Charles University in Prague, Faculty of Science
Special Chemical and Biological Programmes:
Molecular Biology and Biochemistry of Organisms
Department of Genetics and Microbiology

Radmila Hanečková

Genome editing using programmable endonucleases

Editace genomu pomocí programovatelných endonukleáz

Bachelor thesis

Supervisor: doc. RNDr. Radislav Sedláček, Ph.D.
Consultant: Mgr. Petr Kašpárek

Laboratory of Transgenic Models of Diseases
Institute of Molecular Genetics of the ASCR
Prague, 2014

I would like to thank my supervisor Radislav Sedláček for making me a part of his team, and my consultant Petr Kašpárek for his guidance and the trust he has placed in my abilities. I would also like to thank my family, my partner and my friends for their patience and support, and my grandfather for instilling in me a love for science that remains strong to this day.

Prohlášení:
Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 26. 4. 2014

Radmila Hanečková

Abstrakt

Programovatelné endonukleázy jsou proteiny schopné rozpoznat specifické sekvence nukleotidů a následně v rozpoznané sekvenci štěpit obě vlákna DNA. Zinc-finger nukleázy jsou široce využívaným nástrojem v genomové editaci, metodě zavádění změn do genomů buněčných liní nebo celých organismů za účelem studia funkce genů. Nedávno se objevily nové typy programovatelných endonukleáz v podobě transcription activator-like effector (TALE) nukleáz a CRISPR/Cas systému. Tyto systémy se od sebe liší z pohledu mechanizmu funkce, dostupnosti, selektivity, frekvence off-targetů a cytotoxicity. V této práci zinc-finger nukleázy, TALE nukleázy a CRISPR/Cas systém porovnáváme a zkoumáme jejich aktuální a možné budoucí využití v široké oblasti výzkumu, který sahá od vývoje geneticky modifikovaných organismů až po genovou terapii.

Klíčová slova:
genetické inženýrství, editace genomu, zinc-finger nukleázy, TALEN, CRISPR

Abstract

Programmable endonucleases are engineered proteins that recognize specific nucleotide sequences and that are capable of introducing double-strand breaks within these sequences. Zinc-finger nucleases have been used extensively as a tool in genome editing, the practice of introducing changes into genomes of cell lines or whole organisms as a way to study gene function. Recently, new types of programmable endonucleases have emerged in the form of transcription activator like effector (TALE) nucleases and the CRISPR/Cas system. The types differ in respect to their mechanism of function, accessibility, selectivity, frequency of off-target cleavage and cytotoxic effects. Here, we compare zinc-finger nucleases, TALENs and the CRISPR/Cas system and explore their current and possible future applications in a broad spectrum of research ranging from developing genetically modified organisms to gene therapy.

Keywords:
genetic engineering, genome editing, zinc-finger nucleases, TALEN, CRISPR

Glossary

CRISPR	clustered regularly interspaced short palindromic repeat
DSB	double-strand break
ESC	embryonic stem cell
HR	homologous recombination
NHEJ	non-homologous end-joining
PAM	protospacer adjacent motif
RVD	repeat-variable di-residue
TALEN	transcription activator-like effector nuclease
ZFN	zinc-finger nuclease

Table of contents

1. Introduction

Genome editing (or genome engineering) is the practice of introducing specific changes into genomes of model organisms or cell lines. It is a widely used and tried approach to understanding gene function through alteration, removal, fusion with a reporter tag and even replacement with a different gene. There are numerous engineered knock-out or knock-in model organisms and cell lines available to researchers now, for example mouse strains with mutations mirroring those found in humans displaying hereditary diseases are serving as models to elucidate the molecular mechanism behind the phenotype. They provide valuable insight into possible therapy or disease prevention and, most importantly, help understand the complexity of the eukaryotic organism. Biotechnology is greatly profiting from the ability to introduce or remove certain traits in order to engineer expression systems for protein production or increase the agricultural yield of crops. Gene therapy also relies on altering the genetic make-up of patient-derived cell lines in order to use them as a therapeutic agent.

In order to introduce changes into the genome reliably, efficiently, and lastingly, it is desirable to target specific loci within the genome. But to do this safely in for example the human genome (which comprises approx. 3.2 Gbp), the targeted DNA sequence has to be at least 16 bp in length to ensure its uniqueness within the genome. Statistically, a 16 bp long sequence should occur just once within a genome of the size of 4^{16} bp (4.5 Gbp), assuming the bases are randomly distributed. In comparison, type II restriction endonucleases generally recognize sequences only 4–8 bp in length, and their use as a means to introduce double-strand breaks (DSBs) on unique sites in large genomes is severely limited.

The introduction of DSBs is a crucial step in genome editing. DSBs on genomic DNA are repaired by the two prevalent DNA damage repair pathways, non-homologous end-joining (NHEJ) and homologous recombination (HR) (Takata *et al.* 1998). If a DSB is repaired by NHEJ, this error-prone mechanism frequently leads to indels at the site of the DSB. Thus, if the DSB is targeted into a specific locus within an exon of a chosen gene, repair by NHEJ can lead to frameshift-inducing indels and therefore a truncated or absent product. Alternatively, larger deletions can be induced by the introduction of two DSBs flanking a longer

sequence, leading to deletion of the entire flanked segment. Another repair mechanism endogenous to most eukaryotes is homologous recombination (HR) — it requires a template in the form of DNA homologous to the target site, and though it does not not need to be preceded by the introduction of a DSB, its frequency greatly increases if a DSB is present. Defined mutations can be generated by providing a template for homologous recombination (HR), and large fragments of DNA can be inserted into the genome via HR in a similar fashion.

Programmable endonucleases are a powerful tool that enables researchers to introduce DSBs in pre-selected sequences within the genome. Zinc-finger nucleases (ZFNs) have been around for a few decades, but two alternatives have emerged recently and are being developed intensively: Transcription activator-like effector nucleases (TALENs) and the recently utilized CRISPR/Cas system are becoming indispensable tools for genome editing. The nucleases can be engineered to recognize a specific sequence of sufficient length to be unique within even a large genome. Already they have enabled the development of many previously unattainable genetically modified animals and plants from a variety of species, thus aiding several fields of research. Lasting changes to the genome have a distinct advantage over gene-knockdown by RNAi, since introducing changes on the DNA level rather than the transcriptional level prevents the targeted gene being leaky and small amounts of the product still being present within targeted cells (Kim *et al.* 2011). In addition, genome editing allows for reproducible and reliable results.

Here, we examine the mechanism of function of zinc-finger nucleases, TALE nucleases and the CRISPR/Cas system in order to attempt a comparison of these three most widely used programmable endonucleases in respect to their accessibility, effectivity, off-target cleavage frequency and cytotoxicity. We further explore their potential as genome editing tools and their current and possible future applications in the generation of genetically modified animal models and cell lines, and their potential in biomedicine. Finally, we discuss the shortcomings of each particular method, and offer suggestions for future research.

2. Zinc-finger nucleases (ZFNs)

The Cys_2-His_2 zinc-finger domain is a DNA-binding motif commonly present in mammalian transcription factors (Miller *et al.* 1985). Each domain binds a zinc ion, displays a ββα fold and is able to recognize and bind to a 3 bp sequence. The α-helix inserts into the major groove, and due to variations in the amino-acid residues present is able to recognize various combinations of base pairs within the groove (Figure 1). Naturally occurring zinc-finger proteins do not recognize more than 9 bp long sequences using three domains. Although the number of domains present in a zinc-finger protein can be higher, not more than three domains participate in the binding itself (Gupta *et al.* 2014). Based on the structure of Zif268 bound to DNA obtained using x-ray crystallography, so-called polydactyl zinc-finger proteins containing several zinc-finger domains linked together were designed, thus obtaining an artificial DNA-binding domain capable of recognizing and binding longer sequences (Liu *et al.* 1997). Since the binding properties of each individual zinc-finger domain are not entirely stand-alone but rather co-determined by the adjacent domains, methods that enable selection of efficient zinc-finger arrays had to be implemented.

