

## Invited Mini Review

## Recent advances in developing molecular tools for targeted genome engineering of mammalian cells

Kwang-il Lim\*

Department of Medical and Pharmaceutical Sciences, College of Science, Sookmyung Women's University, Seoul 140-742, Korea

Various biological molecules naturally existing in diversified species including fungi, bacteria, and bacteriophage have functionalities for DNA binding and processing. The biological molecules have been recently actively engineered for use in customized genome editing of mammalian cells as the molecule-encoding DNA sequence information and the underlying mechanisms how the molecules work are unveiled. Excitingly, multiple novel methods based on the newly constructed artificial molecular tools have enabled modifications of specific endogenous genetic elements in the genome context at efficiencies that are much higher than that of the conventional homologous recombination based methods. This minireview introduces the most recently spotlighted molecular genome engineering tools with their key features and ongoing modifications for better performance. Such ongoing efforts have mainly focused on the removal of the inherent DNA sequence recognition rigidity from the original molecular platforms, the addition of newly tailored targeting functions into the engineered molecules, and the enhancement of their targeting specificity. Effective targeted genome engineering of mammalian cells will enable not only sophisticated genetic studies in the context of the genome, but also widely-applicable universal therapeutics based on the pinpointing and correction of the disease-causing genetic elements within the genome in the near future. [BMB Reports 2015; 48(1): 6-12]

## INTRODUCTION

Multiple, newly emerging molecular tools are enabling effective genome engineering of mammalian cells. While genome engineering of bacterial cells, whose genome size is much smaller than that of mammalian cells roughly by order of

three, has been performed up to the level of *de novo* synthesis of designed whole genome (1), genome engineering of mammalian cells has been mainly done at the scale of modification of a specific endogenous locus. Genome engineering of mammalian cells has often narrowly involved gene targeting, along with mainly four different purposes: disruption (or deletion), correction, replacement, and additional insertion of a gene of interest.

The conventional means of gene knock-in and knock-out during gene targeting relies on intracellular homologous recombination. However, the efficiency of the conventional method is very low because the cellular process occurs rarely only at a frequency of  $10^{-6}$ , even when donor nucleic acid elements homologous to the target genomic locus are provided within cells (2, 3). Fortunately, when DNA double-strand breaks (DSBs) are additionally made near the target locus, occurrence of local genome modifications can be increased by over 1000-times (4), reaching a level that is practically executable. Such an increase is mediated by cellular DNA repair pathways that can be stimulated by the occurrence of DSBs. These pathways include the homologous recombination (HR) pathway in the presence of homologous DNA donor molecules and the error-prone non-homologous end-joining (NHEJ) pathway mainly in the absence of such donor molecules (5, 6).

Development of novel molecular machinery that can efficiently make DSB(s) at a chosen genome locus has thereby greatly enhanced genome engineering of mammalian cells. Such machinery is generally composed of a targeting domain searching for a specific genome locus and a catalytic domain actually generating DSBs at the targeted locus. Our recently gained ability to sophisticatedly manipulate gene(s) or genetic component(s) in the context of genome as loss/gain-of-function strategies allows better understanding of genotype-to-phenotype relationships from the genetic viewpoint. In addition, the use of genome engineering tools has been further extended to generate new therapeutic strategies, correcting or altering the genomic loci that account for a disease state.

This minireview introduces the most recently spotlighted molecular genome engineering tools. It mainly discusses the key features, potentials, limitations, and genetic and biomedical applications of the molecular tools with a special focus on ongoing modifications.

