to reduce off-target events is to use a pair of ZFNs that have distinct FokI domains that are obligate heterodimers(18, 19). This prevents a single ZFN from binding to two adjacent off-target sites and generating a DSB; rather, the only way an off-target event could occur is if both ZFNs in a pair bind adjacently and thus allow the FokI dimer to form.

**Transcription activator-like effector nucleases**

The recent discovery of a class of proteins called transcription activator-like effectors (TALEs), exclusive to a group of plant pathogens, has led to the characterization of a novel DNA-binding domain, termed TAL repeats. The naturally occurring TAL repeats comprise tandem arrays with 10 to 30 repeats that bind and recognize extended DNA sequences(20). Each repeat is 33 to 35 amino acids in length, with two adjacent amino acids (termed the repeat-variable di-residue, or RVD) conferring specificity for one of the four DNA basepairs(21-25). Thus, there is a one-to-one correspondence between the repeats and the basepairs in the target DNA sequences. Understanding the RVD code has made it possible to create a new type of engineered site-specific nuclease that fuses a domain of TAL repeats to the FokI endonuclease domain, termed
TAL effector nucleases (TALENs)(26, 27)(Fig. 1). TALENs are similar to ZFNs in that they can generate DSBs at a desired target site in the genome and so can be used to knock out genes or knock in mutations in the same way (Fig. 2).

In comparison to ZFNs, TALENs have turned out to be much easier to design. The RVD code has been employed to engineer many TAL repeat arrays that bind with high affinity to desired genomic DNA sequences; it appears that more often than not a de novo engineered TAL repeat array will bind to a desired DNA sequence with high affinity(27, 28). TALENs can be designed and constructed in as short a time as two days and in as large a number as hundreds at a time(29, 30); indeed, a library with TALENs targeting all of the genes in the genome has been constructed(31).

One potential advantage over ZFNs is that the TAL repeat array can be easily extended to whatever length desired. Whereas engineered ZFNs typically bind 9- or 12-bp sequences, TALENs are often built to bind 18-bp sequences or even longer, with the theoretical possibility of achieving greater affinity and specificity with TALENs. Another possible advantage of TALENs over ZFNs is that there appear to be fewer constraints on site selection, with at least a few potential sites available in each 100 bp of genomic DNA. However, it also appears that methylation of the target site attenuates the binding affinity(32), although it is possible to tweak the RVD code to accommodate methylated DNA bases(33, 34).

As with ZFNs, off-target effects are a significant concern with TALENs. A study in which TALENs were used for genome editing in human pluripotent stem cells found low but
measurable rates of mutagenesis at some of 19 possible off-target sites based on sequence similarity to the on-target site(28). Although comparative data is scarce, one study found that for TALENs and ZFNs targeting the same site in the *CCR5* gene, the TALENs produce fewer off-target mutations than the ZFNs at a highly similar site in the *CCR2* gene(35). Furthermore, the ZFNs produced greater cell toxicity (i.e., inhibited their growth) when introduced into cells compared to the TALENs. Thus, there is a general perception in the field that TALENs are “cleaner” than ZFNs. As with ZFNs, TALENs with obligate heterodimer FokI domains are routinely used to minimize the possibility of off-target events.

A clear disadvantage of TALENs is their significantly larger size compared to ZFNs. The typical size for a cDNA encoding a TALEN is ~3 kb, whereas a cDNA encoding a ZFN is only ~1 kb. In principle, this makes it harder to deliver and express a pair of TALENs into cells compared to ZFNs, and the size of the TALENs make them less attractive for therapeutic applications in which they must be delivered in viral vectors with limited cargo size or as RNA molecules. Furthermore, the highly repetitive nature of the TALENs may impair their ability to be packaged and delivered by some viral vectors(36), though this can apparently be overcome by diversifying the coding sequences of the TAL repeats(37).

**CRISPR-Cas9**

The recent discovery of bacterial adaptive immune systems known as clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems have led to the newest set of genome-editing tools. CRISPR-Cas systems use a combination of proteins
and short RNAs to target specific DNA sequences for cleavage. The bacteria collect "protospacers" from foreign DNA sequences (e.g., from bacteriophages), incorporate them into their genomes, and use them to express short guide RNAs, which can then be used by a CRISPR-Cas system to destroy any DNA sequences matching the protospacers.