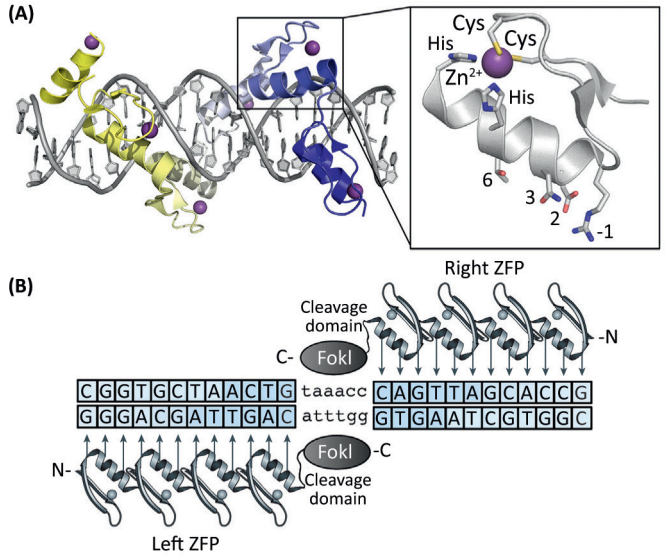

Figure 1. Zinc-finger nuclease architecture. **(A)** Strucure of a zinc-finger array bound to its target DNA with close-up of a single zinc-finger domain, showing the amino acid residues bound to a zinc ion and the four residues (-1, 2, 3, 6) responsible for contacting base pairs within the major groove. **(B)** A pair of zinc-finger nucleases (ZFNs) recognizing a target sequence with the Fok1 nuclease domain bound to the C-terminal end. Adapted from Gaj *et al.* 2013.

Eventually, libraries of zinc-finger domains targeting 5' GNN 3', 5' ANN 3' and 5' CNN 3' were constructed (Gonzalez *et al.* 2010). The modular assembly methods rely on these libraries to construct multi-finger proteins and then using a suitable selection system (such as the bacterial two-hybrid system) to obtain zinc-finger constructs with high binding specificity to the intended target region. These methods treat individual zinc-fingers as units. Sigma-Aldrich has undertaken large-scale research into zinc-finger develope-

ment and now offers custom zinc-fingers made using their proprietary modular assembly technique as a commercial product ("CompoZr® Custom ZFN Service | Sigma-Aldrich" 2014). Another assembly and selection method is the OPEN protocol (oligomerized pool engineering) — this combinatorial approach utilizes a limited library of individual zinc-finger domains and takes into account which individual zinc-fingers work well together when assembled into arrays. The OPEN protocol claims that it should be possible to obtain three-finger zinc-fingers in 8 weeks, the method however does not always yield functional products (Maeder *et al.* 2009).

2.1 The FokI nuclease domain

Fusing an additional effector domain to the C-terminal end of a zinc-finger array couples the target specificity of the zinc finger array with the enzymatic activity of the selected effector domain (Kim *et al.* 1996). The type IIS FokI endonuclease derived from *Flavobacterium okeanokoites* consists of an N-terminal DNA-binding domain and a C-terminal domain exhibiting non-specific DNA cleavage activity (Wah *et al.* 1998). The cleavage domain retains its function when fused to a different DNA-binding protein, such as a zinc-finger array (Durai 2005). Dimerization of two cleavage domains is necessary for successful double-strand breaks to occur (Bitinaite *et al.* 1998). A pair of such hybrid zinc-finger nucleases (ZFNs) is thus required for targeted DNA cleavage, the zinc-fingers binding to sequences flanking the desired target site, enabling the FokI nuclease domain to dimerize and cleave the spacer region between them.

3. Transcription activator-like effector nucleases (TALENs)

Transcription activator-like effectors (TALEs) are virulence factors belonging to members of the bacteria genus *Xanthomonas*, enabling these plant pathogens to infect their host and induce changes beneficial to the pathogen. Injected into the plant cell, they act as transcription factors within the nucleus, activating gene expression of the host (Bai *et al.* 2000, Romer *et al.* 2007).

Members of the TALE family contain a nuclear localization signal, an acidic transcriptional activation domain and a central domain. The central domain is composed of tandem repeats which differ in number and amino-acid composition between the members of the

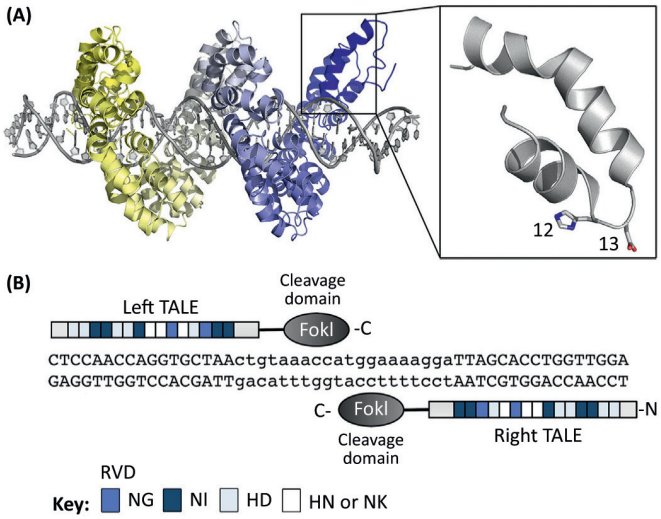

Figure 2. TALE nuclease architecture. **(A)** A TALE domain bound to its target DNA with close-up of a single subunit containing the two repeat-variable di-residues (RVDs) on positions 12 and 13. **(B)** Schematic drawing of a pair of TALE domains with the Fok1 nuclease domains bound to the C-terminal end, dimerizing over a spacer between the recognized sequences. The subunits within the TALE domain each carry a different RVD, allowing it to recognize the corresponding nucleotide. Adapted from Gaj *et al.* 2013.

TALE family and facilitate binding of the TALE to its corresponding up-regulation target. The mechanism of specificity has been uncovered by Boch *et al.* by studying the *Xanthomonas campestris pv. vesicatoria* TAL effector AvrBs3 (Boch *et al.* 2009) — its central domain contains tandem repeats of 34 highly conserved amino acids, with the exception of the amino acids in positions 12 and 13, which display a high level of polymorphism and are called repeat-variable di-residues (RVDs). It has been shown that these RVDs within a TALE repeat recognize and bind a single base pair within the target sequence of the effector following a simple code (Figure 2). The RVDs consisting of the amino acids NG recognize thymine, NI recognize adenine, HD recognize cytosine and both HN and NK recognize guanine with variable accuracy (Moscou and Bogdanove 2009). A known limitation as to the base composition of the target site is the requirement for a 5' T preceding the first RVD-associated base (Gaj *et al.* 2013, Kim *et al.* 2013). Shortly after the TALE code was deciphered, custom engineered TALEs bound to the catalytic domain of FokI have been created, functioning in a similar way as the zinc-finger FokI hybrids — introducing targeted DSBs via a pair of transcription activator-like effector nucleases (TALENs) able to bind to specific sequences upstream and downstream of the intended cleavage site, mediating cleavage within the spacer region by enabling the FokI nuclease to dimerize (Christian *et al.* 2010). Other catalytic domains have been fused to the TALE domain, including recombinases (Mercer *et al.* 2012).

There are several methods available for designing and assembling custom TALE domains. Since the individual tandem repeats within a TALE function independently, custom TALE domains can be designed *in silico* using software freely available online (Doyle *et al.*

2012). One of them utilises the Golden Gate assembly method, allowing for simultaneous joining of several DNA fragments in one reaction (Cermak *et al.* 2011). Type IIS restriction endonucleases cleave several bp upstream of the sequence they recognize, thus one IIS restriction endonuclease can be used to create a variety of different sticky ends in one reaction. A library of plasmids that carry the sequence coding for each of the four RVDs is required. These are further subdivided in groups that correspond to positions within the TALE array and if cleaved by the IIS restriction nuclease result in sticky ends that allow them to be ligated in the desired sequence. Thus, mixing the plasmids from the correct group and type of RVD, digesting and ligation allows for the assembly of the entire TALE domain in one reaction. Further cloning is required to add the C-terminal and N-terminal domains containing the nuclear localization signal and the FokI nuclease domain. Cermak *et al.* claim that this method yields functional TALENs in five days. Another available method of assembly is FLASH, which has been optimized for high-throughput automation (Reyon *et al.* 2012). FLASH relies on a large library of plasmids containing pre-assembled multimers of TALE repeats of various length, with the first repeat bound to a biotin marker that adheres to magnetic beads coated with streptavidin. Another recently developed high-throughput method uses ligation-independent cloning to assemble the TALE repeats via long single-strand overhangs (Schmid-Burgk *et al.* 2012).