\*Corresponding author. Tel: +82-2-2077-7627; Fax: +82-303-0799-0396; E-mail: klim@sm.ac.kr

<http://dx.doi.org/10.5483/BMBRep.2015.48.1.165>

Received 4 August 2014

**Keywords:** Customized gene targeting, Cytotoxicity, DNA binding domains, Genome engineering, Nucleases

## ZINC FINGER NUCLEASES

Zinc finger nucleases (ZFNs) are generally constructed by fusion of a nuclease that works as a dimer for DNA cleavage reaction and multiple zinc fingers that play the guide role toward the intended genomic target locus. Zinc finger has a characteristic structure with two beta sheets and an alpha helix whose multiple residues mainly determine its binding DNA sequences (7). The secondary structures are coordinated by a single zinc ion. Each zinc finger unit normally recognizes a three base pair DNA sequence. When multiple zinc fingers are assembled in tandem they can recognize a longer sequence (8) eventually allowing the targeting of a unique address on large genomes. ZFNs have been used for gene targeting in various cell types from differentiated ones including multiple immortalized cell lines with their own tissue origin (9) to pluripotent stem cells (10). ZFN-mediated gene targeting has provided therapeutic effects for multiple disease including infectious disease. For example, a specific disruption of the C-C chemokine receptor type 5 (CCR5) gene encoding a product that functions as a co-receptor for human immunodeficiency virus type 1 (HIV-1) in CD4+ T cells could make the host cells significantly resistant to the viral infection (11). In addition, use of ZFNs helps develop various animal models of disease by making it possible to edit the genes specifically associated with the disease of interest (12).

Despite its highly enhanced gene targeting efficiency (up to 2-digit percentages) compared with the conventional method, ZFN-mediated engineering of huge genomes as in mammalian cells often suffers from its limited specificity. Insufficient specificity of ZFN can lead to a significant level of cleavage activity for off-target genomic loci. High-throughput sequencing analysis has identified off-target distributions on human genome, showing that zinc fingers in tandem can interact with off-target sequences that differ by at least a few bases from the intended target (13).

Binding of only a fraction of zinc finger units within a ZFN to DNA can lead to unsatisfactory targeting performance. In addition, individual zinc fingers can have a significant fraction of binding affinity to off-target sequences compared with that for their actual target sequences (14-16), possibly increasing the chance of ZFN targeting failure. These off-target effects often account for the observed cytotoxic events in cells during ZFN-based genome engineering (17).

Recent efforts have focused on improving the specificity of ZFNs for safer genome engineering. Multiple research groups have attempted to especially improve the assembly process of zinc finger units given the purpose. Zinc finger units in a ZFN do not work in a perfectly modular fashion, which means that individual zinc fingers can affect the bindings of their neighboring fingers to target sequences (18). Sub-optimal design of ZFNs without considering such context effects of assembled zinc fingers can result in DNA cleavage at unintended genomic loci. A new approach, so-called context-dependent assembly (CoDA)

of zinc finger units, which considers the potential interactions among fingers, has helped generate efficient ZFNs that enable targeted genome engineering for zebrafish embryos (19). In general, ZFNs are designed to cleave DNA via formation of a heterodimer composed of ZFN1 that binds to a half of the target DNA sequence and ZFN2 that binds to the other half. However, unwanted nuclease domain-mediated formation of homodimers, such as ZFN1-ZFN1 and ZFN2-ZFN2, often ends up with cleavage of off-targets on large genome. Rational engineering of the nuclease domain involved in the dimerization could significantly reduce the formation of homodimers, ultimately decreasing the off-target activity of the newly constructed ZFNs (20). On the other hand, engineering of a conventionally used type II restriction endonuclease, FokI, generated a nickase that produces single-strand DNA breaks instead of DSBs that the wild-type nuclease normally produces (21). The resulting DNA lesions were corrected by the cellular HR pathway, rather than the more mutagenic NHEJ pathway (21). Such a DNA repair pathway shift by engineering of nuclease domains was helpful to reduce the cytotoxicity led by ZFN-mediated genome modifications.

## TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES

Another DNA-binding protein, transcription activator-like effector (TALE), has more recently received attention as an alternative guiding molecule for targeted genome engineering. The protein originates from plant pathogenic bacteria, *Xanthomonas*. It contains an array of repeat domains that consist of about 34 amino acids, each recognizing a single DNA base (22). Especially, two residues, called as internal repeat variable di-residues (RVDs), in each repeat domain determine the target DNA nucleotide based on the following code with a certain degree of degeneracy as shown in the format of di-residues:target base: NI:A, NG & IG:T, HD:C, NN:A or G (23, 24). Compared with the array of zinc fingers, where interactions between individual finger units are significant, TALEs assembled for targeting are much less affected by such context effects, a characteristic that facilitates customized gene targeting (25). The TALE DNA binding domains are often fused to a nonspecific nuclease, FokI, for genome engineering of mammalian cells.

A TALE nuclease (TALEN) generated by modular assembly of TALEs and their fusion to FokI efficiently disrupted the CCR5 gene, a human endogenous gene often targeted for testing the performance of genome engineering nuclease tools (26). In the study, truncated variants of TALE molecules were used to construct TALENs for better performance, and the novel TALEN system showed significantly less cytotoxic effects compared with ZFNs (26). Other studies showed that the length of linker between TALE and nuclease is very important to determine the activity as well as specificity of constructed TALEN (2, 26).

Despite the ongoing success of TALEN-mediated genome

engineering, this new method also has some limitations to overcome. For example, methylated cytosine can be resistant to recognition by TALE (27). This DNA methylation sensitivity may hinder the TALEN-mediated cleavage of genomic loci in CpG islands that often contain methylated cytosine bases (28). Interestingly, a TALE variant that loses glycine in its RVDs could efficiently recognize methylated target sequences, potentially broadening the use of TALENs for targeted genome modifications (27). Another hurdle to the use of TALENs for customized genome engineering is the difficulty in assembling an array of TALE molecules due to the resulting multiple DNA repeats that can cause severe recombination problems during cloning steps (29, 30). Different efforts have been recently made to overcome the well-known difficulty. They include the reduction of repetitive nature of TALE-encoding DNA sequences by codon optimization as well as use of advanced molecular biology strategies, such as Golden Gate cloning (29, 30).

## CRISPR-CAS9 SYSTEMS

The most recently emerging genome engineering tool, clustered regularly interspaced short palindromic repeats (CRISPRs)-Cas9 systems, uniquely uses RNA molecules to target a genomic locus. The targeting is based on the Watson-Crick base pairing between the RNA molecules incorporated within Cas9 and the targeted part of the genome locus (31). The conventionally used CRISPR-Cas9 for genome modifications naturally functions to make acquired immune responses in a bacterium, *Streptococcus pyogenes*. CRISPR-Cas9 system was originally composed of three components: CRISPR RNA (crRNA) that plays the main targeting roles, sequence invariant trans-activating crRNA (tracrRNA) that is involved in the processing of the pre-crRNA, and a single protein component with the catalytic functionality, Cas9. Recently engineered CRISPR-Cas9 system often uses the fusion construct of crRNA and tracrRNA, called as a guide RNA (32), for simplicity and simultaneous delivery of the two RNA components into cells (31). Designed CRISPR-Cas9 systems typically target a chosen DNA sequence of around 20 bp, and their target recognition additionally requires the presence of the protospacer-adjacent motif (PAM), a short DNA sequence following the 3' end of the target (33).

One of the most valuable advantages of using CRISPR-Cas9 systems is that it allows us to very easily customize genome engineering for chosen targets just by changing the guide RNA sequences. In comparison, other genome engineering tools often require empirical optimization steps during assembly of their targeting subunits (34). In addition, multiple loci on a certain genome can be simultaneously engineered by applying this new RNA-based tool in a multiplexing fashion (32), a still challenging task for other genome engineering tools. Targeted cleavage by CRISPR-Cas9 has been shown to be insensitive to methylation of target DNA sequences (35). Engineered

CRISPR-Cas9 systems have been used to edit endogenous genes on human genome, including the C4BPB and CCR5 genes (36).

