
Improving CRISPR/Cas9 On-target Specificity

Muhammad Jamal^{1*}, Arif Ullah², Muhammad Ahsan Naeem¹, Rohit Tyagi¹, Zeshan Habib¹ and Khaista Rehman¹

¹State Key Laboratory of Agricultural Microbiology, College of Animal Science and Medicine, Huazhong Agricultural University, Wuhan, The People's Republic of China.

²Institute of Biochemistry and Biotechnology, University of Veterinary and Animal Sciences, Lahore, Pakistan.

*Corresponding author: jamalbiotech@webmail.hzau.edu.cn

<https://doi.org/10.21775/cimb.026.065>

Abstract

The CRISPR/Cas9 has revolutionized the field of molecular biology, medical genetics and medicine. The technology is robust, facile and simple to achieve genome targeting in cells and organisms. However, to propagate these nucleases for therapeutic application, the on-target specificity is of paramount importance. Although the binding and cleavage of off-target sites by Cas9 is an issue of concern, however, the specificity of CRISPR technology is greatly improved in current research employing the use of engineer nucleases, improved gRNA selection, novel Cas9 orthologues and the advancement in methods to detect and screen off-target sites and its effects. Here we summarize the advances in this state-of-the-art technology that will equip the genome editing tools to be applied in clinical research. The researcher should optimize these methods with emphasis to achieving perfection in the specificity.

Introduction

CRISPR (clustered, regularly interspaced, short palindromic repeat)/Cas (CRISPR-associated protein) is a defensive mechanism against the invading pathogen such as phages and plasmids (Bhaya *et al.*, 2011; Jinek *et al.*, 2012; Wiedenheft *et al.*, 2012). The system works by integrating a part of the invading genome into the specialized region in the

genome known as repeats; the incorporated part known as a spacer is inserted in between the two repeats. The expression of the repeat-spacer-repeat produces pre-CrRNA which upon maturation liberates mature crRNA. The degradation of foreign genome is achieved by means of CrRNA (CRISPR RNAs) complementary to sequences within re-invading viral genome or plasmid DNA (Bhaya *et al.*, 2011; Terns and Terns, 2011; Wiedenheft *et al.*, 2012). The crRNA guides Cas9 to the target site on the DNA to produce breaks on both DNA strands thereby removing the foreign genome (Jinek *et al.*, 2012). So far, CRISPR/Cas systems are recognized in a representative set of 703 archaeal and bacterial genomes, of which 310 (44%) encode one or more CRISPR/Cas modules after its initial discovery in *E. coli* in 1987 (Deveau *et al.*, 2010; Koonin and Makarova, 2009).

CRISPR technology has been established by the researchers to regulate or edit any locus of interest. Cas9 is localized by means by a short stretch of guide sequence known as guide RNA (gRNA). From the time of its first application more than 1700 papers have been published and more than 60,000 CRISPR/Cas related plasmids have been shipped by Addgene. The tool has been successfully applied in many organisms for genome editing including *C. elegans* (Friedland *et al.*, 2013; Tzur *et al.*, 2013), zebrafish (Chang *et al.*, 2013), mice (Platt *et al.*, 2014), rats (Li *et al.*, 2013), rabbits

(Yang *et al.*, 2014) and monkeys (Niu *et al.*, 2014), and also applied in cell lines and to treat mutations causing genetic disorders (Khan *et al.*, 2016; Niu *et al.*, 2014) such as β -thalassaemia (Xie *et al.*, 2014), cystic fibrosis (Schwank *et al.*, 2013), haemophilia A (Park *et al.*, 2015), cataracts (Wu *et al.*, 2013), hereditary tyrosinaemia type I (Yin *et al.*, 2014) and Duchenne muscular dystrophy (Long *et al.*, 2014). Moreover, the system is actively used for the removal of viruses such as hepatitis B (Ramanan *et al.*, 2015) and HIV (Ebina *et al.*, 2013) in human cell line. Beyond genome editing the technology is successfully applied for genome regulation (Fineran and Dy, 2014), imaging (Chen *et al.*, 2013), epigenetic regulation (Thakore *et al.*, 2015), fusion protein delivery (Mali *et al.*, 2013) (Fig. 6.1). However, to target any region in large genome, the binding of Cas9 to secondary genomic sites been remained a major complication in this field.

A number of novel strategies to measure and enhance genome wide specificity of Cas9 have been described in last three years which are

comprehensively summarized in recent reviews (Jamal *et al.*, 2016).

In this article we discuss Cas9 specificity profiled by various biased and unbiased methods, and their advantages and disadvantages. We focus on molecular mechanisms of Cas9 targeting, binding and cleavage that defines Cas9 specificity, Cas9 and guide sequence engineering for improving Cas9 specificity (Fig. 6.2). The optimization and consideration of these approaches will greatly benefit the genome editing field to be broadly applied in therapies.

CRISPR off-targets

RNA guided endonuclease (RGENs) are now regarded as important member of engineered nucleases. Certain applications, e.g. therapeutic genome editing in human stem cells, generation of homozygous cell lines for evaluating fundamental genetic variations, demands high level of precision and specificity (Soldner *et al.*, 2011). These

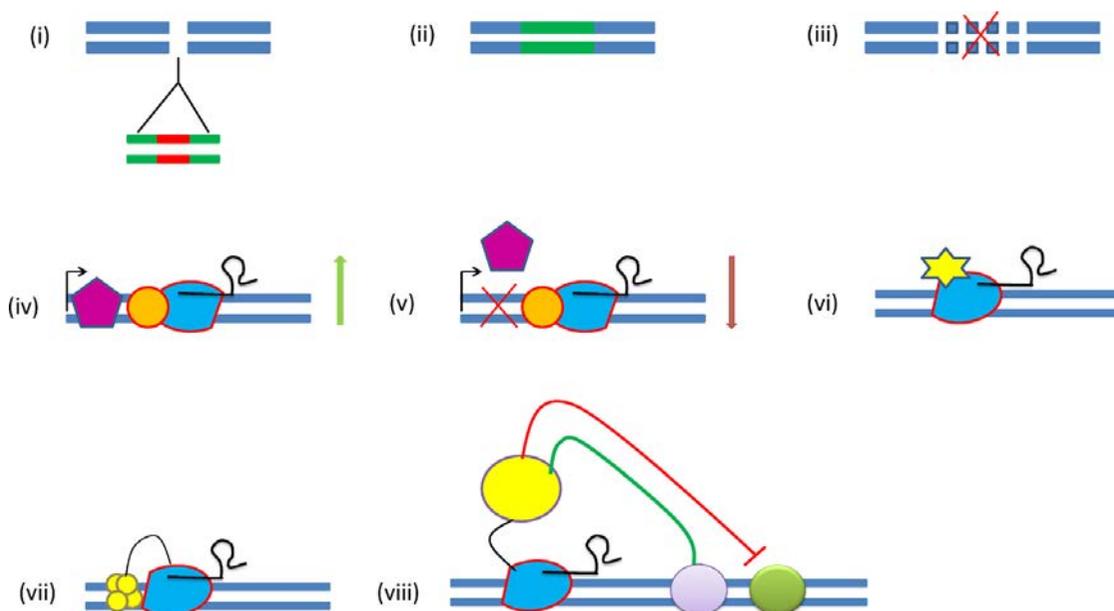


Figure 6.1 Application of CRISPR/Cas9/dCas9 based (A) gene disruption of wild type locus by using HDR template with desired mutation. (B) Gene knock-in to the target site by HR using HDR. (C) Gene knock-out caused by erroneous NHEJ pathways to rectify the DSB induced by Cas9. (D) dCas9 based up-regulation or transcriptional activation of target gene using dCas9 fused with domains like VP64 or omega. (E) Gene silencing or down-regulation of gene using dCas9 fused with repressor like KRAB. (F) Imaging of target genomic site using a colouring compound such as a fluorophore. (G) Antibody epitope tagging or any other fusion protein delivery. (H) Epigenetic modification (histone demethylation repress gene expression or histone acetylation activate gene expression).

