

RESEARCH HIGHLIGHT

A CRISPR way to engineer the human genome

Sivaprakash Ramalingam, Narayana Annaluru and Srinivasan Chandrasegaran*

Abstract

RNA-guided genome engineering based on the type II prokaryotic CRISPR/Cas system provides an efficient and versatile method for targeted manipulation of mammalian genomes.

Ever since the discovery of restriction enzymes in 1970, the holy grail of molecular biologists has been site-specific manipulation of mammalian genomes, including the human genome. Restriction enzymes are not useful for this task since they recognize relatively small DNA sequences, which occur much too frequently within a genome. Even though the genome sequences of many organisms have now been determined, understanding the functions of their constituent genes requires sequence editing by deleting or modifying genes, and then studying the resulting phenotypes. Powerful new techniques had to be developed in order to achieve the targeted modification of sequences in the human genome. Scientists reasoned that they could use the highly conserved universal process of homologous recombination (HR) to this end. Cells use HR to mediate site-specific recombination and maintain their genome integrity, especially during the repair of a double-strand break (DSB), which would otherwise be lethal to the cells. DSB repair of a damaged chromosome by HR works via the copy-and-paste mechanism, which uses the homologous DNA segment from the undamaged chromosomal partner as template; it is the most accurate form of repair. Gene targeting - the process of replacing a gene by HR - uses the investigator-provided extrachromosomal donor DNA and invokes the cell's own repair machinery for gene conversion. With the exception of mouse cells, gene targeting is not an efficient process in mammalian cells - only one in a million treated cells undergoes the desired genome modification. However, when a defined genomic

DSB is introduced, HR is induced efficiently at that site in a large fraction of cells in a population. Thus, generation of a targeted genomic DSB has been the rate-limiting step in the development of HR technology for gene-modification or genome engineering. The challenge then was to develop a general means of delivering a targeted genomic DSB at a unique chromosomal locus in order to stimulate homology-directed repair at that site with the exogenously added donor DNA.

Two powerful techniques have emerged for site-specific manipulation of mammalian genomes by delivering a targeted DSB, followed by repair of the DSB by non-homologous end joining (NHEJ) or by HR (Figure 1). First came the use of gene-editing nucleases, with the creation of designer zinc finger nucleases (ZFNs) in 1996 [1,2]; this was followed by transcription-activator-like effector nucleases (TALENs) in 2010 [3] with the discovery of TALE DNA-binding modules that are found in plant virulence factors from *Xanthomonas* bacteria [4,5], and building on the experience gained from ZFN development. More recently, several groups have successfully employed RNA-guided genome engineering based on the type II prokaryotic CRISPR/Cas system [6-10].

Gene targeting using ZFNs and TALENs

We reported the creation of designer ZFNs [1], and then, in collaboration with Dana Carroll's laboratory in Utah, showed the utility of ZFNs in gene targeting using frog oocytes as a model system [2]. Custom-designed ZFNs - proteins designed to cut at specific DNA sequences - combine the non-specific cleavage domain (N) of the FokI restriction endonuclease with zinc finger proteins (ZFPs). The Cys₂His₂ zinc finger (ZF) motif can target specific sequences by virtue of its unique 30 amino acid ββα structure, which is stabilized by a zinc ion. Each ZF motif usually recognizes 3 to 4 bp, and binds DNA by inserting the α-helix into the major groove of the double helix. Amino acids within the α-helix (positions -1, +1, +2, +3, +5, +6) of the ZF motif can be changed, while conserving the remaining amino acids as a consensus backbone, to generate ZF motifs with new sequence specificities. Most ZF motifs make contact with their target 3-bp site; however, when there is an aspartic acid residue present at the +2 position of the α-helix, the ZF motif makes contact with a base outside the 3-bp site,

*Correspondence: schandra@jhsp.edu
Department of Environmental Health Sciences, Bloomberg School of Public Health, Johns Hopkins Medical Institutions, 615 North Wolfe Street, Baltimore, MD 21205, USA