In early 2013, four groups demonstrated that heterologous expression of a CRISPR-Cas system from *Streptococcus pyogenes*, comprising the Cas9 protein along with guide RNA(s) (either two separate RNAs, as found in bacteria, or a single chimeric RNA), in mammalian cells results in DSBs at target sites with (1) a 20-bp sequence matching the protospacer of the guide RNA and (2) an adjacent downstream NGG nucleotide sequence (termed the protospacer-adjacent motif, or PAM)(38-41). This occurs via the formation of a ternary complex: Cas9 binds the non-protospacer portion of the guide RNA, the protospacer of the guide RNA hybridizes with one strand of the genomic DNA, and Cas9 binds to the PAM in the DNA. Cas9 then catalyzes the DSB in the DNA at a position three basepairs upstream of the PAM(40)(Fig. 1).

In contrast to ZFNs and TALENs, which must be built from scratch for each new target site, CRISPR-Cas9 can be easily adapted to target any genomic sequence by changing the 20-bp protospacer of the guide RNA, which can be done with simple molecular biology. The Cas9 protein component remains unchanged. This ease of use for CRISPR-Cas9 is a significant advantage over ZFNs and TALENs, especially in generating a large set of vectors to target numerous sites(39) or even genome-wide libraries(42-44). Another potential advantage of CRISPR-Cas9 is the ability to multiplex, i.e., to use multiple guide RNAs in parallel to target multiple sites simultaneously in the same cell(38, 39). This makes it straightforward to mutate
multiple genes at once or to engineer precise deletions in a genomic region, although it should be noted that simultaneous use of multiple ZFN or TALEN pairs can achieve the same outcomes.

One potential disadvantage of CRISPR-Cas9 is site selection, though in this regard it compares favorably with ZFNs and TALENs. Even with the most flexible version of the \textit{S. pyogenes} CRISPR/Cas system, site selection is limited to 23-bp sequences on either strand that end in an NGG motif (the PAM for \textit{S. pyogenes} Cas9), which occur on average once every 8 bp\cite{38}.

However, CRISPR-Cas systems from other species are starting to be employed in mammalian cells\cite{38, 45, 46}, and their versions of Cas9 have different PAM requirements, which allows for targeting of sites in the genome for which the \textit{S. pyogenes} system is not optimal. For example, the canonical \textit{Neisseria meningitides} Cas9 PAM has been reported to be NNNNGATT, although it appears to be more tolerant of variation in the PAM compared to \textit{S. pyogenes} Cas9\cite{45, 46}.

Another disadvantage of CRISPR-Cas9 is the size of the Cas9 protein. The cDNA encoding \textit{S. pyogenes} Cas9 is ~4 kb in size, making it somewhat larger than a TALEN and much larger than a ZFN. This size makes it challenging to deliver via viral vectors or as an RNA molecule. The chimeric version of the guide RNA is only ~100 nucleotides in size, but it needs to be delivered in parallel with Cas9, either as a separate RNA molecule or via a DNA cassette with a separate promoter (typically a RNA polymerase III promoter such as U6). Here again, the emerging availability of CRISPR-Cas systems from other species may prove helpful. The cDNA encoding \textit{N. meningitidis} Cas9 is ~3.2 kb in size and so should allow for easier delivery, which may be important for therapeutic applications.
Perhaps the biggest concern regarding CRISPR-Cas9 is the issue of off-target effects. It has recently been demonstrated that although that each nucleotide within the 20-nt protospacer contributes to overall *S. pyogenes* Cas9 binding and specificity, single mismatches are often well tolerated, and multiple mismatches can sometimes be tolerated depending on their locations in the protospacer (47-50). Systematic analysis of the effect of alterations in the protospacer reveals an increasing tolerance for mismatches with increasing distance from the PAM. A number of studies in mammalian cells have documented off-target mutations occurring at significant rates at sites with sequence similarity to the on-target sites, occasionally rivaling or even surpassing mutagenesis at the on-target sites (47-52). It has been posited that alternative CRISPR-Cas systems such as that from *N. meningitides* may offer better targeting specificity by virtue of their longer protospacers (24 nt for *N. meningitides*) and longer PAMs. Experimental confirmation of improved specificity in mammalian cells remains to be shown. Early results with the *N. meningitides* CRISPR-Cas9 system suggest that it may be less tolerant of mismatches in the protospacer compared to the *S. pyogenes* system (45).