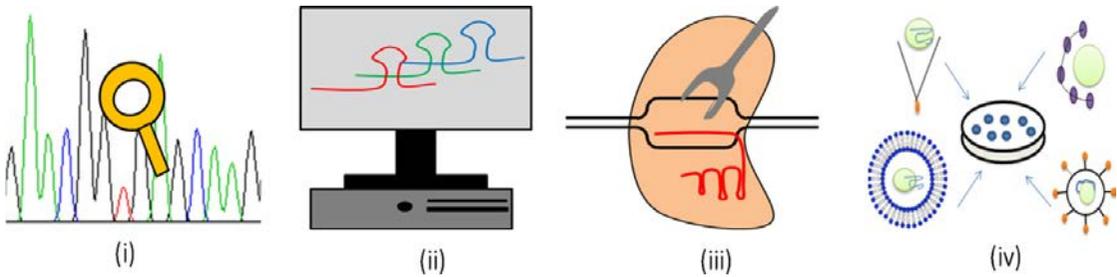


Figure 6.2 Schematic representation of the strategies used to increase CRISPR/Cas9 specificities, (A) analysis of off-target sites using biased and unbiased methods. (B) Web based designed of gRNA with minimum off-targets in the genome. (C) Engineering of Cas9 and gRNA. (D) Delivery methods of Cas9–sgRNA ribonucleoprotein complex to allow transient expression of Cas9-sgRNA ribonucleoprotein complex.

nucleases with off-target DSB activity could induce undesirable mutations with potentially deleterious effects, an unacceptable outcome in most clinical settings. Similarly, due to the permanent nature of genomic modifications, specificity is of paramount concern to sensitive applications such as for studies aimed at linking specific genetic variants with biological processes, disease phenotypes, clinical applications and gene therapy. An ideal nuclease engineered nuclease should have single site cleavage specificity, but Cas9 has been shown to possess off-target cleavage activities.

Analysing off-target effects

Prediction of off-target sites is difficult, as the off-target cleavage is not fully understood. Various methods including restriction fragment length polymorphism (RFLP), PCR amplification of off-target sites followed by next generation sequencing (NGS), or mismatch sensitive nuclease (Surveyor assay and T7 endonuclease assay). The first method used for detection of off-targets was RFLP in cells (Urnov *et al.*, 2005). Surveyor assay and T7 endonuclease assays have been used to detect indel mutations generated as a result of erroneous pathway NHEJ in a population of cells.

Sanger sequencing is used to reveal mutations at off- and on-target sites induced by nucleases. Similarly, high-throughput sequencing predicts indel mutation (frequency ranging from 0.01% to 1%) up to hundreds at off- and on-target sites. However, a great deal of precautions should be taken to remove the false positive clones obtained during PCR artefacts (Cho *et al.*, 2014). These endonucleases recognize and cleave non-perfectly

matched DNA (heteroduplexes) that are obtained as a result of annealing of wild-type and mutant DNA sequences. These digested DNA are then run on gel. The size and intensity of the bands can provide a clue about the mutation frequency. The T7E1 endonuclease assay is more sensitive than CEL-I to indels with deletion substrates (Vouillot *et al.*, 2015), while Surveyor is suited for detecting single nucleotide changes.

Genome wide assays for interrogation of off-target sites

The methods used to detect off-target sites are mostly based on binding and cleavage of these nucleases either in cell or out of cells (*in vitro*) are described in Fig. 6.3.

In-cell

The sequencing techniques such as sequencing the entire genome of the edited cells, known as whole genome sequencing, can be used to profile Cas9 specificity (Iyer *et al.*, 2015; Smith *et al.*, 2014). The sequencing result of the genome pre- and post-editing can reveal about the generation of novel mutation near the target site. This platform is useful when applied to single cell, clones non-mosaic F1 genome edited animals. However, the method is limited owing to low sensitivity; especially off-target with low frequencies are not detected, but also off-target in large population of cells (Smith *et al.*, 2014; Veres *et al.*, 2014).

The method is unbiased to gauge specificity of engineered nucleases for single cell-clones. However, the poor sensitivity to off-target sites especially those which occur at low frequency in a

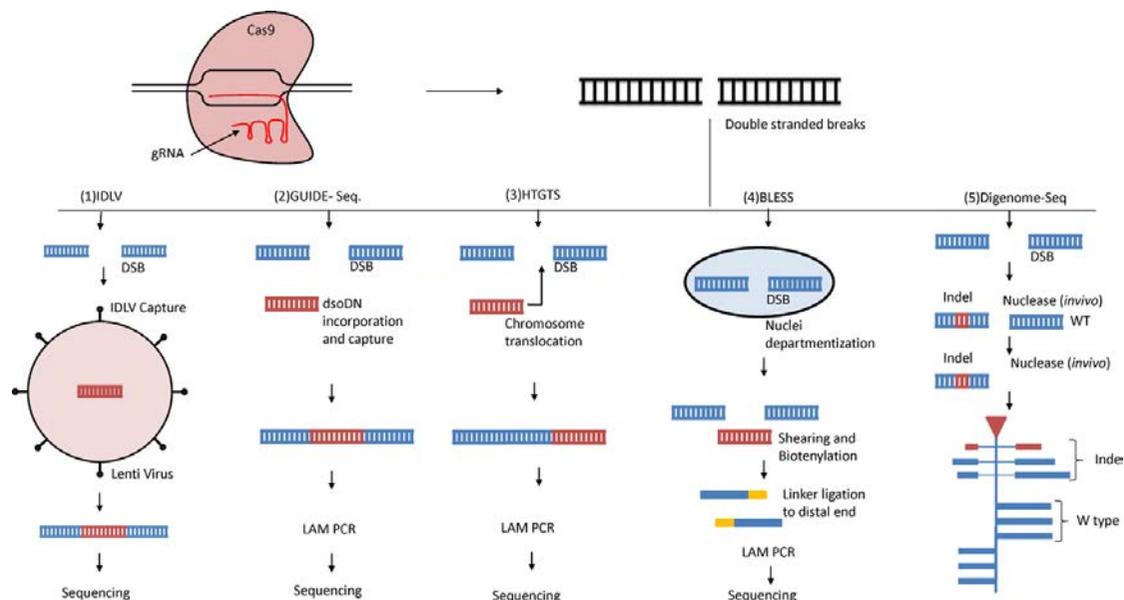


Figure 6.3 Schematic overview of off-target detection methods, the mechanism of each method is described in the text.

cell population (Tsai and Joung, 2014). DSB capturing method is based on incorporation of IDLV into the DSBs via non-homologous end joining (NHEJ) and tags those transient DSBs. The tagged sites are recovered by linear amplification-mediated PCR (LAM-PCR) and then mapped using high-throughput sequencing. The advantage of the IDLV capture method is that it can recognize the DSB induced by nuclease. However, the technique has some limitations such as less integration efficiencies requiring positive selection to resolve, and also random integration into sites free of nuclease induced DSBs (Gabriel *et al.*, 2011).