4. The CRISPR/Cas system

Some species of bacteria and archaea possess an adaptive defensive mechanism against viral infection. Unlike the well-known restriction-modification system, the mechanism in question functions very similarly to RNA-interference described in eukaryotes and displays the capability of acting as an adaptive immune system (Wiedenheft *et al.* 2012). If an intracellular parasite (such as a phage, but also a plasmid) invades the bacterial or archaeal host cell, the host cell is capable of storing short fragments of the foreign nucleic acid originating from the parasite in a designated locus within its own genome. These loci, termed clustered regularly interspaced short palindromic repeats (CRISPRs) have been discovered in several prokaryotic species, but their function remained unknown for several years. They consist of an AT-rich leader sequence, followed by direct repeats around 20–50 bp long, separated with spacers of varying sequence but similar length (Figure 3 A). Recently, it has been discovered that the CRISPR locus gets transcribed and processed into short RNA segments

termed crRNAs. The previously encountered pathogen-derived nucleic acid that has been incorporated into the CRISPR locus each becomes part of an individual crRNA. A complex forms that incorporates the crRNA and uses it as a guide for cleavage of invading nucleic acids. Proteins required for processing and cleavage are encoded upstream of the CRISPR locus. These CRISPR-associated (*cas*) genes can vary in number and function — several types of CRISPR-mediated systems have been described in bacteria and archaea that differ in mechanism of function and contain a different set of associated *cas* genes.

Different Cas proteins in the different systems described take over the steps necessary to ensure integration of new spacers into the CRISPR locus, transcribing and processing the primary transcript into crRNAs and crRNA binding, recognition and finally cleavage of the recognized foreign nucleic acid (Figure 3 B). The simplest family of systems is the type

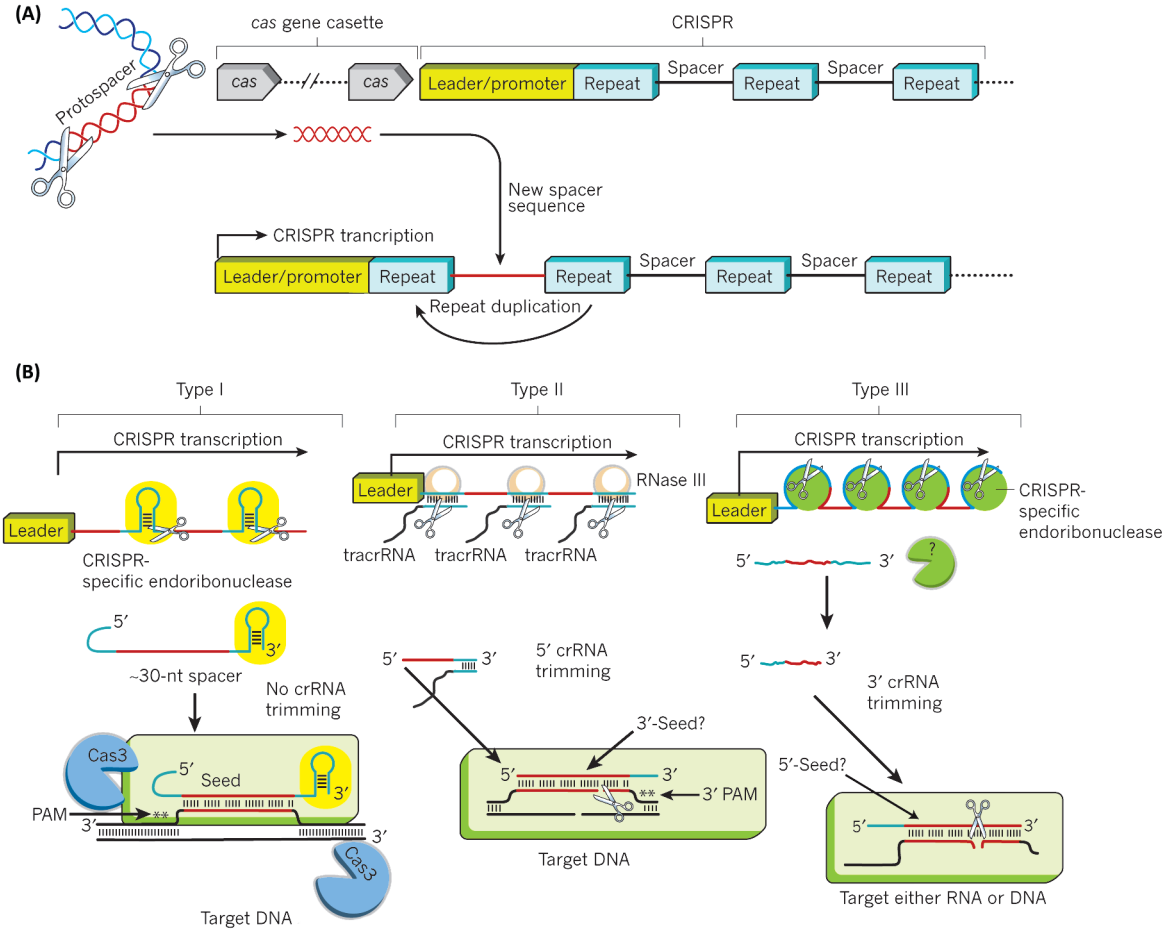

Figure 3. The CRISPR locus architecture and three types of the CRISPR/Cas system. **(A)** The clustered regularly interspaced short palindromic repeat (CRISPR) locus contains a cas gene cassette with a variable number of cas genes, a leader/promoter sequence and several repeats flanking variable spacers. Foreign (protospacer) DNA gets processed and inserted between repeats. **(B)** Mechanisms of function fall ino three types that differ in respect to number and types of enzymes involved in CRISPR transcription, RNA processing and protospacer recognition and cleavage. Adapted from Wiedenheft *et al.* 2012.

II, consisting of four different *cas* genes. The primary transcript composed of spacers and direct repeats (also called pre-crRNA) gets processed by RNase III. This cleavage is guided by so-called trans-acting antisense RNA (tracrRNA), which is complementary to the spacers within the CRISPR locus and also encoded within the locus. Further processing yields crRNA which is then free to bind to Cas9, a product of one of the three *cas* genes present in the type II system. It is a large protein with multiple catalytic sites, displaying nuclease as well as helicase activity (Sternberg *et al.* 2014).

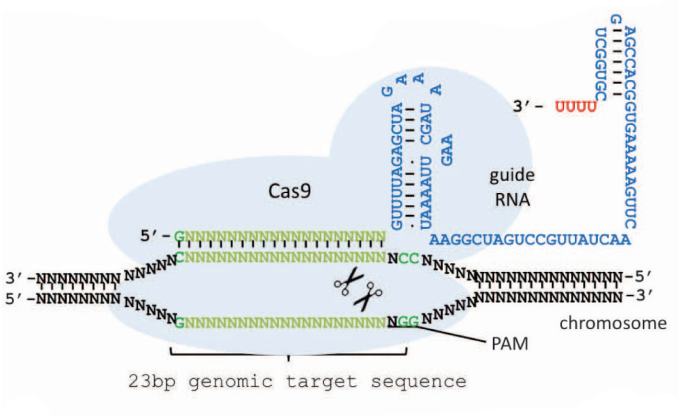

Figure 4. The Cas9-gRNA complex recognizing its target DNA sequence. The protospacer adjacent motif (PAM) is located 3' downstream of the genomic target sequence on the DNA strand complementary to the one recognized by gRNA. Adapted from Mali *et al.* 2013.

Mali *et al.* took the CRISPR type II locus from the bacteria *Streptococcus pyogenes SF370* and successfully attempted to adapt the procaryotic immune system to function as a programmable tool for introducing double-strand breaks in eukaryotic chromosomes *in vivo* (Mali *et al.* 2013). To express tracrRNA and pre-crRNA (containing two direct repeats flanking a single spacer) in the eucaryotic cell, the U6 snRNA promotor recognized by Pol III can be used. This ensures that the transcribed RNA doesn't undergo capping, polyadenylation, and subsequent export from the nucleus as mRNA but instead remains within the nucleus. Cas9 and RNase III genes have been modified to account for the codon bias difference and equipped with nuclear localization signals, so that the proteins are imported into the nucleus where they encounter the tracrRNA and pre-crRNA. Cong *et al.* have demonstrated that tracrRNA and pre-crRNA can be fused together into a chimeric guide RNA (gRNA) which is sufficient for Cas9 recognition and target sequence cleavage (Figure 4), thus eliminating the need for RNase III altogether (Cong *et al.* 2013). This two-component system is very flexible in terms of target sequence. The only requirement for Cas9 to function properly in relation to the gRNA supplied is the presence of a protospacer adjacent motif (PAM) located 3' from the target sequence (which is referred to as the protospacer in some of the literature), in the form of 5' NGG 3'. Mali *et al.* used 5' $GN_{20}GG$ 3' as target sequence, the 5' G not being crucial. The PAM however seems to be absolutely necessary for Cas9 target sequence (protospacer) recognition.