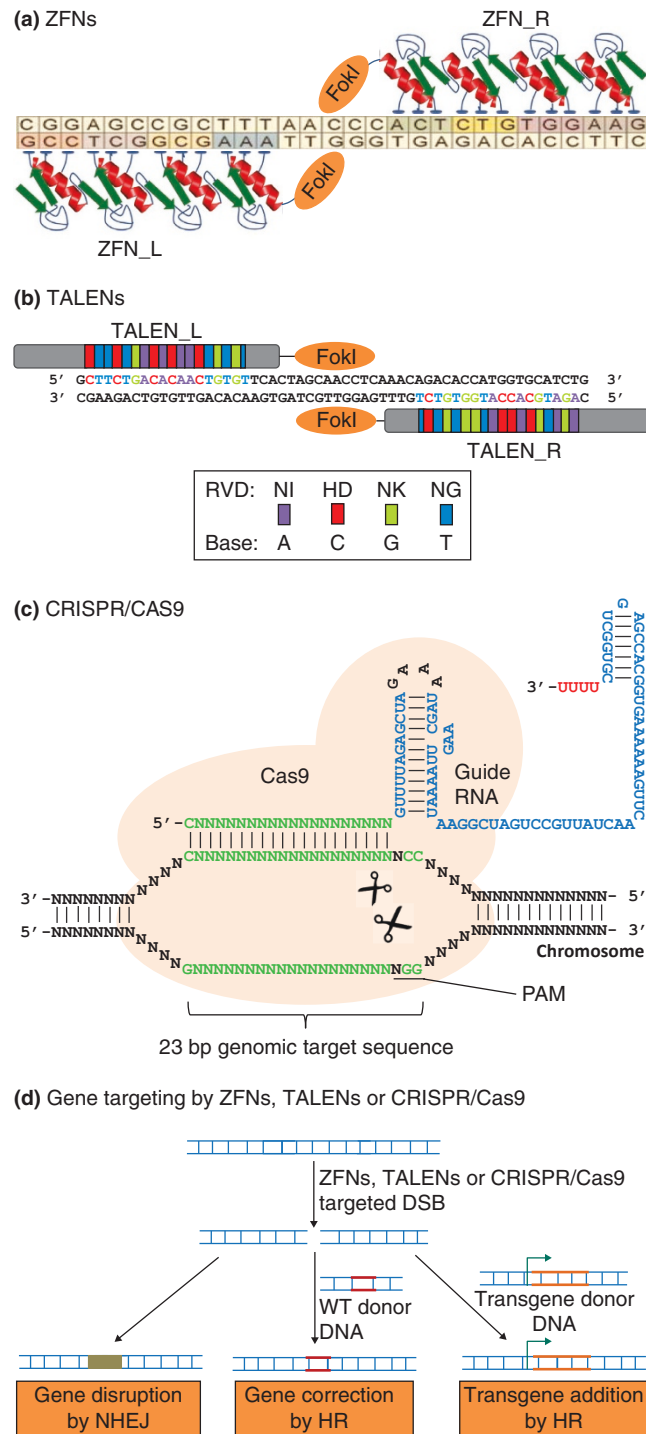


Figure 1. Schematic representation of zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated protein 9 (Cas9) bound to their cognate sites to deliver a targeted genomic double-strand break. (a) ZFNs, (b) TALENs and (c) CRISPR/Cas9. Both ZFNs and TALENs possess a FokI nuclease domain for DNA cleavage, whereas the CRISPR-based guide RNA recruits a Cas nuclease for DNA cleavage. All three systems use components for DNA binding that can be readily engineered for sequence specificity: ZFNs and TALENs possess twin DNA-binding domains, arranged in an inverted dimer, for DNA recognition (zinc finger and transcription activator-like effector domains, respectively); the CRISPR/Cas9 system recognizes DNA through an RNA-DNA interaction between the target site and a CRISPR-based synthetic guide RNA. **(d)** Targeted genome engineering by non-homologous end joining (NHEJ) or homologous recombination (HR) using the gene editing tools. In the homology-directed repair experiment, cells are co-transfected with both the gene editing tool and a wild-type (WT) or transgene donor DNA.

changing its recognition sequence to a 4-bp site. This ZF motif contact outside the 3-bp site also influences the specificity of neighboring ZF motifs, complicating the generation of ZFPs by simple modular design, where each ZF motif recognizes a triplet sequence. Therefore, design and selection of each ZF motif has to be done in a context-dependent fashion to obtain highly sequence-specific ZFPs, and this is laborious and time consuming. Normally, four such ZF motifs are linked together in tandem to generate a ZFP that binds to a 12-bp site. Binding of two four-finger ZFN monomers each recognizing a 12-bp inverted site is necessary because dimerization of the FokI cleavage domain is required to produce a DSB (Figure 1a). Therefore, four-finger ZFNs effectively have a 24-bp recognition site, which is long enough to specify a unique address within the human genome. Because the recognition specificities of ZFPs can be manipulated experimentally, ZFNs offer a general way to deliver a targeted DSB to the human genome. Owing to the high conservation of DNA repair mechanisms, application of ZFN-mediated gene targeting has been successful in numerous species including *Xenopus laevis*, *Arabidopsis*, wheat, rice, *Drosophila*, *Caenorhabditis elegans*, zebrafish, silkworm, rodents, mice, pigs, cows, butterflies and various human cell types, including immortalized cell lines, primary somatic cells, embryonic stem cells and induced pluripotent stem cells.