BLESS

BLESS (breaks labelling, enrichment on streptavidin and next-generation sequencing) is based on *in vivo* DNA cleavage, but *in vitro* DSB capturing detection method that involves fixation of cells, chromatin purification, nuclease digestion, ligation with biotinylated linkers and targeted deep sequencing. The process is beneficial as no exogenous bait is introduced to the cells also it can detect DSBs from tissues in *in vivo* (Ran *et al.*, 2015). The process is limited as the DSB captured by BLESS is at one and specific time at a specific time, during fixation but cannot detect the DSBs formed pre-fixation. The

complex and time consuming protocol and also the requires a large number of cells (more than 10 million cells) are the disadvantages associated with this technique.

GUIDE-seq

GUIDE-seq (genome-wide unbiased identification of double stranded breaks enabled by sequencing) is a method used to detect the off-target effects in cells (Tsai *et al.*, 2015). The method involves the incorporation of a blunt, end-protected double-stranded oligodeoxynucleotide (dsODN) tag, specific amplification of the tag followed by high-throughput sequencing. The sequencing reads are compared with reference genome to locate the DSBs. The method is simple, efficient and sensitive, and can detect off-target sites that are mutagenized with frequencies of 0.1% or lower in a population of cells. The method precisely detects off-target sites that were not detected by other tools (Heigwer *et al.*, 2014) and also the computational pipelines available online for bioinformatics analysis (Tsai *et al.*, 2016). However, the disadvantage of the method is the requirement of cost-effective delivery method of dsODN to avoid its toxicity in some cell lines as observed.

HTGTS

HTGTS (high-throughput genome-wide translocation sequencing) (Frock *et al.*, 2015; Hu *et al.*, 2015) is an unbiased method for off-target site detection using known site of the genome as bait to capture prey DNA sequences that are translocated to on-target sites in live cells. The genome wide libraries containing bait–prey interactions are generated and are cloned using LAM-PCR, adapter ligation and PCR amplification of library followed by deep sequencing.

The drawback of this method is that translocation events occur very rarely depending on the three dimensional organization of genome, since translocations occur frequently on the same chromosome or a chromosome that lies in close proximity.

***In vitro* genome-wide assay for the detection of off-targets sites**

The genomic DNA free from cells can be used to analyse the off-target effects *in vitro*.

Digenome-seq

The method used known as digested genome sequencing (Digenome-seq) is based on cleaving genomic DNA *in vitro* with Cas9 resulting in the same 5' end at cleavage sites, adapter ligation followed by WGS. The cleaved sites are identified by read alignment (Kim *et al.*, 2015). The method offers advantage over GUIDE-seq as it can be applied to any sort of cell type without considering factors such as epigenetics, chromatin or subnuclear localization because digestion is performed on genomic DNA *in vitro*, and is much more sensitive than GUIDE-seq because increasing the concentration of Cas9–gRNA complex will enable the detection of sites cleaved with low frequency. The problem associated with Digenome-seq is the sequencing cost, moreover the method still needs to determine the difference between spontaneous mutation and that of the mutation induced by the off-target effect of the nuclease, so to this end the sequencing reads must be compared with other method to validate the cleaved sites.

SELEX

SELEX (systematic evolution of ligands by exponential enrichment) is an *in vitro* method based on

the investigation of Cas9 binding to the target site rather than target cleavage. A set of oligonucleotides matching target sites in the genome are used and the bound sequences are then searched to a library of oligonucleotides to anticipate potential off-target sites under controlled conditions (Gabriel *et al.*, 2011).

ChIP-seq

ChIP-seq (chromatin immunoprecipitation) uses the binding of dCas9 to target site to assess off-target sites (Frock *et al.*, 2015; Wu *et al.*, 2014). Although ChIP is a useful unbiased method, unfortunately one should not correlate this method for Cas9 and dCas9 as DNA binding and cleavage are independent events, with cleavage more stringent than binding.

In summary, each and every method has disadvantages so using these methods in complementary manner will help understand potential off-target sites and its effects that will in turn help to develop novel approaches with optimized specificity. The biased method depends on bioinformatics algorithms to identify off-target sites, and on the other hand the unbiased method captures DSB generated by nucleases. The biased method is complemented by the unbiased method as it fails to predict about large number of off-target sites. For example, using *in vitro* test, off-target sites can be discovered (e.g. Digenome-seq), in cells (e.g. HTGTS, GUIDE-seq), or *in vivo* (e.g. BLESS) prior to validation. The combinations of these methods HTGTS, GUIDE-seq, and Digenome-seq revealed a set of potential off-target sites used for one sgRNA targeting VEGF-A site. In this regard the use of larger datasets can accurately define off-targeting *in vivo* and in cell that in turn will help in the development of bioinformatics tools used to design and select target sites.

gRNA designing tools

Experimental identification of off-target sites by *in vitro* or *in vivo* methods is laborious. Therefore, *in silico* prediction tools are broadly applied. The first generation *in silico* tools were mostly designed according to the degree of similarity between the target site and secondary sites in the genome. The recently developed tools predict unique genomic

sites with minimum sequence homology to the secondary genomic sites (Frock *et al.*, 2015; Heigwer *et al.*, 2014).

While designing an sgRNA the search for potential off-targets in the genome is important.

The web based tools developed by research groups are described in Table 6.1. The basic feature of each tool is to look for PAM and seed sequence in a genomic DNA. Tools such as CasOT (available online: <http://eendb.zfgenetics.org/casot/>) and sgRNACas9 (available online: www.biotoools.com) also predict the number and position of mismatches.

Cas-OFFinder [available online: www.rgenome.net/cas-offinder/ (accessed on 14 October 2015)] and COSMID [available online: <https://crispr.bme.gatech.edu> (accessed on 14 October 2015)] incorporate input sequence and search for 20 nt target sequences near to the canonical or non-canonical PAM (NGG or NRG), giving output off-target sites.

The web tools search off-target sites based on mismatches not on the basis of indels. However, it has been showed that gRNA with few indels may

induce cleavage (CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences). Therefore, to look for off-target sites in a given sequence it is also necessary to check presence of indels as well as nucleotide mismatches, tools like Cas-OFFinder, CRISPRdirect and COSMID provide these features. The online tools can be applied for limited organisms but the off-line tools selects target sites from many organisms.

How to improve the specificity of CRISPR/Cas9 system?

Off-target mutations can be caused both by Cas9 and gRNA. Therefore, improvement of the current generation of CRISPR/Cas platform is mostly focused on these two factors. Several efforts aiming to improve CRISPR specificity should be performed prior to using CRISPR/Cas9-mediated genome editing, including the rational design of gRNA, selection of eligible CRISPR nuclease, choice of suitable target sites and delivery of Cas9-sgRNA into cells followed by rapid degradation

Table 6.1 Web based tools to design gRNA

Name of tool	No. of organism genomes supported	Length of input sequence	Website
ZiFiT	9	Target sequence (more than 1 kb)	http://zifit.partners.org/ZiFiT/Disclaimer.aspx
Cas-OFFinder	More than 30	Designed guide sequence (15 to 25 nt)	http://www.rgenome.net/cas-offinder/
E-CRISP	More than 30	Target sequence (more than 10 kb)	http://www.e-crisp.org/E-CRISP/designcrispr.html
CHOPCHOP	More than 25	More than 10 kb	https://chopchop.rc.fas.harvard.edu
CasOT	Any	More than 10 kb	http://eendb.zfgenetics.org/casot/
Blenching	22		https://benchling.com
CRISPR-P	More than 25 plants	More than 5 kb	http://cbi.hzau.edu.cn/crispr/
COSMID	More than 8	Out sgRNA designed is 10 to 55 nt	https://crispr.bme.gatech.edu/
DNA 2.0 CRISPR gRNA Design Tool	Any	Input target sequence more than 10 kb	https://www.dna20.com/eCommerce/cas9/input
CRISPR-direct	More than 20	Input target sequence more than 10 kb	http://crispr.dbcls.jp/
MIT CRISPR design tool	More than 10	Input target sequence more than 10 kb	http://crispr.mit.edu/
CRISPR-ERA	More than 10	Input target sequence more than 5 kb	http://crispr-era.stanford.edu/

of the nuclease. Additionally, understanding the molecular mechanism of binding and cleavage of Cas9–gRNA complex can improve the specificity of nuclease.