Mashiko *et al.* have succeeded in introducing indels via NHEJ in targeted loci (Cetn1 and Prm1) in mouse zygotes by microinjection of all the components required for CRISPR/Cas-mediated cleavage (Mashiko *et al.* 2013). They used a circular plasmid that contained both the *hCas9* (humanized Cas9) expression cassette and the guide-RNA (gRNA) expression cassette, thus demonstrating that it is not necessary to introduce Cas9 in the form of RNA. A high efficiency has been achieved with this method: they claim to have obtained 40–60% knock-out animals, even several homozygotes with mutations on both alleles in the *Cetn1* locus. Subsequent screening for *hCas9* gene integration into the genome yielded a negative result. The CRISPR/Cas system seems to be emerging as a new alternative to both zinc-finger nucleases and TALE nucleases.

5. Comparison of ZFNs, TALENs, and the CRISPR/Cas system

The programmable nucleases described before all share the main principle of function — recognition of a specific DNA sequence and subsequent introduction of a double-strand break adjacent to the recognized sequence. However, the systems differ greatly in several aspects: While ZFN subunits recognize 3 bp each, TALE repeats each correspond to a single base-pair of the recognized sequence. Furthermore, neighboring ZFN subunits influence each other, while TALE repeats constitute independent units, which has implications for both the ease with which each type of programmable nuclease can be designed and assembled and the specificity with which it binds to its target sequence. While both ZFNs and TALENs bind in pairs, allowing the FokI nuclease domain to be activated by dimerization, the CRISPR/Cas system is based on a single RNA-binding protein that mediates both recognition and cleavage of the target sequence. This difference in mechanism causes differences in off-target cleavage activity and influences target recognition specificity. Finally, the RNA-reliant mechanism of the CRISPR/Cas system offers an advantage over both ZFNs and TALENs by making multiplex genome engineering possible.

5.1 Differences in respect to mechanism, accessibility, and flexibility

While Zinc-finger nucleases (ZFNs) have proven to be reliable, their main disadvantage is their relatively low flexibility and cost of production. Since the individual domains recognize a 3 bp sequence, a large library of variations of ZF domains is necessary to cover all the possible combinations and thus allow for a desired sequence to be targeted. From this

library, an optimal combination has to be selected and further screening and selection is required to find those arrays that bind to the targeted locus. A further complication results from the fact that the binding of individual subunits to their target DNA is influenced by the adjacent subunits within the ZF array. This labor-intensive selection and screening process can be avoided by directly ordering custom ZFNs as a commercial product, but this keeps the overall cost of the method relatively high and thus renders it inaccessible for many researchers with a limited budget at their disposal.

Transcription activator-like effector nucleases (TALENs) however have the distinct advantage of each of the individual subunit being identical with the exception of the two repeat-variable di-residues (RVDs) within each repeat. Since the RVDs bind specifically to a single base-pair within the major groove of the target DNA, only four different TALE subunits are necessary. Also, as opposed to ZF subunits, TALE repeats seem to act as individual units. These two factors combined render TALENs more flexible in terms of targeting a specific sequence and essentially eliminate the necessity for binding specificity and efficiency screening. There are accessible online tools available for *in silico* TALE design, which when provided with a desired target sequence will automatically assemble a series of possible TALE repeats, taking into account the requirement of a 5' T preceding the RVD-associated base. Using one of the assembly methods described, such as the combinatorial DNA assembly method Golden Gate, enables researchers to design, assemble, verify and use their own TALENs routinely with relative ease and without need to purchase pre-assembled final products - this keeps both cost and labor-intensity of TALENs (as opposed to ZFNs) relatively low.

TALENs and ZFNs both rely on the same principle of function — both act as pairs of multi-subunit proteins containing arrays of DNA-binding domains recognizing sequences that flank the intended cleavage site, with FokI nuclease domains fused to the C-terminal ends that dimerize and cleave the target. The CRISPR/Cas system mechanism of function however shows similarity to RNA interference (RNAi), a pathway found in eukaryotes that inhibits gene expression on the mRNA level. Part of the RNAi pathway is the RISC complex, which incorporates siRNA complementary to the targeted mRNA which is then cleaved by the RISC complex (Tijsterman and Plasterk 2004). Unlike ZFNs and TALENs, where both target recognition and cleavage are mediated by protein subunits, the CRISPR/Cas system acts as a protein-RNA complex. The recently determined 3D structure of Cas9 from *Strep-*

tococcus pyogenes and *Actinomyces naeslundii* using x-ray crystallography reveal that it displays a bi-lobed architecture, where one lobe contains the nuclease domains (Figure 5). Without guide-RNA (gRNA), the enzyme adopts an auto-inhibited state in which the central channel that enables Cas9 to bind and cleave its target DNA remains inaccessible (Jinek *et al.* 2014).

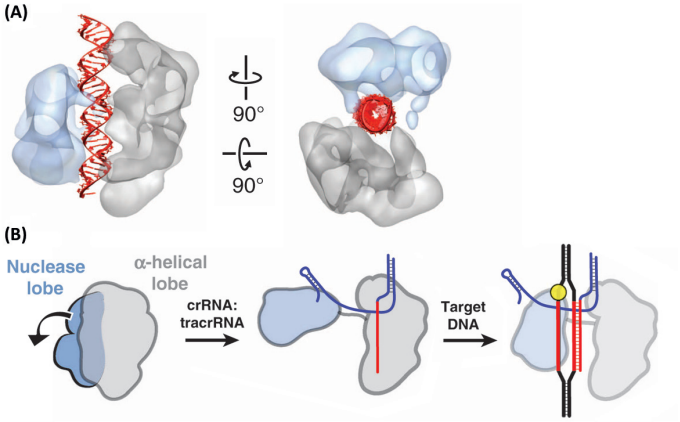

Figure 5. Cas9 structure and conformational change induced by the binding of crRNA and target DNA. **(A)** The channel formed by the nuclease lobe and the α-helical lobe envelops the RNA-DNA duplex. **(B)** Proposed model for Cas9 activation. The protospacer adjacent motif (PAM) is depicted as a yellow circle. Adapted from Jinek *et al.* 2014.

Another distinct feature of the CRISPR/Cas system is the necessity of the 3 bp protospacer adjacent motif (PAM) located downstream from the target sequence (Sternberg *et al.* 2014). Sternberg *et al.* have demonstrated that this tri-nucleotide sequence is unexpectedly crucial in all of the steps leading to DNA cleavage. Cas9 displays significantly higher binding lifetimes to DNA containing its respective PAM (5' NGG 3') even if these are not preceded by a sequence complementary to the Cas9-bound gRNA (also referred to as crRNA). gRNA-complementary sequences without an adjacent PAM do not display higher binding frequency. This indicates that Cas9 uses the PAM as its primary way to seek out its target via preferentially binding to PAMs. The recognition is highly selective, even a single mutation (for example 5' NCG 3') will prevent cleavage of the target site, even if the gRNA bound to Cas9 is fully complementary to the DNA sequence preceding the mutated PAM. Interestingly, though the Cas9-gRNA complex displays similar binding affinities with target ssDNA and dsDNA, ssDNA is cleaved much slower, by several orders of magnitude. The PAM is recognized on the non-complementary strand to the gRNA bound to Cas9, thus ssDNA contains only its complement. This stringent regulation by the PAM might function as protection from self-cleavage of the bacterial genome within the CRISPR locus, since the spacers there are not followed by a PAM. It does however somewhat limit the flexibility of the CRISPR/Cas system as a tool for genome editing.