Efforts to improve the specificity of CRISPR-Cas9 in mammalian cells are in progress. One strategy has been to use a mutant version of Cas9 that can only introduce a single-strand nick into the target DNA, rather than a DSB. Use of a pair of “nickase” CRISPR-Cas9 complexes with binding sites on opposite strands flanking the target site can produce the equivalent of a DSB with 5’ overhangs (Fig. 1), which is then repaired by NHEJ or HDR and can result in an on-target alteration. At an off-target site, a single-strand nick would be fixed by a different mechanism (base excision repair pathway) that is much less likely to result in a mutation. Because the likelihood of two nickases binding near each other elsewhere in the genome is very
low, the off-target mutation rate should be dramatically reduced. Indeed, testing of this strategy in mammalian cells has demonstrated a reduction in off-target activity by up to three orders of magnitude with at most a modest reduction in on-target efficacy (49, 52, 53). Another strategy to reduce off-target effects is to reduce the length of the protospacer portion of the guide RNA, which makes it less tolerant of mismatches and thus can preserve the on-target efficacy while reducing off-target mutagenesis (54).

**Genome editing in mammalian models**

Although the creation of mouse lines with genetic alterations such as gene knockouts or conditional alleles has long been feasible with traditional homologous recombination employed in mouse embryonic stem cells, the last few years have seen the application of novel genome-editing tools for the generation of genetically modified mice with unprecedented ease and efficiency. Furthermore, these tools have made it possible to genetically modify animals for which embryonic stem cell lines are not widely available.

Initial studies of the efficacy of genome-editing tools in the mutagenesis of mammalian embryos were performed with rats. Inspired by studies in which injection of RNAs encoding ZFNs directly into the embryos of fruit flies and zebrafish yielded stable, heritable genomic alterations, injection of ZFN-encoding RNAs into one-cell rat embryos successfully generated monoallelic and biallelic frameshift mutations resulting in gene knockout (55, 56). Numerous knockout rats have since been generated using this ZFN strategy. Subsequently, both TALENs and CRISPR-Cas9 have been used in similar fashion to generate knockout rats (57-59).
A particular advantage is that it is possible to obtain knockout animals in the first generation (assuming the targeted gene is not embryonic lethal), dramatically speeding up the time needed to do genetic studies in animals. Another advantage of this approach is that embryos from any of a variety of animal strains can be used; in the case of mice, there is no longer a restriction to a limited number of embryonic stem cell lines that necessitate backcrossing to an inbred strain of choice. Embryos from that inbred strain can be used to directly generate the knockout mice. Similarly, embryos from a strain that already carries genetic alterations can be used, relieving the need for many generations of interbreeding to obtain mice with multiple genetic alterations. The ability to perform multiplex gene targeting with CRISPR-Cas9 is also helpful in this regard.

All three engineered nucleases outlined above have proven effective at producing targeted mutations in mouse embryos(60-68). The efficiencies vary wildly depending on the nuclease, target site in the genome, and amount of RNAs injected. The most striking demonstration of efficiency has been with CRISPR-Cas9, with simultaneous targeting of both alleles of two genes in 80% of mice(66). CRISPR-Cas9 has also been used along with ssODNs or double-strand DNA donor vectors in mouse embryos to knock in tags and fluorescent markers into endogenous gene loci and, most impressively, to generate conditional knockout mice in one step by simultaneously knocking in two loxP sites flanking an exon of a gene(67).

Finally, the high efficiencies of the genome-editing tools, particularly CRISPR-Cas9, has made it possible to generate targeted mutations in animals far beyond the reach of the traditional homologous recombination/embryonic stem cell approach. Both TALENs and CRISPR-Cas9
have now been used to generate genetically modified monkeys (69, 70), in each case targeting genes involved in human diseases. This is a remarkable accomplishment that suggests that there is no technical barrier to using genome-editing tools to modify human embryos, notwithstanding the profound social and ethical repercussions that would result if such attempts were to be made.