Specificity achieved by using modified gRNA

The truncated gRNAs (tru-gRNAs), with shorter sequences (<20 nucleotides, usually 17 or 18) can effectively improve the specificity of Cas9 (Fu *et al.*, 2014b). The off-target mutagenesis was largely reduced about 5000-fold by using tru-gRNA (Fu *et al.*, 2014b). GUIDE-seq showed that numbers of off-target cleavage sites across the genome were decreased by 2- to 5-fold for the same sites digested with Cas9 guided by full length gRNA. Intriguingly, truncation at the 5' end of gRNA increase the on-target specificity, while truncation at the 3' end of gRNA reduces the on-target cleavage activity of Cas9 nucleases (Fu *et al.*, 2013; Hwang *et al.*, 2013). Moreover, truncation of the standard gRNA by one to three nucleotides showed comparable activity to that of the full length gRNA however deleting more than three nucleotides (16 nucleotides gRNA) resulted in the loss of activity (Fu *et al.*, 2014b).

Recently tru-RFNs (combination of tru-gRNA pair with RNA guided FOKI nucleases) showed high activity in human cancer cell lines and embryonic stem cells with reduced undesirable off-target effects (Wyvekens *et al.*, 2015). The tru-gRNA provides Cas9 with the necessary energy for binding to target site and reduces the affinity to off-target site. Usually, the full length gRNA provides more energy than the required, allowing the Cas9 to bind to unspecific sequences. In addition, the efficiency of tru-gRNA can be further enhanced when truncated gRNAs are conjugated with paired Cas9 nickases (Fu *et al.*, 2014b).

Another strategy is using the gRNA containing additional guanine bases at the 5' end (5'-GGX20 or 5'-GGGX19) makes it more specific to the target by an unknown exact mechanism. the possible mechanism might be changing the gRNA stability and secondary structure or stabilizing protein interaction with 5' end of the gRNA (Cho *et al.*, 2014; Kim *et al.*, 2015) (shown in Fig. 6.1C). In some cases it has been reported that these extensive gRNA exhibit lower on-target activity compared to gRNA of normal length (Cho *et al.*, 2014; Kim *et al.*, 2015). In another strategy the introduction of

an A–U base pair to the gRNA scaffold stabilizes the gRNA scaffold. Besides, extension of the base pair in the hairpin increases the binding rate of gRNA to target site and finally enhances its on-target specificity (Chen *et al.*, 2013).

The synthesis of gRNA with chemical modification at 5' and 3' termini increased Cas9 specificity allowing successful genome editing in human T cells and CD34+ haematopoietic stem and in progenitor cells the modification consists of adding 2'-O-methyl (M), 2'-O-methyl 3'phosphorothioate (MS), or 2'-O-methyl 3' thioPACE (MSP) to the last three nucleotides at both 5' and 3' ends (Hendel *et al.*, 2015). However using these modified gRNA revealed off-target activity for four predicted sites using deep sequencing.

In another approach Rahdar *et al.* (2015) synthesized a 29-nucleotide synthetic crRNA (ScrRNA) with methyl modifying the 5' end base pairs and bases and interspersed 2'-fluoro modifications towards the 3' end (Rahdar *et al.*, 2015). These modified guides although showed enhanced stability and specificity but showed a reduced the on-target effects.

Lesson learnt from biochemical, structural and functional studies

The dilemma of understanding the mechanism of how Cas9 recognize and cleave the target DNA is of pivotal importance to achieve enhanced specificity in genome editing. Recent good models describe a comprehensive mechanism and insights into the conformational changes when Cas9 recognize and binds to its target in complex with gRNA (Jiang *et al.*, 2015) and associated with gRNA and the target strand without (Nishimasu *et al.*, 2014) or with the PAM sequence (Anders *et al.*, 2014). Cas9 comprises of two lobes; a nuclease lobe containing two nuclease domains, namely HNH and RuvC that performs the duty of double-stranded DNA cleavage, and a recognition lobe that binds sgRNA to the target DNA where HNH and RuvC cleaves the complementary strand of DNA and the non-target DNA respectively (Nishimasu *et al.*, 2014). The nuclease lobe also interacts with protospacer adjacent motif (PAM) by mean of presence of a carboxyl terminal. A DNA sequence of 2–5 bp (mostly 5'-NGG-3', where N can be any nucleotide) is crucial for CRISPR-based cutting of DNA located

adjacent to the target sequence (protospacer). In the context of recognition the Cas9 interacts with repeat anti-repeat region of sgRNA when recognizing PAM proximal position on target DNA, and then recognizing the complementary sequence on the target DNA by means of 20 bp gRNA sequence. The protospacer interacting (PI) domain identifies the PAM sequence leading to the formation of R-loop. The formation of R-loop crucial for making gRNA-DNA-Cas9 (ternary complex) activates the NHN and RuvC domains to cleave the phosphate group on non-complementary and on complementary strands respectively (Jinek *et al.*, 2012).

PAM interactions are important for high-affinity Cas9 binding to target DNA (Jinek *et al.*, 2012). Cas9 in complex with gRNA is pre-organized to make PAM-interacting contacts (Jiang *et al.*, 2015). The interaction of Cas9 with PAM sequence on the target non-target strand changes the conformation resulting in the unwinding of DNA (Anders *et al.*, 2014). Moreover, future engineering of the PI domain may allow us to program the PAM specificity, improve the target site recognition fidelity, and increase the versatility of the Cas9 genome engineering platform.

Specificity achieved by Cas9 engineering

Cas9 mutant and its variant

The mutated Cas9 (Cas9n) with one point mutation, either H840A in the HNH domain or D10A in the RuvC domain, was found to be more specific and efficient for genome targeting purposes (Ran *et al.*, 2013) (shown in Fig. 6.1F). Cas9n has the ability to create offset nicks instead of a DSB at the target site guided by pair of gRNAs resulting in the formation of indel mutations (Cong *et al.*, 2013; Ran *et al.*, 2013). Additionally, a pair of double nicks at two sites by four customized gRNAs–Cas complex successfully deleted genomic fragments of up to 6kb in HEK 293FT cells (Ran *et al.*, 2013). Cas9n reduces the off-targeting by 50- to 1000-fold without sacrificing the on-target effect (Ran *et al.*, 2013). Thus combination of two gRNAs is preferred when the specificity is of paramount importance (e.g. the creation of engineered cell lines). The delivery of two gRNAs reduces the likelihood that their off-target sites are close enough to

cause double stranded breaks. This approach has led the foundation to edit stem cells for therapeutic purposes (Ran *et al.*, 2013).