As with the aforementioned genome engineering tools, CRISPR-Cas9 systems are also reactive to multiple off-targets on genomes of mammalian cells. Approaches combining randomized mutations in target sequences and DNA cleavage profiling through high-throughput sequencing have provided clues about the sequence ranges of potential off-targets given CRISPR-Cas9 systems (34). Only 7 to 12 base matches between target DNA and guide RNA may be sufficient for target recognition, likely resulting in a high chance of off-target cleavages (34). CRISPR-Cas9 off-targets that differ by up to 5 bases from the intended target have been also reported (37). As expected, there will be more off-targets for CRISPR-Cas9 on larger genomes, such as those of mammalian species (34). In a study of endogenous gene targeting CRISPR-Cas9 systems disrupted not only human CCR5 gene and hemoglobin  $\beta$  gene as the intended targets, but also CCR2 and hemoglobin  $\delta$  genes as off-targets (38).

A significant fraction of the early studies on CRISPR-Cas9 system has focused on enhancing the specificity and reducing the cytotoxicity of the new genome engineering tools. As one example, a Cas9 nickase obtained from the D10A mutation did not generate detectable mutations on human genome after target locus cleavage (33). This performance enhancement seemed to be due to a shift in the underlying host DNA repair pathway toward the less error-prone homology-directed repair (HDR) pathway (33, 39). It has been also proposed that decrease of the concentration of Cas9 in cells can increase the specificity of the enzymatic cleavage reaction (35). However, such a benefit might be compromised to a certain degree by the decrease in the overall efficiency of target cleavage. Prior knowledge about potential off-targets given a genome of interest and a target locus (or target sequence) will be useful to fine-tune the guide RNA sequence for higher specificity and efficiency of genome engineering. To achieve such a goal, a computational tool has been developed (40). The bioinformatics tool searches for off-targets with considerations of the target sequence, number of mismatches, type of genome, and type of Cas9 that recognizes its own PAM sequence as user inputs (40).

## HOMING ENDONUCLEASES

Unlike artificially constructed ZFNs and TALENs, where separate DNA binding protein units are fused to nucleases, homing endonucleases naturally have both DNA binding and catalytic domains in them. Homing endonucleases, so-called meganucleases, originate from various species including bacteria, yeast and archaea. The enzymes inherently recognize their own DNA target sequences ranging from 12 bp to 45 bp, a characteristic that confers specificity sufficient to target a unique locus on human genome (41). In addition, monomeric

homing endonucleases, such as I-SceI and I-Anil, do not require the formation of dimer of the enzymes for DNA cleavage reactions (42, 43). This feature is helpful to reduce off-target cleavage that can be potentially led by the actions of the homodimer of the nucleases unwantedly formed instead of the intended heterodimer, a key problem during ZFN and TALEN-mediated genome editing.

Because natural recognition sequences of homing endonucleases mostly quite differ from our intended targets on mammalian genomes, either the enzymatic inherent recognition sequences need to be inserted into our target locus, or the sequences need to be changed via modifications of their DNA recognition domains prior to being used for genome engineering applications. Most recent efforts have focused on the second scenario to extend the use of homing endonucleases. For example, two protein engineering schemes have been applied. The first is a rational or random mutagenesis of the key domains and, more specifically, amino acid residues of individual meganucleases involved in the inherent target recognition. The second involves domain recombinations of different meganucleases (41, 44). The products of such approaches could process novel DNA target sequences showing that homing endonucleases can be also used for customized gene targeting. Structural information on the DNA binding domain, catalytic domain, interdomain interface of individual enzymes, and the enzyme dimer interfaces can provide a useful guide to accelerate the development of novel homing endonucleases (45).

It is supposed that off-target cleavage problems are less serious for the genome engineering based on homing endonucleases, because the enzymes have relatively long recognition sequences ranging up to 45 bp (41). However, there is still a risk of targeting failure that can eventually cause cytotoxic effects on genome manipulated cells (46). For example, I-SceI generated cleavage-mediated DNA integrations at an off-target site on the genome of HT-1080 cells more frequently than at the target site, even though the off-target differed by four bases from the target (46). To further enhance the specificity of homing endonucleases, TALEs have been fused to the enzymes (47). Such a modification scheme has yielded nucleases with improved specificity and higher activity. The underlying mechanism would involve an event where the inserted TALE increases the chance of formation of the enzyme-DNA substrate complex (47). This novel nuclease tool can modify better the genomic locus that is simultaneously recognized by both assembled TALEs and the chosen homing nuclease.