5.2 Specificity and off-target effects

A very important attribute of all the discussed programmable endonucleases is their selectivity with respect to target cleavage. Both ZFNs and TALENs work in pairs, their respective recognized target sequences flanking a spacer (around 10 bp with ZFNs and 20 bp with TALENs) above which the FokI nuclease domains dimerize. The length of recognized sequences used as targets by individual ZFNs and TALENs is very similar, around 18 to 20 bp (individual ZF domains recognize tri-nucleotides, thus the recognized sequence has to be a multiple of three). Statistically, this length is sufficient for the sequence to be unique in a genome of the size corresponding to the most widely used eukaryotic model organisms. This fact, combined with the necessity of pairs doubling the sequence targeted and the need of FokI nuclease domains to dimerize in order to be able to cleave acts as a safeguard against ZFNs and TALENs cleaving elsewhere within the genome.

Employing bioinformatics methods in the target site selection process is another option to avoid or at least significantly reduce off-target activity. Kim *et al.* for example have utilized this strategy as a part of their effort to assemble a TALEN library spanning the human genome by screening for possible off-target sites with a suitably tailored algorithm for TALEN design, selecting target sequences that do not have partial homologies elsewhere within the genome (Kim *et al.* 2013). Since both designing and assembling TALENs as opposed to ZFNs is a more straightforward process, implementing this kind of off-target screening should prove more simple, even though there is no guarantee that it eliminates all off-target activity, since it does not account for differences in the target genome to the referenced genome available in the database, and other, yet undiscovered factors.

The Cas9-gRNA complex has been successfully used to target 20 bp sequences preceding a PAM (Mali *et al.* 2013). The conformational change induced by the binding of Cas9 to gRNA, the protospacer target DNA and PAM combined lead to activation of Cas9 nuclease domains and thus induce cleavage. In this case, both nuclease domains are active sites of a single protein, Cas9, and no dimerization is therefore necessary as is the case with TALEs and ZFs fused to FokI nuclease domains. According to Jinek *et al.*, Cas9 is auto-inhibitory in its unbound state, which is corroborated by their x-ray crystallography structure data on unbound apo forms of SpyCas9 and AnaCas9 (Jinek *et al.* 2014). Sternberg *et al.* emphasize the role of PAM and propose the theory that it acts as an allosteric regulator of the nuclease activity of the Cas9-gRNA complex (Sternberg *et al.* 2014). This generates a potential risk

of off-target cleavage within PAM-rich regions of the genome, since PAM-Cas9 interaction is implied in destabilizing the structure and exposing the protospacer. Since the Cas9-gRNA complex displays very low binding affinity to a sequence that is complementary to the gRNA but lacks an adjacent PAM sequence, the fact that the CRISPR/Cas system relies on a shorter recognition sequence as opposed to ZFNs and TALENs might not turn out to be a major factor in relation to off-target cleavage. An interesting approach to mitigate off-target effects of the CRISPR/Cas system and at the same time retain the advantage of the flexible RNA-guided binding mechanism has been explored recently: Tsai *et al.* attempted to construct dimeric RNA-guided FokI nucleases (RFNs) that combine the RNA-based DNA recognition domain of Cas9 with the FokI nuclease domain used by ZFNs and TALENs to allow for a more stringent control over DNA cleavage (Tsai *et al.* 2014).

Off-target effects along with toxicity and other adverse effects resulting from introducing the programmable endonucleases into cells could be partly attributed to the method of delivery. There are several methods available, using plasmid or viral vectors carrying the genes encoding the respective pair of ZFNs or TALENs, via lipofection, electroporation or microinjection. Each delivery method has its advantages and specificities, rendering it applicable in one case and unsuccessful in another. For example, lentiviral vectors as a means to deliver TALENs into cells have failed to perform as well as adenoviral vectors in one study (Holkers *et al.* 2013). Not every cell type is suitable for the introduction of DNA with the aforementioned methods, and the presence of foreign DNA within the cell can have adverse side-effects such as increased off-target effects due to random integration of the introduced DNA into the genome. This can be countered by introducing the nucleases in the form of *in vitro* prepared mRNA, or avoiding nucleic acids entirely and introducing purified proteins directly (Gaj *et al.* 2012). The necessity to prepare mRNA or purify proteins however carries its own specific disadvantages (such as the increased care necessary to avoid RNA degradation prior to delivery), and the limitations of available delivery systems are a major factor in restricting the use of all three described types of programmable endonucleases.

5.3 Multiplex genome engineering

Modifying two or more loci within the genome of one single organism has the potential to provide insight into gene interaction and redundancy in respect to function. Instead of

crossing specimen carrying a single mutation each in order to obtain the desired genotype, multiplex genome engineering allows for both loci to be modified simultaneously in one organism and thus eliminates the necessity for laborious and time-consuming breeding. Here, the architecture of the CRISPR/Cas system seems to be the most advantageous, since unlike ZFNs and TALENs, which require a pair of enzymes for every targeted locus, Cas9 is a single enzyme that can be supplied with a variety of guide RNAs (gRNAs) that mediate targeted cleavage at several loci simultaneously. Cong *et al.* have successfully used two gRNAs targeting two protospacers situated close to each other, which led to excision of the protospacer-flanked segment in-between (Cong *et al.* 2013). Multiple independent genome loci have been targeted simultaneously by using Cas9 in combination with several gRNAs in the zebrafish by Jao *et al.* and in the monkey *Macaca fascicularis* by Niu *et al.* (Jao *et al.* 2013, Niu *et al.* 2014).

6. Examples of current and possible future applications

The emergence and subsequent development of programmable endonucleases has greatly impacted several fields: Zinc-finger nucleases (ZFNs) and their potential as genome editing tools have contributed to the mouse being established as the most widely used complex eukaryote animal model. TALENs and the CRISPR/Cas system are offering new possibilities as a faster and more affordable means to generate genetically modified mouse strains, but their increased efficiency has had a far greater impact by allowing researchers to produce previously unattainable genetically engineered specimen of a vast variety of animal and plant species. Biotechnology can profit from employing programmable nucleases as a means to genetically alter economically relevant livestock and crops, and there is potential for clinical application of highly precise programmable nucleases in gene therapy.

6.1 Mouse genome editing

Mouse genetics in particular is profiting from the opportunities offered by programmable endonucleases. The mouse is a widely used model organism, with many applications spanning basic research to biomedicine. The mouse genome was sequenced in 2002, and today many transgenic and knock-out strains are available to researchers (Chinwalla *et al.* 2002). To provide insight into gene function, two international initiatives have been launched with the aim of providing a repository of mouse strains harboring mutations in

every gene — the Knock-Out Mouse Project KOMP (Austin *et al.* 2004) and European Conditional Mouse Mutagenesis EUCOMM (Auwerx *et al.* 2004).

A significant fraction of currently available genetically modified mouse strains have been created via modifying embryonic stem cells (ESCs) and pronuclear injection of DNA into zygotes (Glaser *et al.* 2005). In order to establish a homozygous strain carrying the desired mutation, ESCs that have been modified via transfection, electroporation or microinjection of a suitable DNA construct are added to blastocysts which are then transferred surrogate mothers producing chimeric offspring in the first generation. Through backcrossing a suitable founder that produces gametes carrying the mutation, a stable strain can be established. The alternative method consists of modifying zygotes by microinjecting the DNA construct into the male pronucleus (Ittner and Götz 2007). Both methods heavily rely on the mechanism of homologous recombination (HR) as means to alter the recipient genome and therefore have been held back by the low frequency with which HR occurs naturally. Since the introduction of a double-strand break significantly increases the HR frequency, the use of programmable endonucleases is having a great impact on mouse genome engineering. Whereas five years ago, establishing a strain could take several years, now biallelic founders carrying the desired mutations can be obtained within weeks (Mashiko *et al.* 2013).

Double knock-out mouse models have generally been obtained by crossing single knock-out strains with the desired mutation — a strategy that is very difficult to implement in the case of both genes being situated on the same chromosome. Due to infrequent crossing-over events, it is possible to obtain offspring carrying both mutations in a large enough cohort. If the distance between the genes is too small to warrant a high enough crossing-over frequency however, this strategy cannot be used — an alternative path is being made possible by employing TALENs and the CRISPR/Cas system as a means to induce both mutations simultaneously, relying on their high efficiency. Thus, new previously unattainable mouse models could be generated utilising highly effective programmable endonucleases for multiplex genome engineering.