The TALENs designed in a modular style similar to ZFNs are also able to target unique loci in complex genomes. While ZFNs use ZF motifs as DNA-binding modules, TALENs utilize the central repeat domain of TAL effectors (TALEs) for DNA recognition (Figure 1b); both use the FokI catalytic domain as a DNA cleavage module [3]. The TALE central repeat domain consists of repeating units of 33 to 35 amino acids. Each repeat is largely identical, except for two highly variable amino acids at positions 12 and 13, referred to as the repeat variable di-residues (RVDs). Whereas each ZF motif recognizes three to four bases, each TALE motif recognizes a single nucleotide, with the recognition specificity determined by the RVD (for example, NI recognizes A, HD recognizes C, NG or HG recognizes T, and NN recognizes G or A) [4,5]. Unlike the ZF motifs, the TALE modules each appear to bind DNA without interference from neighboring modules. The DNA recognition code thus provides a one-to-one correspondence between the array of amino acid repeats and the nucleotide sequence of the DNA target. This simple DNA recognition code and its modular nature make TALEs an ideal platform for constructing custom-designed artificial DNA nucleases. Some reports suggest that TALENs have the same efficiency of cutting, but markedly lower cytotoxicity, compared with ZFNs targeted to the same genomic locus.

Gene targeting using the bacterial CRISPR/Cas system

Bacteria and archaea have evolved an adaptive defense mechanism that uses a CRISPR/Cas system to degrade complementary sequences present within invading viral and plasmid sequences. The type II CRISPR/Cas system relies on integration of foreign DNA fragments into clustered regularly interspaced short palindromic repeat (CRISPR) loci. Upon transcription and processing, these inserts produce short CRISPR RNAs (crRNAs), which then anneal to a *trans*-activating crRNA (tracrRNA), enabling CRISPR-associated (Cas) proteins to direct sequence-specific degradation of the foreign DNA. It turns out that Cas9 endonuclease-mediated cleavage can also function efficiently using a fusion of crRNA and tracrRNA to form a synthetic guide RNA (gRNA). Several groups have now shown that they can engineer the type II bacterial CRISPR/Cas9 system to function with custom gRNA in human cells *in vitro* to direct sequence-specific cleavage (Figure 1c). For the endogenous *AAVS1* locus, Mali *et al.* [6] achieved targeting efficiency of 10% to 25% in 293T cells, 13% to 8% in K562 cells, and 2% to 4% in induced pluripotent stem cells. Cong *et al.* [7] independently reported that the CRISPR/Cas system is able to mediate genomic cleavage with comparable or superior efficiency to a pair of TALENs targeting the same *EMX1* locus. Cho *et al.* [8] used the CRISPR/Cas system to cleave two human genomic sites, *CCR5* (C-C motif chemokine receptor type 5) and *C4BPB* (complement component 4 binding protein, beta), in a targeted manner for genome editing, but not at related off-target sites within the human genome that are most homologous to the 23-bp target sequences. All three groups show that the simultaneous introduction of multiple gRNAs into human cells can achieve multiplex gene editing of multiple targeted loci, establishing the potential of the CRISPR/Cas system for high-throughput applications. Mali *et al.* [6] and Cong *et al.* [7] also used a Cas9-mutant nuclease, nickase, to generate only single-strand breaks at a target locus, thereby promoting HR while minimizing NHEJ-mediated mutagenesis.

Two other independent studies report RNA-guided genome engineering in microbial organisms and in zebrafish, demonstrating broad utility of the CRISPR/Cas technology (Figure 1d). Hwang *et al.* [9] tested the CRISPR/Cas9 system *in vivo* using zebrafish embryos and showed that it can induce targeted genetic modifications with efficiencies similar to that of ZFNs and TALENs. The authors successfully targeted more than 80% of the tested sites in zebrafish. They monitored the toxicity induced by gRNA:Cas9-encoding mRNA in zebrafish by observing the numbers of deformed and dead embryos. These frequencies were similar to those observed previously using ZFNs and TALENs. Jiang *et al.*

[10] show the versatility of the CRISPR technique for bacterial genome engineering by introducing precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli*. Although the range of targetable sequences for the CRISPR/Cas system is somewhat constrained by the requirement of an NGG 3-bp motif (known as the protospacer adjacent motif, or PAM, sequence) in genomic DNA just 3' to the target site, it could potentially be expanded by using homologs with different PAM requirements.