## RECOMBINASES AND RETROVIRAL INTEGRASES

Recombinase-based genome engineering tools originate from multiple species including fungi, bacteria, and bacteriophage. As homing endonucleases do, recombinases also respond rigidly to their own natural recognition sequences during the cut-and-paste of DNA substrates, which limits their usefulness in genome engineering (48). Therefore, recent approaches

have focused on the weakening and, ultimately, removal of the inherent sequence recognition rigidity of recombinases. The relevant main strategies have employed randomization of the catalytic domain of recombinases, likely involved in the DNA recognition, and selection of the generated enzyme variants toward gaining of significantly altered recognition patterns (48-50). For example, multiple libraries of Cre recombinase mutants have been constructed and directly evolved to achieve a biomedical goal, obtaining an ability to excise HIV-1 provirus from the human genome (50). The study employed gradual library evolution processes containing the intermediate combination and shuffling of selected pools of recombinase variants between evolution stages to accomplish the challenging goal. The finally obtained variant from the long series of 126 evolution steps could target a site within HIV-1 genome, called loxLTR, that is composed of an asymmetric sequence with around 50% similarity to the original recognition site of Cre recombinase having a symmetric sequence feature (50). Along with the directed evolution approach, attaching zinc finger domains to recombinases has much shortly enabled specific endogenous gene targeting for human cells by shifting the recombinase sequence recognition to the customized targets (48, 49).

In contrast to recombinases, retroviral integrases do not need the presence of specific recognition sequences prior to genome engineering. In addition to this advantage, retroviral integrases have a high efficiency for integration reaction sufficient to possibly make very high percentages of genome-engineered mammalian cell populations, up to 100% depending on the cell type. However, retroviral insertion of genetic elements into mammalian genomes occurs in non-random fashions. Two representative forms of retroviral vectors, MLV-based and HIV-1 based ones, insert genetic elements into the host genome preferring transcriptional start sites (TSS) (51-53) and the regions throughout transcriptional units (53, 54), respectively. Such inherent integration preferences for the gene regulatory or gene encoding regions can lead to detrimental effects in the host by activating oncogenes or disrupting tumor suppressor genes (51, 53). To shift the integration preference of a retrovirus, MLV, to safer genomic regions, we engineered the retroviral integrase. We constructed novel integrases by introducing multiple zinc finger complexes into two permissive sites within the integrase. The retroviruses with the enzyme variants produced altered integration patterns without significant preference for TSS (55). Although we could not observe integration events at the intended genomic loci given the zinc finger complexes, the successful strategy to effectively shift integration patterns toward safer regions not enriched in gene regulatory elements will extend the use of retroviral integrases for genome engineering.

## CONCLUDING REMARKS

Targeted genome engineering of mammalian cells can enable

**Table 1.** Key features of recently spotlighted molecular genome engineering tools

	ZFN	TALEN	CRISPR-Cas9
Origin of DNA recognition unit	DNA binding motifs within transcription factors	Structural repeats within the secreted transcriptional regulatory proteins from plant pathogenic bacteria	Pathogen-targeting RNAs of the adaptive immune systems of prokaryotes
Molecular type of DNA recognition unit	Proteins	Proteins	RNAs
Size of DNA recognition unit	Around 30 amino acids per unit targeting a 3 bp sequence	Around 34 amino acids per unit targeting a single base	Around 20 bases targeting the equivalent size of DNA sequence
Modular assembly of DNA recognition units	Possible, but limited by the interactions between neighboring units	Possible	Not relevant, use of a single guide RNA in general
Targeting mechanism	Protein-DNA interactions	Protein-DNA interactions	Watson-Crick base pairing between guide RNA and target DNA
Presence of repeated units	Yes, as zinc finger domains in tandem	Yes, as TALE DNA binding domains in tandem	No
Presence of DNA cleavage activity	Yes	Yes	Yes