Another interesting possibility programmable nucleases offer in this respect is targeted transgenesis (DeKelver *et al.* 2010). The introduction of exogenous genes into selected cell lines or organisms is a laborious process largely due to low efficiency when relying on

introducing the transgene randomly into the genome. Perez-Pinera et al. have engineered ZFNs targeting a well-studied mouse safe-harbor locus, ROSA26 (Perez-Pinera *et al.* 2012). Integration of a transgene into this locus ensures its ubiquitous expression over most tissues and cell types and prevents disruption of an endogenous pathway. By using ZFNs to introduce DSBs within the ROSA26 locus and co-transfecting the cells with a vector containing a multiple-cloning site (MCS) flanked by 800 bp fragments that were homologous to the sequences flanking the ZFN target site, Perez-Piera *et al.* have greatly increased the frequency of homologous recombination (HR), leading to integration of the MCS into the ROSA26 locus. They further showed that if an expression cassette containing EGFP fused to Zeomycin resistance was inserted into the MCS of the vector, it could be introduced into the ROSA26 locus this way. The yield of positive transgenic cells has been assayed by Perez-Piera et al. using flow cytometry and estimated at 10%. TALENs and the CRISPR/Cas system could be used in to facilitate gene targeting into safe-harbor loci in a similar fashion.

6.2 Generation of novel genetically modified organisms

Programmable endonucleases have not only been successfully used in a vast variety of model organisms such as mouse, rat, fruit fly and zebrafish, but also in economically relevant crops and livestock including rice, maize, tobacco, and pig, displaying potential not only in basic research but also applied biotechnology (Urnov *et al.* 2010, Gaj *et al.* 2013, Sun and Zhao 2013). The number of species successfully targeted with programmable endonucleases is rapidly increasing, with new species and new genes being added to the repertoire. To a large extent, programmable endonucleases are what drives this increase.

One of the reasons why the mouse has become such a widely used model organism is due to the availability of mouse embryonic stem cells (ESCs) *in vitro*, and the fact that these can be induced to mature into cell-culture derived animals (Nagy *et al.* 1993). This allowed for generating genetically modified mouse strains from ESCs, but restricted the method to only those species from which ESCs can be cultured and induced to mature into fully developed animals. With the development of programmable endonucleases however, an alternative has emerged in the form of directly modifying zygotes via DNA microinjection as a way to introduce mutations with high efficiency, thus removing the necessity of culturing and modifying ESCs. A notable breakthrough is the generation of targeted knock-out rats using ZFNs (Geurts *et al.* 2009). Rats are better suited than mice for several research

applications, not only due to their larger size, but also due to their highly developed cognitive abilities — a notable example of a field that could greatly benefit from the availability of transgenic rat strains is neurobiology and optogenetics (Zalocusky and Deisseroth 2013). One of the most recent additions to the compendium of successfully genetically modified species using programmable endonucleases is the cynomolgus monkey, *Macaca fascicularis* — monkeys have resisted the use of both ZFNs and TALENs as means to induce genetic modifications, but Niu *et al.* report to have successfully introduced mutations in two genes (*Ppar-γ* and *Rag1*) by co-injecting Cas9 mRNA with gRNA targeting several genes into fertilized eggs (Niu *et al.* 2014). Another economically relevant organism that was recently successfully targeted with TALENs is the silkworm, the primary producer of silk (Takasu *et al.* 2013).

6.3 Gene therapy and other biomedical applications

Recently, a library of TALENs that target protein-coding genes within the human genome has been assembled (Kim *et al.* 2013). Kim *et al.* screened for suitable target sites within these genes and assembled 18740 plasmids carrying corresponding TALENs using a modified high-throughput Golden Gate assembly method. The success rate of the assembled TALENs in a pilot study on HEK293 cells was 101 out of 103 target sites, with an average mutation frequency of 16%. There was a high variability in respect to performance, with the best performing TALEN pair displaying a mutation frequency of 54%. The fact that Kim *et al.* identified the two unsuccessfully targeted sites as having a high prevalence of methylated CpG islands suggests that factors such as methylation, DNA conformation and chromatin structure could play a role in influencing TALEN performance. The authors argue that in conjunction with a suitable screening assay for selecting successfully targeted cells, this TALEN library could be used to prepare knock-out cell lines that would provide more reliable insight into gene function than knock-down RNAi with its relatively high off-target rate and reproducibility issues.

This study may aid in another field in which readily available highly specific programmable endonucleases could be expected to make an impact: gene therapy. Both ZFNs (Lombardo *et al.* 2011), and recently TALENs and the CRISPR/Cas system have already been used as a genome editing tool on human induced and embryonic pluripotent stem cells with varying degrees of success (Li *et al.* 2014). For utilizing their po-

tential in biomedical and clinical applications, a much more stringent assessment of off-target mutations and cytotoxic effects of each respective nuclease is crucial. The relative ease of controlled assembly of both TALENs and the CRISPR/Cas system opens up another interesting path that may lead to clinical applications in the future — the potential of programmable nucleases as a tool to correct genetic disease (Schwank *et al.* 2013). Wu *et al.* demonstrate that it is possible to rescue the phenotype dominant cataract disorder in mice, by injecting Cas9 mRNA along with gRNA targeting the deficient locus within the *Cryc* gene responsible for the disorder into zygotes that were obtained by crossing wild-type females with homozygous males carrying the *Cryc* gene mutation (Wu *et al.* 2013).

7. Discussion

There is high variability in the nuclease-induced mutation frequencies reported in literature, but it is safe to say that TALENs surpass ZFNs in terms of efficiency, especially if the ease and speed with which they can be designed is taken into account. Their biggest advantage and potential lies in their accessibility, enabling their use in laboratories with access to basic molecular biology tools and with a limited budget to successfully employ TALENs in their research. In order to allow researchers to determine which system is the most suitable to be used in their respective setting though, it is crucial to uncover the mechanism and therefore the reasons behind off-target cleavage and the kind of variability in respect to activity reported by Kim *et al.* in their TALEN library. Kim *et al.* have successfully uncovered that CpG methylation can impair TALEN activity, but this accounts only for a fraction of the variability of activity displayed by the TALENs in their pilot study. Heterochromatin structure, DNA secondary structure, histone acetylation, method of delivery into the cell and a myriad other unexpected factors could play a significant role in influencing not only TALEN performance, but the performance of the CRISPR/Cas system as well. The same can be said about factors influencing their off-target activity and toxicity, elimination of which is a crucial step towards clinical applications of programmable nucleases. Since both ZFNs and TALENs rely on a peptide-nucleotide bond when recognizing their cleavage target and the CRISPR/Cas system relies RNA-DNA bonding, it is safe to assume that their performance and vulnerability to certain conditions will reflect this difference.

Large-scale comparative studies that could offer more insight into these issues and aid further development would greatly benefit from improving the use of programmable endonucleases in cell culture, a field that has somewhat lagged behind their use as a means to produce genetically modified organisms. Efficiency of cleavage varies greatly between cell types and to obtain cells that have been successfully edited requires labor-intensive cultivation of many clonal colonies and subsequent screening for mutations. There are several ways currently used to assess TALEN and CRISPR/Cas activity, notably the T7E1 assay which relies on PCR amplification from both wild-type DNA and DNA from the targeted cells, subsequent hybridization and heteroduplex cleavage (Niu *et al.* 2014). The drawbacks of using the T7E1 assay in cell culture screening is the need to isolate and amplify the DNA, which has to be preceded by generation of clonal colonies. Another strategy to screen for successful targeting is cotransfection of a reporter plasmid along with the ZFNs, TALENs or Cas9 and gRNA. The reporter plasmid that Kim *et al.* employ has a copy of the target sequences for both TALENs inserted into an mRFP-EGFP fusion expression cassette between mRFP and EGFP so that in its default state, EGFP is out of frame and thus only mRFP is expressed (Kim *et al.* 2011). If the nuclease is active within the cell, it cleaves its target on both chromosomal DNA and reporter plasmid, and reconstitutes EGFP expression in some cases where NHEJ leads to frameshift-inducing indels. Mashiko *et al.* use a similar reporter plasmid that relies on HR to restore EGFP expression as a result of nuclease activity (Mashiko *et al.* 2013). We are currently working on improving on these concepts by preparing a series of reporter plasmids that enable direct selection of successfully targeted cells from cell culture that can be used in conjunction with both TALENs and the CRISPR/Cas system.

8. Conclusion

Zinc-finger nucleases have been studied and used as a genome editing tool for several decades now, whereas the mechanism and function of TALE nucleases has been uncovered relatively recently. TALENs have greatly benefited from the volume of research in respect to performance, off-target activity, and mechanism that has already been done on ZFNs due to the fact that both systems use the same FokI nuclease domain. The functionally different CRISPR/Cas system however has been adapted as a genome editing tool only two years ago and is currently undergoing a surge of popularity (Shen 2013), even though there

is still relatively little insight into the mechanism underlying its function. This is being swiftly remedied as there is great interest to develop the method further. It is safe to say that both TALENs and the CRISPR/Cas system have not yet reached their full potential, despite the considerable impact their emergence has had in the few years since their discovery and deployment as genome editing tools.