Gene editing nucleases versus the CRISPR/Cas system

The target specificity of the CRISPR/Cas system is programmed with gRNA, without any need for enzyme engineering. Only one customized gRNA is required to target a specific sequence; the same Cas9 enzyme is suitable for all other sequence targets of the human genome. In contrast, ZFNs and TALENs require the design and assembly of two nucleases, one for each half of the target site. Furthermore, the target specificity of gRNAs is achieved by a 20-bp RNA-DNA interaction, which is encoded by short sequences of approximately 100 bp. gRNAs are therefore much simpler and easier to engineer than are ZFNs or TALENs. The short length of gRNA sequences also avoids difficulties associated with delivering longer and highly repetitive ZFN/TALEN-encoding vectors into cells. The specificity and versatility of gene editing by the CRISPR/Cas9 system, coupled with the ease-of-use, will likely encourage wider applications of the technology, especially by smaller laboratories with limited resources.

But the most important question, especially for clinical translation, is as follows: how specific are these methods? In other words, do they cleave at sites in the genome other than the ones they are designed to target? Potential wide-ranging use of ZFNs, TALENs or CRISPR gene editing tools in clinical trials (ongoing clinical trials include ZFNs designed to knock-out *CCR5* in human T cells and CD34+ stem cells to render these cells resistant to infection by HIV) warrants a systematic and careful evaluation of their cleavage specificity, through the determination both of the locations and of the frequencies of unwanted off-target events on a genome-wide scale. Off-target cleavage likely occurs at sites whose sequences differ slightly from the target sites or at partial target sites. To ensure safety, off-target cleavage analysis is essential for each and every newly targeted loci of the human genome. Of the three methods, off-target cleavage is most well characterized for ZFNs and to a much lesser

extent for TALENs and the CRISPR/Cas9 system. Deep sequencing and whole-genome sequencing may reveal off-target mutations induced by these gene editing tools. Such studies will be of utmost importance for safe human genome modification and potential clinical translation to gene therapy.

Abbreviations

bp, base pair; Cas, CRISPR-associated protein; CRISPR, clustered regularly interspaced short palindromic repeat; crRNA, short CRISPR RNA; DSB, double-strand break; gRNA, guide RNA; HR, homologous recombination; HDR, homology-directed repair; NHEJ, non-homologous end joining; PAM, protospacer adjacent motif; RVD, repeat variable di-residues; TALE, transcription activator-like effector; TALEN, transcription activator-like effector nuclease; tracrRNA, *trans*-activating crRNA; ZF, zinc finger; ZFN, zinc finger nuclease; ZFP, zinc finger protein.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The research in our laboratory was supported by a grant from National Institutes of Health (GM077291). SR is supported by an Exploratory Research Grant from MSCRF and NA is supported by a grant from National Science Foundation (MCB 0718846).

Published: X Month 2013

References

1. Kim YG, Cha J, Chandrasegaran S: **Hybrid restriction enzymes: Zinc finger fusions to FokI cleavage domain.** *Proc Natl Acad Sci U S A* 1996, **93**:1156-1160.
2. Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim Y-G, Chandrasegaran S: **Stimulation of homologous recombination through targeted cleavage by chimeric nucleases.** *Mol Cell Biol* 2001, **21**:289-297.
3. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF: **Targeting DNA double-strand breaks with TAL effector nucleases.** *Genetics* 2010, **186**:757-761.
4. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U: **Breaking the code of DNA binding specificity of TAL-type III effectors.** *Science* 2009, **326**:1509-1512.
5. Moscou MJ, Bogdanove AJ: **A simple cipher governs DNA recognition by TAL effectors.** *Science* 2009, **326**:1501.
6. Mali P, Yang L, Esvelt KM, Aach J, Guell M, Dicarlo JE, Norville JE, Church GM: **RNA-guided human genome engineering via Cas9.** *Science* 2013. doi:10.1126/science.1232033.
7. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F: **Multiplex genome engineering using CRISPR/Cas systems.** *Science* 2013. doi:10.1126/science.1231143.
8. Cho SW, Kim S, Kim JM, Kim JS: **Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.** *Nat Biotechnol* 2013. doi:10.1038/nbt.2507.
9. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK: **Efficient genome editing in zebrafish using a CRISPR-Cas system.** *Nat Biotechnol* 2013. doi:10.1038/nbt.2501.
10. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA: **RNA-guided editing of bacterial genomes using CRISPR-Cas systems.** *Nat Biotechnol* 2013. doi:10.1038/nbt.2508.

doi:10.1186/gb-2013-14-2-107

Cite this article as: Ramalingam S, *et al.*: A CRISPR way to engineer the human genome. *Genome Biology* 2013, **14**:107.