not only sophisticated genetic studies in the context of genome but also widely applicable universal therapeutics based on the pinpointing and correction of the disease-causing genetic elements on genome. However, the conventional HR-based genome modification has an efficiency that is insufficient to translate the ambitious idea to reality. Fortunately, we have recently witnessed the rapid emergence of various novel molecular genome engineering tools. While the molecular tools work by following their own distinct mechanisms, they have consistently shown greatly improved genome editing efficiency up to the level of allowing the targeted modifications of endogenous genetic elements at two digit frequencies. Several defects mostly in the specificity of the molecular targeting, however, still limit the use of the genome engineering tools in clinically related applications. Therefore, since the rise of the tools, a significant fraction of research efforts has been invested to improve their targeting specificity, ultimately aiming for the minimization of the cytotoxic effects from off-target genome editing. This minireview has introduced the frontline of recent molecular genome engineering tools with their key features (as partly summarized in Table 1) and ongoing modifications. Attempts to develop and enhance novel genome editing tools will be continued and even accelerated. The following success anticipated in the near future will let us take the full benefits from customized genome engineering of mammalian cells in better understanding biology and advancing medicine and biotechnology.

## REFERENCES

- Gibson DG, Glass JL, Lartigue C et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52-56
- Kim Y, Kweon J, Kim A et al (2013) A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* 31, 251-258
- Pruett-Miller SM, Connelly JP, Maeder ML, Joung JK and Porteus MH (2008) Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther* 16, 707-717
- Liang F, Romanienko PJ, Weaver DT, Jeggo PA and Jasin M (1996) Chromosomal double-strand break repair in Ku80-deficient cells. *Proc Natl Acad Sci U S A* 93, 8929-8933
- Ramalingam S, Annaluru N and Chandrasegaran S (2013) A CRISPR way to engineer the human genome. *Genome Biol* 14, 107
- Mao Z, Bozzella M, Seluanov A and Gorbunova V (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair* 7, 1765-1771
- Cornu TI, Thibodeau-Beganny S, Guhl E et al (2008) DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* 16, 352-358
- Lam KN, van Bakel H, Cote AG, van der Ven A, and Hughes TR (2011) Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucl Acids Res* 39, 4680-4690
- Toscano MG, Anderson P, Munoz P et al (2013) Use of zinc-finger nucleases to knock out the WAS gene in K562 cells: a human cellular model for Wiskott-Aldrich syndrome. *Dis Models Mechanisms* 6, 544-554
- Hockemeyer D, Soldner F, Beard C et al (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol* 27, 851-857
- Perez EE, Wang J, Miller JC et al (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* 26, 808-816
- Cui X, Ji D, Fisher DA, Wu Y, Briner DM and Weinstein EJ (2011) Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol* 29, 64-67
- Pattanayak V, Ramirez CL, Joung JK and Liu DR (2011) Revealing off-target cleavage specificities of zinc-finger