**Genome editing in human cells**

To date, there have been a number of reports demonstrating the feasibility of performing genome editing in human pluripotent stem cells (hPSCs) with ZFNs, TALENs, and CRISPRs(15, 28, 39, 53, 71-75). Genetically altered hPSCs offer the possibility of differentiating wild-type and mutant cell lines into whatever somatic cell type desired, potentially giving new insights into disease pathophysiology. In one such study, the investigators generated induced pluripotent stem cells (iPSCs) from patients with Parkinson disease caused by the G2019S mutation of the LRRK2 gene as well as control individuals(75). Upon differentiation into midbrain dopaminergic neurons, the cell lines displayed striking differences in whole-genome gene expression patterns, with clustering analysis showing that in some cases a patient line and a control line were more closely matched than lines generated from two different patients. Indeed, even iPSC lines generated from the same patient failed to cluster together, demonstrating the high degree of heterogeneity among iPSC lines. As an alternative approach, the investigators used ZFNs to correct the G2019S mutation in three of the patient-derived iPSC lines and to insert the mutation into a control iPSC line. They found that the matched sets of wild-type/mutant cell lines clustered together very closely, confirming the superiority of the genome-editing strategy for disease modeling studies. The investigators consistently found that mutant neurons displayed less
neurite outgrowth and more apoptosis in response to oxidative stress than matched wild-type neurons.

Other human cell types have proven to be amenable to genome editing. In one study, the investigators isolated intestinal stem cells from cystic fibrosis patients homozygous for the common delta508 mutation in the **CFTR** gene\(^{(76)}\). They used CRISPR-Cas9 targeting the site of the mutation, along with a double-strand DNA donor vector, to correct one mutant allele (sufficient to “cure” the disease in this recessive disorder). They then used the mutant and corrected stem cells to create intestinal organoids in culture. Whereas the mutant organoids failed to respond to forskolin treatment by swelling, consistent with a lack of functional CFTR protein, the corrected organoids did respond by swelling, demonstrating a functional rescue.

The remarkable efficiency and ease of use of CRISPR-Cas9, where only 20 nucleotides in the guide RNA need be changed to retarget the nuclease, has led to the development of genome-wide “CRISPR interference” or “CRISPRi” libraries with the potential to knock out each of the genes in the genome. Three groups have performed proof-of-principle, genome-wide knockout screens in cells, two in human cells\(^{(42, 43)}\) and one in mouse cells\(^{(44)}\). The results of the screens compared favorably with traditional genome-wide RNA interference screens, establishing a powerful new complementary approach to RNA interference to probe gene function in an unbiased fashion.

**Conclusion**
The rapid development and improvement of genome editing tools provides investigators with three well-characterized options for experiments as diverse as forward genetic screens to correction of pathogenic mutations in iPSC-derived human cells. ZFNs, TALENs, and CRISPRs can all generate site-specific double stranded breaks with varying degrees of specificity and efficiency. The early uses of these systems have demonstrated remarkable new possibilities and allowed for the creation of model systems in a wide variety of organisms. With each iteration the technology has improved, and the prospects for the study of human disease with genome editing has never been better.

Acknowledgements

Given the explosive nature of the genome-editing field in recent years, especially in the past year, we regret that due to space limitations we were unable to cite and describe many worthy studies, and we apologize for their omission.

Competing Interests

The authors declare that they do not have any competing or financial interests.

References


<table>
<thead>
<tr>
<th>Origin</th>
<th>Typical genomic target site</th>
<th>Flexibility in site selection</th>
<th>Ease of use/affordability</th>
<th>Size (ability to be packaged in viruses)</th>
<th>Efficacy</th>
<th>Specificity/lack of off-target effects</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFNs</td>
<td>Zinc finger proteins (widely found in nature)</td>
<td>Pair of 9- or 12-bp sequences</td>
<td>+</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>(6, 9, 11-17, 72)</td>
</tr>
<tr>
<td></td>
<td>TAL effector proteins in plant pathogens</td>
<td>Pair of 13-bp or longer sequences (no length limitation)</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>(27-30)</td>
</tr>
<tr>
<td>TALENs</td>
<td>Bacterial immune system (S. pyogenes, other species)</td>
<td>20-nt protospacer + 3-nt PAM (S. pyogenes)</td>
<td>++</td>
<td>+++</td>
<td>+</td>
<td>+++ (depending on strategy used)</td>
<td>(37-41, 48-52, 54, 66)</td>
</tr>
</tbody>
</table>

+ indicates least favorable, +++ indicates most favorable