It has been shown that Cas9n mutagenizes certain target sites at high frequency, presumably because the nicks created at these sites can be converted into double-strand breaks. While targeting by Cas9n the off-target mutation induced by the second gRNA in a pair is not precisely understood, to this end the off-target effects of both gRNA should be considered. The method discussed to detect the nuclease profile of Cas9n targeting RAG1 (recombination activating gene 1), HTGTS revealed the detection of few translocation junctions detected genome wide. The problem associated with the use of Cas9n is its high frequency of inducing point mutations at certain target sites (Cho *et al.*, 2014; Fu *et al.*, 2014b; Tsai *et al.*, 2014) and also low efficiency of targeting with some gRNAs.

Specificity achieved by alanine substitution in Cas9

Recently researcher of Zhang group developed eSpCas9 and eSaCas9 (enhanced specificity SpCas9 and SaCas9) respectively based on structural engineering of SpCas9 and SaCas9 (Slaymaker *et al.*, 2016). These nucleases are different from their parent nuclease by three and four codon substitutions respectively with increased specificity detected by BLESS. These mutations are presumably weakening the interaction and binding of Cas9 proteins to the non-target DNA.

An analogous approach used by researchers developed another variant of Cas9 described as HF (High Fidelity)-SpCas9 based on previous structural profile of Cas9 (Kleinstiver *et al.*, 2015). This type of nuclease contains four alanine substitutions in the residues that interact with phosphate backbone of the target DNA. The HF-SpCas9 possesses high efficiency at on-target compared with Cas9 when observed for gRNA targeting sites where the wild type Cas9 showed off-target mutation. The disadvantage of these variants is the induction of mutation at some off-target sites.

Another strategy involved the engineering of Cas9 DNA-binding ability by fusing it with DNA-binding domain to produce a chimera complex known as Cas9-pDBD. This complex reduced improved the precision and specificity nearly 150-fold, this is because binding of the domain

increases the concentration of SpCas around the target site (Bolukbasi *et al.*, 2015). The wide range of targeting is probably provided by skipping the formation of R-loop followed by PAM recognition (Szczelkun *et al.*, 2014).

Another approach is modifying the PAM recognition sequence in Cas9. PAM plays a key role in the recognition of target DNA sequence. This will help to narrow down target site selection and also confer specificity to this type of variant. This may be achieved by altering the protospacer interacting domain (PID) to recognize a different PAM sequence as described by (Fonfara *et al.*, 2014; Nishimasu *et al.*, 2014) where the Cas9 sequences from *Staphylococcus pyogenes* and from *Streptococcus thermophilus* CRISPR were 60% identical. These variants can be guided by dual gRNAs without affecting the specificity and efficiency (Fonfara *et al.*, 2014).

dCas9 specificity

The dead Cas9 generated as a result of double mutations, fused with effectors domains, is used for transcriptional modulation, chromosomal labelling and epigenetics regulation (Koneremann *et al.*, 2013). The dCas9 lacks cleavage activity but retains the binding ability.

dCas9–FOKI

The fusion of FOKI domain to the carboxy terminal of zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) achieved mark specificity (Bogdanove and Voytas, 2011). Here wild type nuclease FokI domain fused to the amino terminal of catalytically inactive Cas9 (dCas9) protein known as RFNs. The dCas9 is itself unable to cut DNA, but the fused FokI endonuclease cuts the DNA as a dimer again assisted by two sgRNAs with closely juxtaposed target site. The process is proved to be more specific as the RFNs complex does not induce any DNA lesions. RFNs offer a better recognition fidelity and specificity of ≥ 140 -fold greater than Cas9 and 4-fold lower to that of Cas9n (Guilinger *et al.*, 2014b). RFNs are governed by the half-sites bound by each gRNA/FokI–dCas9 complex in a particular relative orientation (PAMs-out) in which PAM site lies outside of the intervening spacer (14 to 17bp). The requirement of two gRNAs for precise orientation of the target sites and for correct spacing between

half-sites is more inflexible for the dCas9–FokI fusion construct than that for Cas9 nickase (Tsai *et al.*, 2014). Such a tough requirement selection gRNAs will likely reduce the number of available on-target sites. Moreover, the dimerization of paired dCas9 binding to two precise genomic targetable half-sites should render dCas9–FokI fusion nucleases substantially improved on-target specificity.

Coupling tru-gRNA with dCas9–FOKI (tru-RFNS) reduced off-target mutagenesis by 40% compared to RFNs in human cells (Wyvekens *et al.*, 2015). These tru-RFNS reduce the possibility of point mutations, tru-gRNAs provide necessary energy for binding to target site compare to full length gRNA which provide extra energy, thus allowing the Cas9 to bind to unspecific sequences.

The non-target DNA (non-complementary to gRNA) released first after Cas9 cleavage, the use of single stranded donor DNA complementary to the non-target strand showed an increased rate of HDR compared to the donor complementary to the target strand (Richardson *et al.*, 2016).

RFNs strictly depend upon appropriately targeted gRNAs. In a broader sense, the development of dimeric FokI–dCas9 fusions may prove a significant turning point towards practical mammalian genome manipulations with high efficiency and accuracy. However, a downside of both systems is that both double-nicking and dimeric FokI–dCas9 is reported to have lower efficiency than wild type Cas9, and also the large size of these genes (more than 8kb), which is hard to carry for the vectors (capacity of AAV is 4.2 kb). Moreover, the requirement for PAM sites within a specified region limits the possible target sites. To this point the improvement in RFN by engineering the FOKI domain like heterodomain FOKI and gRNA sequence can aid to reduce or eliminate the off-target effects observed in single gRNA guiding.

Looking for Cas9 orthologues and CRISPR proteins across the bacterial kingdom

Recently discovered Cas9 orthologue from *Staphylococcus aureus* (SaCas9) four times smaller than SpCas9 allow easily packaging into AAV for delivery (Friedland *et al.*, 2015; Ran *et al.*, 2015). The requirement of a guide sequence and PAM is different for SpCas9 and SaCas9. The former needs

a guide sequence of 20 nt and 5'-NGG PAM, the later needs a guide sequence of 21 to 24 nt and 5' NNGRRT as PAM sequence. SaCas9 guided by 5' (NGG)RRT has been showed to induce lower indel mutations at off-target sites (Ran *et al.*, 2015).

The Cas9 discovered from *Neisseria meningitidis* (NmCas9) with a more complex PAM sequence (5'-NNNNGATT) and a 21 to 24 bp guide sequence has been used in mammalian cell genome editing (Mali *et al.*, 2013a; Müller *et al.*, 2016). The long PAM sequence limits the number of off-target sites, as the frequency of short PAM 5'-NGG occurs after every 8 bp whereas lengthy PAM 5'-NNNNGATT occur after 128 bp. However, the longer PAM sequence limits the target site range (Lee *et al.*, 2016).

The presence of CRISPR system in diverse prokaryotic kingdom allows scientists to discover CRISPR system across different bacterial genome. Recently discovered nuclease known as Cpf1 (CRISPR from *Prevotella* and *Francisella* 1) belongs to class 2 type V-A CRISPR system has been efficiently used for human genome editing (Zetsche *et al.*, 2015). The PAM sequence for Cpf1 is 5'TTN present upstream of the target site. The target DNA is cleaved at different position, producing a staggered cut unlike Cas9, which produces blunt-ended DNA (Fonfara *et al.*, 2014; Yamano *et al.*, 2016). The genome editing specificity of Cpf1 was found to be comparable to that of Cas9 profiled Digenome-seq and GUIDE-seq (Kim *et al.*, 2016).