- nucleases by in vitro selection. *Nat Methods* 8, 765-770
14. Dreier B, Segal DJ and Barbas CF 3rd (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Molec Biol* 303, 489-502
  15. Dreier B, Beerli RR, Segal DJ, Flippin JD and Barbas CF 3rd (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 276, 29466-29478
  16. Dreier B, Fuller RP, Segal DJ et al (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 280, 35588-35597
  17. Pruett-Miller SM, Reading DW, Porter SN and Porteus MH (2009) Attenuation of zinc finger nuclease toxicity by small-molecule regulation of protein levels. *PLoS Genet* 5, e1000376
  18. Fu F and Voytas DF (2013) Zinc Finger Database (ZiFDB) v2.0: a comprehensive database of C(2)H(2) zinc fingers and engineered zinc finger arrays. *Nucl Acids Res* 41, D452-455
  19. Sander JD, Dahlborg EJ, Goodwin MJ et al (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* 8, 67-69
  20. Szczypek M, Brondani V, Buchel J, Serrano L, Segal DJ and Cathomen T (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* 25, 786-793
  21. Kim E, Kim S, Kim DH, Choi BS, Choi IY and Kim JS (2012) Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res* 22, 1327-1333
  22. Bogdanove AJ (2014) Principles and applications of TAL effectors for plant physiology and metabolism. *Curr Opin Plant Biol* 19C, 99-104
  23. Boch J, Scholze H, Schornack S et al (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 326, 1509-1512
  24. Moscou MJ and Bogdanove AJ (2009) A simple cipher governs DNA recognition by TAL effectors. *Science* 326, 1501
  25. Cermak T, Doyle EL, Christian M et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39, e82
  26. 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
  27. Valton J, Dupuy A, Daboussi F et al (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem* 287, 38427-38432
  28. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13, 484-492
  29. Weber E, Gruetzner R, Werner S, Engler C and Marillonnet S (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One* 6, e19722
  30. Zhang F, Cong L, Lodato S, Kosuri S, Church GM and Arlotta P (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* 29, 149-153
  31. Mali P, Esvelt KM and Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods* 10, 957-963
  32. Mali P, Yang L, Esvelt KM et al (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823-826
  33. Cho SW, Kim S, Kim Y et al (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 24, 132-141
  34. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA and Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31, 839-843
  35. Hsu PD, Scott DA, Weinstein JA et al (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827-832
  36. Cho SW, Kim S, Kim JM and Kim JS (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31, 230-232
  37. Fu Y, Foden JA, Khayter C et al (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 31, 822-826
  38. Cradick TJ, Fine EJ, Antico CJ and Bao G (2013) CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res* 41, 9584-9592
  39. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819-823
  40. Bae S, Park J and Kim JS (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473-1475
  41. Grizot S, Epinat JC, Thomas S et al (2010) Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Res* 38, 2006-2018
  42. Bolduc JM, Spiegel PC, Chatterjee P et al (2003) Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes Dev* 17, 2875-2888
  43. Moure CM, Gimble FS and Quijoco FA (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J Molec Biol* 334, 685-695
  44. Wang Y, Khan IF, Boissel S et al (2014) Progressive engineering of a homing endonuclease genome editing reagent for the murine X-linked immunodeficiency locus. *Nucleic Acids Res* 42, 6463-6475
  45. Stoddard BL (2005) Homing endonuclease structure and function. *Q Rev Biophys* 38, 49-95
  46. Petek LM, Russell DW and Miller DG (2010) Frequent endonuclease cleavage at off-target locations in vivo. *Mol Ther* 18, 983-986
  47. Boissel S, Jarjour J, Astrakhan A et al (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res* 42, 2591-2601

48. Gaj T, Mercer AC, Sirk SJ, Smith HL and Barbas CF 3rd (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res* 41, 3937-3946
49. Gersbach CA, Gaj T, Gordley RM and Barbas CF 3rd (2010) Directed evolution of recombinase specificity by split gene reassembly. *Nucleic Acids Res* 38, 4198-4206
50. Sarkar I, Hauber I, Hauber J and Buchholz F (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science* 316, 1912-1915
51. Hacein-Bey-Abina S, Von Kalle C, Schmidt M et al (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302, 415-419
52. Wu X, Li Y, Crise B and Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300, 1749-1751.
53. Lim KI (2012) Retroviral integration profiles: their determinants and implications for gene therapy. *BMB Rep* 45, 207-212
54. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR and Bushman F (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110, 521-529
55. Lim KI, Klimczak R, Yu JH and Schaffer DV (2010) Specific insertions of zinc finger domains into Gag-Pol yield engineered retroviral vectors with selective integration properties. *Proc Natl Acad Sci U S A* 107, 12475-12480