9. Literature

Austin, Christopher P., James F. Battey, Allan Bradley, Maja Bucan, Mario Capecchi, Francis S. Collins, William F. Dove, Geoffrey Duyk, Susan Dymecki, and Janan T. Eppig. "The Knockout Mouse Project." *Nature Genetics* 36, no. 9 (2004): 921–24.

Auwerx, Johan, Phil Avner, Richard Baldock, Andrea Ballabio, Rudi Balling, Mariano Barbacid, Anton Berns, Allan Bradley, Steve Brown, and Peter Carmeliet. "The European Dimension for the Mouse Genome Mutagenesis Program." *Nature Genetics* 36, no. 9 (2004): 925–27.

Bai, Jianfa, Seong-Ho Choi, Grisel Ponciano, Hei Leung, and Jan E. Leach. "Xanthomonas Oryzae Pv. Oryzae Avirulence Genes Contribute Differently and Specifically to Pathogen Aggressiveness." *Molecular Plant-Microbe Interactions* 13, no. 12 (2000): 1322–29.

Bitinaite, Jurate, David A. Wah, Aneel K. Aggarwal, and Ira Schildkraut. "FokI Dimerization Is Required for DNA Cleavage." *Proceedings of the National Academy of Sciences* 95, no. 18 (1998): 10570–75.

Boch, J., H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas. "Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors." *Science* 326, no. 5959 (October 29, 2009): 1509–12. doi:10.1126/science.1178811.

Cermak, T., E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, and D. F. Voytas. "Efficient Design and Assembly of Custom TALEN and Other TAL Effector-Based Constructs for DNA Targeting." *Nucleic Acids Research* 39, no. 12 (July 1, 2011): e82–e82. doi:10.1093/nar/gkr218.

Chinwalla, Asif T., Lisa L. Cook, Kimberly D. Delehaunty, Ginger A. Fewell, Lucinda A. Fulton, Robert S. Fulton, Tina A. Graves, LaDeana W. Hillier, Elaine R. Mardis, and John D. McPherson. "Initial Sequencing and Comparative Analysis of the Mouse Genome." *Nature* 420, no. 6915 (2002): 520–62.

Christian, M., T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, and D. F. Voytas. "Targeting DNA Double-Strand Breaks with TAL Effector Nucleases." *Genetics* 186, no. 2 (July 26, 2010): 757–61. doi:10.1534/genetics.110.120717.

"CompoZr® Custom ZFN Service | Sigma-Aldrich." Accessed April 24, 2014. http://www.sigmaaldrich.com/life-science/zinc-finger-nuclease-technology/custom-zfn.html.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, et al. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339, no. 6121 (January 3, 2013): 819–23. doi:10.1126/science.1231143.

DeKelver, R. C., V. M. Choi, E. A. Moehle, D. E. Paschon, D. Hockemeyer, S. H. Meijsing, Y. Sancak, et al. "Functional Genomics, Proteomics, and Regulatory DNA Analysis in Isogenic Settings Using Zinc Finger Nuclease-Driven Transgenesis into a Safe Harbor Locus in the Human Genome." *Genome Research* 20, no. 8 (August 1, 2010): 1133–42. doi:10.1101/gr.106773.110.

Doyle, E. L., N. J. Booher, D. S. Standage, D. F. Voytas, V. P. Brendel, J. K. VanDyk, and A. J. Bogdanove. "TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: Tools for TAL Effector Design and Target Prediction." *Nucleic Acids Research* 40, no. W1 (July 1, 2012): W117–W122. doi:10.1093/nar/gks608.

Durai, S. "Zinc Finger Nucleases: Custom-Designed Molecular Scissors for Genome Engineering of Plant and Mammalian Cells." *Nucleic Acids Research* 33, no. 18 (October 12, 2005): 5978–90. doi:10.1093/nar/gki912.

Gaj, Thomas, Charles A. Gersbach, and Carlos F. Barbas III. "ZFN, TALEN, and CRISPR/Cas-Based Methods for Genome Engineering." *Trends in Biotechnology* 31, no. 7 (2013): 397–405.

Gaj, Thomas, Jing Guo, Yoshio Kato, Shannon J Sirk, and Carlos F Barbas. "Targeted Gene Knockout by Direct Delivery of Zinc-Finger Nuclease Proteins." *Nature Methods* 9, no. 8 (July 1, 2012): 805–7. doi:10.1038/nmeth.2030.

Geurts, A. M., G. J. Cost, Y. Freyvert, B. Zeitler, J. C. Miller, V. M. Choi, S. S. Jenkins, et al. "Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases." *Science* 325, no. 5939 (July 24, 2009): 433–433. doi:10.1126/science.1172447.

Glaser, Stefan, Konstantinos Anastassiadis, and A Francis Stewart. “Current Issues in Mouse Genome Engineering.” *Nature Genetics* 37, no. 11 (November 2005): 1187–93. doi:10.1038/ng1668.

Gonzalez, Beatriz, Lauren J Schwimmer, Roberta P Fuller, Yongjun Ye, Lily Asawapornmongkol, and Carlos F Barbas. “Modular System for the Construction of Zinc-Finger Libraries and Proteins.” *Nature Protocols* 5, no. 4 (April 1, 2010): 791–810. doi:10.1038/nprot.2010.34.

Gupta, A., R. G. Christensen, H. A. Bell, M. Goodwin, R. Y. Patel, M. Pandey, M. S. Enuameh, et al. “An Improved Predictive Recognition Model for Cys2-His2 Zinc Finger Proteins.” *Nucleic Acids Research*, February 12, 2014. doi:10.1093/nar/gku132.

Holkers, M., I. Maggio, J. Liu, J. M. Janssen, F. Miselli, C. Mussolino, A. Recchia, T. Cathomen, and M. A. F. V. Goncalves. “Differential Integrity of TALE Nuclease Genes Following Adenoviral and Lentiviral Vector Gene Transfer into Human Cells.” *Nucleic Acids Research* 41, no. 5 (March 1, 2013): e63–e63. doi:10.1093/nar/gks1446.

Ittner, Lars M, and Jürgen Götz. “Pronuclear Injection for the Production of Transgenic Mice.” *Nature Protocols* 2, no. 5 (May 2007): 1206–15. doi:10.1038/nprot.2007.145.

Jao, L.-E., S. R. Wente, and W. Chen. “Efficient Multiplex Biallelic Zebrafish Genome Editing Using a CRISPR Nuclease System.” *Proceedings of the National Academy of Sciences* 110, no. 34 (August 20, 2013): 13904–9. doi:10.1073/pnas.1308335110.

Jinek, M., F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, et al. “Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation.” *Science* 343, no. 6176 (March 14, 2014): 1247997–1247997. doi:10.1126/science.1247997.

Kim, Hyojin, Eunji Um, Sung-Rae Cho, Chorong Jung, Hyongbum Kim, and Jin-Soo Kim. “Surrogate Reporters for Enrichment of Cells with Nuclease-Induced Mutations.” *Nature Methods* 8, no. 11 (October 9, 2011): 941–43. doi:10.1038/nmeth.1733.

Kim, Yang-Gyun, Jooyeun Cha, and Srinivasan Chandrasegaran. “Hybrid Restriction Enzymes: Zinc Finger Fusions to Fok I Cleavage Domain.” *Proceedings of the National Academy of Sciences* 93, no. 3 (1996): 1156–60.

Kim, Yongsub, Jiyeon Kweon, Annie Kim, Jae Kyung Chon, Ji Yeon Yoo, Hye Joo Kim, Sojung Kim, et al. “A Library of TAL Effector Nucleases Spanning the Human Genome.” *Nature Biotechnology* 31, no. 3 (February 17, 2013): 251–58. doi:10.1038/nbt.2517.

Li, M., K. Suzuki, N. Y. Kim, G.-H. Liu, and J. C. Izpisua Belmonte. "A Cut above the Rest: Targeted Genome Editing Technologies in Human Pluripotent Stem Cells." *Journal of Biological Chemistry* 289, no. 8 (February 21, 2014): 4594–99. doi:10.1074/jbc.R113.488247.