C2c2 is a member of class2, type VI present in *Leptotrichia shahii* provides immunity against RNA viruses. Researchers from MIT have used it for programmable repression of target RNA guided by crRNA (Abudayyeh *et al.*, 2016). The C2c2 binds to target sites and acts as RNase inhibiting RNA. However, the system is not completely known and the factors that govern the specificity (PAM, protein structure, and guide sequence) are not completely elaborated.

Concentration of Cas9-gRNA

The rate of off-target mutagenesis also increases with increase in the concentration of TALEN as described by Guilinger *et al.* (2014a). This reflects that the lower concentration of Cas9 reduces the

extra energy for DNA binding to avoid the binding to off-target sites *in vitro* and in cells (Fu *et al.*, 2014a; Hsu *et al.*, 2013; Pattanayak *et al.*, 2013).

There are several methods for the delivery of Cas9-gRNA into the cells; the most common method is the transfection of plasmid containing gRNA and Cas9 cassettes. The method for delivery of the CRISPR system is of great concern as it allows controlled optimum concentration of Cas9 and sgRNA in a cell that in turn is important for increasing the specificity (Hsu *et al.*, 2013; Pattanayak *et al.*, 2013). Recently, a positive correlation was observed between Cas9/sgRNA expression level and mutagenesis efficiency in plants (Mikami *et al.*, 2015), as indicated by increased specificity when lowering the amount of plasmid DNA in transfection. While in excessive enzyme concentration, mismatches are tolerated (Pattanayak *et al.*, 2013). However, it is noteworthy that reducing the Cas9/gRNA complex also decreases the on-target site cleavage efficiency (Pattanayak *et al.*, 2013; Ran *et al.*, 2013). The delivery of purified Cas9 in complex with gRNA via electroporation or lipofection achieved efficient genome editing with low off-target effects (Cho *et al.*, 2013; Lin *et al.*, 2014; Zuris *et al.*, 2015).

The use of a weak or inducible promoter is thought to regulate the Cas9 expression to optimal level. The constitutive expression of these endonuclease increases the on-targeting editing but might also lead to higher off-target effects, as observed for zinc finger nucleases (ZFNs) (Gaj *et al.*, 2012). The expression timing of Cas9 is an additional factor affecting off-target, because short lived and timed delivery of Cas9 allowed minimum off-target effects (Lin *et al.*, 2014).

The regulation of Cas9 concentration at transcriptional and post-transcriptional level is important to allow successful genome editing; using low or inducible promoter can achieve this control expression. To achieve post-transcriptional control of Cas9 in cells Davis *et al.* (2015) developed Cas9 regulated by a small molecule known as intein that undergoes protein splicing when 4-hydroxyamoxefin is available. This conditionally active chimera achieved successful genome editing in cell line with 25-fold for 11 off-target sites already assessed (Davis *et al.*, 2015).

It has been shown that cell cycle synchronization of the nuclease in G2 increases HDR efficiency

while reducing unwanted NHEJ events when supplied with RNP (Lin *et al.*, 2014). Researchers have used inhibitors of the NHEJ repair pathway. Maruyama *et al.* (2015) found that genome editing efficiency was increased up to 19-fold by preventing NHEJ when the DNA ligase IV (a principal enzyme involved in NHEJ repair pathway) was antagonized by its inhibitor Scr7.

What else should be considered in avoiding off-target mutagenesis?

Cell fate

The DSB induced Cas9 and its repair is highly cell specific depending on the cell DNA repair machinery (Duan *et al.*, 2014), for instance, the vulnerability of off-target mutation is high in transformed human cell lines where the repair system is not intact compared to healthy human cell line with a regulated DSB repair pathway (Smith *et al.*, 2014; Veres *et al.*, 2014). The epigenetic state of the cell also impacts the off-target effects; DNA methylation at CpG island may interfere with the binding of Cas9. The methods to detect off-target effects have been applied to cancer cell lines such as HEK293, U2OS, and K562 where the nuclease-induced off-targets were detectable compare to stem cells (ES, iPCS) where the off-target were not detectable. The higher number of off-targets in cancer cell lines might reflect the abnormal DNA repair mechanism, as observed by GUIDE-seq (enables genome-wide profiling of off-target cleavage by CRISPR/Cas nucleases) where DSB were observed in the absence of CRISPR/Cas9.

Protospacer adjacent motif (PAM)

The protospacer-adjacent motif (PAM), three nucleotides (NGG) canonical PAM sequence, identified by one of the Cas9 domains is strictly required to be immediately next to the 3' end of the target sequence. The PAM binding is required for opening and cleavage of the target site (Nishimasu *et al.*, 2014). The number of possible potential gRNAs for a gene is proportional to the length of the gene. The *S. pyogenes* PAM site is repeated every 8 nucleotide on average. Using the frequency of 'GG' equal to 5.21% in the reference human genome

there would be an expected 161,284,793 NGG PAM sites in the human genome, or roughly one 'GG' dinucleotide every 42 bases. The most important triplet nucleotide sequence (NGG) that itself is not included in the sgRNA sequence, is identified by one domain of Cas9 enzyme (Nishimasu *et al.*, 2014). The PAM sequence NRG (where R can be G or A) is identified by Cas9-mediated DNA cleavage at the human EMX locus (Hsu *et al.*, 2013; Jiang *et al.*, 2013) although with lower binding efficiency (one-fifth) than canonical PAM sequence.

Therefore, the designing of gRNA should be based on the fact that there should be minimum or no homology between off-target and on-target sites and the off-target should not be followed by the canonical or non-canonical PAM.

PAM distal versus PAM proximal nucleotides

The seed sequence (5–12 nt) near the PAM sequence is critical for Cas9-sgRNA binding; mismatches in the PAM-distal nucleotides are less important and generally have less effect on disturbing the sgRNA–target DNA hybrid than those in the PAM-proximal nucleotides (PAMPNs). Therefore, it suggests that PAM-proximal nucleotides are major determinants of Cas9 binding with target sites (Cong *et al.*, 2013; Fu *et al.*, 2013; Jinek *et al.*, 2012; Kucsu *et al.*, 2014; Mali *et al.*, 2013c). Moreover, the seed sequence with multiple U also increases the specificity as more likely resulting in decrease sgRNA abundance also terminating the transcription of sgRNA (Wang *et al.*, 2014; Wu *et al.*, 2014).

GC content of sgRNA

It has been shown that mutagenesis efficiency is associated with GC content of PAM proximal nucleotides (PAMPNs) in gRNA. gRNA with three or fewer GCs in PAMPNs rarely reach 60% heritable mutation rate compare to more than 60% heritable mutation of gRNA with four GCs in PAMPNs (Ren *et al.*, 2014). This is because high GC content stabilizes the sgRNA–DNA hybrid allowing maximum Cas9 efficiency. The sgRNA containing G at the first position of the seed sequence is preferable; on the other hand the presence of C but not guanine at fifth position proximal to PAM and A in the middle of sgRNA is ideal.

Future directions

The specificity is of paramount importance in therapeutic applications. The problems of poor specificity are overcome by the latest approaches discussed in this article. The progress to determine the methods to detect the off-target mutations is remarkable, but still there is room for improvement, especially improving the sensitivity for the detection of off-target sites, as these biased and unbiased methods have detection limit of 0.1% yet these methods are insensitive to certain off-target sites. To this end, the development of more accurate and highly sensitive methods that can detect the off-target site below 0.1% is desperately needed. Combination of these methods using the same gRNA and same target site would reinforce the detection of off-target sites.