Liu, Qiang, David J. Segal, Jayant B. Ghiara, and Carlos F. Barbas. "Design of Polydactyl Zinc-Finger Proteins for Unique Addressing within Complex Genomes." *Proceedings of the National Academy of Sciences* 94, no. 11 (1997): 5525–30.

Lombardo, Angelo, Daniela Cesana, Pietro Genovese, Bruno Di Stefano, Elena Provasi, Daniele F Colombo, Margherita Neri, et al. "Site-Specific Integration and Tailoring of Cassette Design for Sustainable Gene Transfer." *Nature Methods* 8, no. 10 (August 21, 2011): 861–69. doi:10.1038/nmeth.1674.

Maeder, Morgan L, Stacey Thibodeau-Beganny, Jeffry D Sander, Daniel F Voytas, and J Keith Joung. "Oligomerized Pool Engineering (OPEN): An „Open-Source" Protocol for Making Customized Zinc-Finger Arrays." *Nature Protocols* 4, no. 10 (September 17, 2009): 1471–1501. doi:10.1038/nprot.2009.98.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church. "RNA-Guided Human Genome Engineering via Cas9." *Science* 339, no. 6121 (January 3, 2013): 823–26. doi:10.1126/science.1232033.

Mashiko, Daisuke, Yoshitaka Fujihara, Yuhkoh Satouh, Haruhiko Miyata, Ayako Isotani, and Masahito Ikawa. "Generation of Mutant Mice by Pronuclear Injection of Circular Plasmid Expressing Cas9 and Single Guided RNA." *Scientific Reports* 3 (November 27, 2013). doi:10.1038/srep03355.

Mercer, A. C., T. Gaj, R. P. Fuller, and C. F. Barbas. "Chimeric TALE Recombinases with Programmable DNA Sequence Specificity." *Nucleic Acids Research* 40, no. 21 (November 1, 2012): 11163–72. doi:10.1093/nar/gks875.

Miller, J., A. D. McLachlan, and A. Klug. "Repetitive Zinc-Binding Domains in the Protein Transcription Factor IIIA from Xenopus Oocytes." *The EMBO Journal* 4, no. 6 (1985): 1609.

Moscou, M. J., and A. J. Bogdanove. "A Simple Cipher Governs DNA Recognition by TAL Effectors." *Science* 326, no. 5959 (October 29, 2009): 1501–1501. doi:10.1126/science.1178817.

Nagy, Andras, Janet Rossant, Reka Nagy, Wanda Abramow-Newerly, and John C. Roder. "Derivation of Completely Cell Culture-Derived Mice from Early-Passage Embryonic Stem Cells." *Proceedings of the National Academy of Sciences* 90, no. 18 (1993): 8424–28.

Niu, Yuyu, Bin Shen, Yiqiang Cui, Yongchang Chen, Jianying Wang, Lei Wang, Yu Kang, et al. "Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos." *Cell* 156, no. 4 (February 2014): 836–43. doi:10.1016/j.cell.2014.01.027.

Perez-Pinera, P., D. G. Ousterout, M. T. Brown, and C. A. Gersbach. "Gene Targeting to the ROSA26 Locus Directed by Engineered Zinc Finger Nucleases." *Nucleic Acids Research* 40, no. 8 (April 1, 2012): 3741–52. doi:10.1093/nar/gkr1214.

Reyon, Deepak, Shengdar Q Tsai, Cyd Khayter, Jennifer A Foden, Jeffry D Sander, and J Keith Joung. "FLASH Assembly of TALENs for High-Throughput Genome Editing." *Nature Biotechnology* 30, no. 5 (April 8, 2012): 460–65. doi:10.1038/nbt.2170.

Romer, P., S. Hahn, T. Jordan, T. Strauss, U. Bonas, and T. Lahaye. "Plant Pathogen Recognition Mediated by Promoter Activation of the Pepper Bs3 Resistance Gene." *Science* 318, no. 5850 (October 26, 2007): 645–48. doi:10.1126/science.1144958.

Schmid-Burgk, Jonathan L, Tobias Schmidt, Vera Kaiser, Klara Höning, and Veit Hornung. "A Ligation-Independent Cloning Technique for High-Throughput Assembly of Transcription Activator–like Effector Genes." *Nature Biotechnology* 31, no. 1 (December 16, 2012): 76–81. doi:10.1038/nbt.2460.

Schwank, Gerald, Bon-Kyoung Koo, Valentina Sasselli, Johanna F. Dekkers, Inha Heo, Turan Demircan, Nobuo Sasaki, et al. "Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients." *Cell Stem Cell* 13, no. 6 (December 2013): 653–58. doi:10.1016/j.stem.2013.11.002.

Shen, Helen. "CRISPR Technology Leaps from Lab to Industry." *Nature*, December 3, 2013. doi:10.1038/nature.2013.14299.

Sternberg, Samuel H., Sy Redding, Martin Jinek, Eric C. Greene, and Jennifer A. Doudna. "DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9." *Nature* 507, no. 7490 (January 29, 2014): 62–67. doi:10.1038/nature13011.

Sun, Ning, and Huimin Zhao. "Transcription Activator-like Effector Nucleases (TALENs): A Highly Efficient and Versatile Tool for Genome Editing." *Biotechnology and Bioengineering* 110, no. 7 (2013): 1811–21.

Takasu, Yoko, Suresh Sajwan, Takaaki Daimon, Mizuko Osanai-Futahashi, Keiro Uchino, Hideki Sezutsu, Toshiki Tamura, and Michal Zurovec. “Efficient TALEN Construction for Bombyx Mori Gene Targeting.” Edited by Mikael Rørdam Andersen. *PLoS ONE* 8, no. 9 (September 18, 2013): e73458. doi:10.1371/journal.pone.0073458.

Takata, Minoru, Masao S. Sasaki, Eiichiro Sonoda, Ciaran Morrison, Mitsumasa Hashimoto, Hiroshi Utsumi, Yuko Yamaguchi-Iwai, Akira Shinohara, and Shunichi Takeda. “Homologous Recombination and Non-Homologous End-Joining Pathways of DNA Double-Strand Break Repair Have Overlapping Roles in the Maintenance of Chromosomal Integrity in Vertebrate Cells.” *The EMBO Journal* 17, no. 18 (1998): 5497–5508.

“The Nobel Prize in Physiology or Medicine 1978.” Accessed April 25, 2014. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1978/.

Tijsterman, Marcel, and Ronald HA Plasterk. “Dicers at RISC: The Mechanism of RNAi.” *Cell* 117, no. 1 (2004): 1–3.

Tsai, Shengdar Q, Nicolas Wyvekens, Cyd Khayter, Jennifer A Foden, Vishal Thapar, Deepak Reyon, Mathew J Goodwin, Martin J Aryee, and J Keith Joung. “Dimeric CRISPR RNA-Guided FokI Nucleases for Highly Specific Genome Editing.” *Nature Biotechnology*, April 25, 2014. doi:10.1038/nbt.2908.

Urnov, Fyodor D., Edward J. Rebar, Michael C. Holmes, H. Steve Zhang, and Philip D. Gregory. “Genome Editing with Engineered Zinc Finger Nucleases.” *Nature Reviews Genetics* 11, no. 9 (September 2010): 636–46. doi:10.1038/nrg2842.

Wah, David A., Jurate Bitinaite, Ira Schildkraut, and Aneel K. Åggarwal. “Structure of FokI Has Implications for DNA Cleavage.” *Proceedings of the National Academy of Sciences* 95, no. 18 (1998): 10564–69.

Wiedenheft, Blake, Samuel H. Sternberg, and Jennifer A. Doudna. “RNA-Guided Genetic Silencing Systems in Bacteria and Archaea.” *Nature* 482, no. 7385 (February 15, 2012): 331–38. doi:10.1038/nature10886.

Wu, Yuxuan, Dan Liang, Yinghua Wang, Meizhu Bai, Wei Tang, Shiming Bao, Zhiqiang Yan, Dangsheng Li, and Jinsong Li. “Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9.” *Cell Stem Cell* 13, no. 6 (December 2013): 659–62. doi:10.1016/j.stem.2013.10.016.

Zalocusky, Kelly, and Karl Deisseroth. “Optogenetics in the Behaving Rat: Integration of Diverse New Technologies in a Vital Animal Model.” *Optogenetics* 1 (January 8, 2013): 1–17. doi:10.2478/optog-2013-0001.