The engineering of Cas9 (discovery of eSpCas9 and SpCas9-HF) and gRNA has further strengthened the specificity. These new variants increased reduced tolerance to mismatched DNA sites. Moreover the discoveries new nucleases Cpf1, C2c2, Cas10 and NgAgo equipped the genome editing toolbox to be used broadly in genome engineering. However the off-target effects of these nucleases are not described yet. Other approaches like use of safe and efficient delivery methods to allow transient expression of Cas9 and sgRNA will reduce the safety risk during clinical and therapeutic application built on genome-editing nucleases. Comprehensive consideration and propagation of these approaches will greatly facilitate genome editing technology to achieve a marked specificity and efficiency.

Glossary

CRISPR: A system present in bacteria providing immunity against invading agents such as viruses and plasmids.

Cas9: CRISPR-associated protein component of CRISPR system responsible for breaking double-stranded DNA when guided by gRNA.

CrRNA: CRISPR RNA is the transcribed product of spacer-repeat-spacer. The spacer is complementary to the target DNA sequence.

tracrRNA: Stands for trans-activating CrRNA, is the product of DNA sequences present near the CRISPR locus. The activation of CRISPR system

requires the hybridization of crRNA into tracrRNA to form guide RNA (gRNA).

Protospacer: The target DNA sequence where Cas9 produces cleavage.

PAM: PAM stands for protospacer adjacent motif and this is the sequence recognized by Cas9 for licensing the DNA binding and cleavage.

NHEJ: Non-homologous end joining is a repair pathway used by the cell to rectify the double-stranded break induced by nucleases in the absence of a homology-directed repair template. The resultant correction may bring point mutations into the genome, often leading to gene knock-out.

HDR: Homology directed repair is another pathway used when a donor DNA template is provided containing the gene of interest to knock-in into the target after the DSB induced by nuclease. This type of pathway results in the formation of knock-in/-out.

Cas9 nickases: A mutant version of Cas9 with a mutation in one of its DNA-cleaving domains, the resultant Cas9 nick single strand of double-stranded DNA.

Dead Cas9 (dCas9): Engineering Cas9 by mutation in both nuclease domains; as a result it cannot cut the DNA but retains the DN-binding ability.

High-throughput sequencing: A DNA sequencing platform with high accuracy compared to Sanger sequencing method, which can be applied to sequence with short reads.

IDLV: The incorporation of integrase-deficient lentivirus (IDLV) into DSB in live cells can capture the DSB sites in live cell. These DSB are tagged and PCR amplified and analysed using high-throughput sequencing.

ChIP-seq: Chromatin immunoprecipitation, a method profiling the genome-wide protein interaction with DNA, the DNA associated with protein is sheared into pieces, unlinked, purified and sequenced.

SELEX: Systematic evolution of ligands by exponential enrichment is a method that analyses the binding of Cas9 to the target site. A random pool of oligo matching target sites are used, the unbound set of oligos is removed and the bound oligos are PCR amplified and sequenced.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 31371106,

31421064), Hubei Provincial Natural Science Foundation of China (Grant No. 36113011), Doctoral Fund of Ministry of Education of China (Grant No. 131012) and the Huazhong Agricultural University Scientific and Technological Self innovation Foundation (Program No. 52204-13002).

Conflict of interest

The authors disclose no conflict of interest.

References

- Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., and Minakhin, L. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* aaf5573.
- Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573. <http://dx.doi.org/10.1038/nature13579>
- Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297. <http://dx.doi.org/10.1146/annurev-genet-110410-132430>
- Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. *Science* 333, 1843–1846. <http://dx.doi.org/10.1126/science.1204094>
- Bolukbasi, M.F., Gupta, A., Oikemus, S., Derr, A.G., Garber, M., Brodsky, M.H., Zhu, L.J., and Wolfe, S.A. (2015). DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* 12, 1150–1156. <http://dx.doi.org/10.1038/nmeth.3624>
- Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.-W., and Xi, J.J. (2013). Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.* 23, 465–472. <http://dx.doi.org/10.1038/cr.2013.45>.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., and Qi, L.S. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. <http://dx.doi.org/10.1016/j.cell.2013.12.001>.
- Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24, 132–141. <http://dx.doi.org/10.1101/gr.162339.113>
- Cho, S.W., Lee, J., Carroll, D., Kim, J.-S., and Lee, J. (2013). Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* 195, 1177–1180. <http://dx.doi.org/10.1534/genetics.113.155853>
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., and Marraffini, L.A. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. <http://dx.doi.org/10.1126/science.1231143>
- Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A., and Liu, D.R. (2015). Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* 11, 316–318. <http://dx.doi.org/10.1038/nchembio.1793>
- Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* 64, 475–493. <http://dx.doi.org/10.1146/annurev.micro.112408.134123>
- Duan, J., Lu, G., Xie, Z., Lou, M., Luo, J., Guo, L., and Zhang, Y. (2014). Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res.* 24, 1009.
- Ebina, H., Misawa, N., Kanemura, Y., and Koyanagi, Y. (2013). Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Scientific Rep.* 3.
- Fineran, P.C., and Dy, R.L. (2014). Gene regulation by engineered CRISPR-Cas systems. *Curr. Opin. Microbiol.* 18, 83–89. <http://dx.doi.org/10.1016/j.mib.2014.02.007>
- Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.-L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* 42, 2577–2590. <http://dx.doi.org/10.1093/nar/gkt1074>
- Friedland, A.E., Baral, R., Singhal, P., Loveluck, K., Shen, S., Sanchez, M., Marco, E., Gotta, G.M., Maeder, M.L., and Kennedy, E.M. (2015). Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol.* 16, 1.
- Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J.A. (2013). Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* 10, 741–743. <http://dx.doi.org/10.1038/nmeth.2532>
- Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 33, 179–186. <http://dx.doi.org/10.1038/nbt.3101>
- Fu, B.X.H., Hansen, L.L., Artiles, K.L., Nonet, M.L., and Fire, A.Z. (2014a). Landscape of target-guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res.* gku1102.
- Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31, 822–826. <http://dx.doi.org/10.1038/nbt.2623>
- Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014b). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 32, 279–284. <http://dx.doi.org/10.1038/nbt.2808>
- Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaepffel, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., and Friedman, G. (2011). An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 29, 816–823. <http://dx.doi.org/10.1038/nbt.1948>
- Gaj, T., Guo, J., Kato, Y., Sirk, S.J., and Barbas, C.F. (2012). Targeted gene knockout by direct delivery of zinc-finger

- nuclease proteins. *Nat. Methods* 9, 805–807. <http://dx.doi.org/10.1038/nmeth.2030>
- Guilinger, J.P., Pattanayak, V., Reyon, D., Tsai, S.Q., Sander, J.D., Joung, J.K., and Liu, D.R. (2014a). Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat. Methods* 11, 429–435. <http://dx.doi.org/10.1038/nmeth.2845>
- Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014b). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 32, 577.
- Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. *Nat. Methods* 11, 122–123. <http://dx.doi.org/10.1038/nmeth.2812>
- Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., and Wilkens, A.B., *et al.* (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat. Biotechnol.* 33, 985–989. <http://dx.doi.org/10.1038/nbt.3290>
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., and Shalem, O., *et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. <http://dx.doi.org/10.1038/nbt.2647>
- Hu, J., Zhang, Y., Zhao, L., Frock, R.L., Du, Z., Meyers, R.M., Meng, F.-l., Schatz, D.G., and Alt, F.W. (2015). Chromosomal loop domains direct the recombination of antigen receptor genes. *Cell* 163, 947–959. <http://dx.doi.org/10.1016/j.cell.2015.10.016>
- Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Kaini, P., Sander, J.D., Joung, J.K., Peterson, R.T., and Yeh, J.-R.J. (2013). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One* 8, e68708. <http://dx.doi.org/10.1371/journal.pone.0068708>
- Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., and Skarnes, W.C. (2015). Off-target mutations are rare in Cas9-modified mice. *Nat. Methods* 12, 479. <http://dx.doi.org/10.1038/nmeth.3408>
- Jamal, M., Khan, F.A., Da, L., Habib, Z., Dai, J., and Cao, G. (2016). Keeping CRISPR/Cas on-target. *Curr. Issues. Mol. Biol.* 20, 1-20.
- Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* 348, 1477–1481.
- Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233–239. <http://dx.doi.org/10.1038/nbt.2508>
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. <http://dx.doi.org/10.1126/science.1225829>
- Khan, F.A., Pandupuspitasari, N.S., Chun-Jie, H., Ao, Z., Jamal, M., Zohaib, A., Khan, F.A., Hakim, M., and Shujun, Z. (2016). CRISPR/Cas9 therapeutics: a cure for cancer and other genetic diseases. *Oncotarget.* 7, 52541–52552.
- Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.-I., and Kim, J.-S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12, 237–243.
- Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.-H., and Kim, J.-S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* 34, 863–868.
- Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., and Yeh, J.-R.J. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485. <http://dx.doi.org/10.1038/nature14592>
- Konermann, S., Brigham, M.D., Trevino, A., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500, 472–476.
- Koonin, E.V., and Makarova, K.S. (2009). CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000. Biol. Rep.* 1.
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 32, 677–683.
- Lee, C.M., Cradick, T.J., and Bao, G. (2016). The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol. Ther.* 24, 645–654. <http://dx.doi.org/10.1038/mt.2016.8>
- Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., and Lu, X. (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 681–683. <http://dx.doi.org/10.1038/nbt.2661>
- Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 3, e04766. <http://dx.doi.org/10.7554/eLife.04766>
- Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA. *Science* 345, 1184–1188. <http://dx.doi.org/10.1126/science.1254445>
- Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31, 833–838.
- Mali, P., Esvelt, K.M., and Church, G.M. (2013b). Cas9 as a versatile tool for engineering biology. *Nat. Methods* 10, 957–963. <http://dx.doi.org/10.1038/nmeth.2649>
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013c). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. <http://dx.doi.org/10.1126/science.1232033>
- Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* 33, 538–542.
- Mikami, M., Toki, S., and Endo, M. (2015). Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. *Plant. Cell. Rep.* 1–9.
- Müller, M., Lee, C.M., Gasiunas, G., Davis, T.H., Cradick, T.J., Siksnys, V., Bao, G., Cathomen, T., and Musolino, C. (2016). *Streptococcus thermophilus* CRISPR-Cas9

- systems enable specific editing of the human genome. *Mol. Ther.* 24, 636–644.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949. <http://dx.doi.org/10.1016/j.cell.2014.02.001>
- Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Kang, Y., Zhao, X., Si, W., and Li, W. (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156, 836–843. <http://dx.doi.org/10.1016/j.cell.2014.01.027>
- Park, C.-Y., Kim, D.H., Son, J.S., Sung, J.J., Lee, J., Bae, S., Kim, J.-H., Kim, D.-W., and Kim, J.-S. (2015). Functional correction of large factor VIII gene chromosomal inversions in hemophilia A patient-derived iPSCs using CRISPR-Cas9. *Cell Stem Cell* 17, 213–220. <http://dx.doi.org/10.1016/j.stem.2015.07.001>
- Pattanayak, V., Lin, S., Guiling, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31, 839–843. <http://dx.doi.org/10.1038/nbt.2673>
- Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., and Jovanovic, M. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159, 440–455. <http://dx.doi.org/10.1016/j.cell.2014.09.014>
- Rahdar, M., McMahon, M.A., Prakash, T.P., Swayze, E.E., Bennett, C.F., and Cleveland, D.W. (2015). Synthetic CRISPR RNA-Cas9–guided genome editing in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, E7110–E7117. <http://dx.doi.org/10.1073/pnas.1520883112>
- Ramanan, V., Shlomai, A., Cox, D.B., Schwartz, R.E., Michailidis, E., Bhatta, A., Scott, D.A., Zhang, F., Rice, C.M., and Bhatia, S.N. (2015). CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. *Sci. Rep.* 5, 10833. <http://dx.doi.org/10.1038/srep10833>
- Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., and Makarova, K.S. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520, 186–191. <http://dx.doi.org/10.1038/nature14299>
- Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., and Zhang, Y. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389. <http://dx.doi.org/10.1016/j.cell.2013.08.021>
- Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S.-J., Qiao, H.-H., Wang, X., and Hu, Q. (2014). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. *Cell. Rep.* 9, 1151–1162. <http://dx.doi.org/10.1016/j.celrep.2014.09.044>
- Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* 34, 339–344.
- Schwank, G., Koo, B.-K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., and van der Ent, C.K. (2013). Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* 13, 653–658. <http://dx.doi.org/10.1016/j.stem.2013.11.002>
- Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88. <http://dx.doi.org/10.1126/science.1232727>
- Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., and Cheng, L. (2014). Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* 15, 12.
- Soldner, F., Laganière, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., and Lindquist, S. (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell* 146, 318–331. <http://dx.doi.org/10.1016/j.cell.2011.06.019>
- Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9798–9803. <http://dx.doi.org/10.1073/pnas.1402597111>
- Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14, 321–327. <http://dx.doi.org/10.1016/j.mib.2011.03.005>
- Thakore, P.I., D’Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., and Gersbach, C.A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* 12, 1143–1149. <http://dx.doi.org/10.1038/nmeth.3630>
- Tsai, S.Q., and Joung, J.K. (2014). What’s changed with genome editing? *Cell Stem Cell* 15, 3–4. <http://dx.doi.org/10.1016/j.stem.2014.06.017>
- Tsai, S.Q., Topkar, V.V., Joung, J.K., and Aryee, M.J. (2016). Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.* 34, 483–483.
- Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 32, 569–576. <http://dx.doi.org/10.1038/nbt.2908>
- Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., and Le, L.P. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197. <http://dx.doi.org/10.1038/nbt.3117>
- Tzur, Y.B., Friedland, A.E., Nadarajan, S., Church, G.M., Calarco, J.A., and Colaiácovo, M.P. (2013). Heritable custom genomic modifications in *Caenorhabditis elegans* via a CRISPR–Cas9 system. *Genetics* 195, 1181–1185.
- Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435, 646–651.

- Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell. Stem. Cell.* 15, 27–30. <http://dx.doi.org/10.1016/j.stem.2014.04.020>
- Vouillot, L., Th  lie, A., and Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3: Genes Genomes Genet.* 5, 407–415.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80–84. <http://dx.doi.org/10.1126/science.1246981>
- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338. <http://dx.doi.org/10.1038/nature10886>
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., and Chen, S. (2014). Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32, 670–676. <http://dx.doi.org/10.1038/nbt.2889>
- Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., Yan, Z., Li, D., and Li, J. (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell. Stem. Cell.* 13, 659–662. <http://dx.doi.org/10.1016/j.stem.2013.10.016>
- Wyvekens, N., Topkar, V.V., Khayter, C., Joung, J.K., and Tsai, S.Q. (2015). Dimeric CRISPR RNA-guided FokI-dCas9 nucleases directed by truncated gRNAs for highly specific genome editing. *Hum. Gene Ther.* 26, 425–431. <http://dx.doi.org/10.1089/hum.2015.084>
- Xie, F., Ye, L., Chang, J.C., Beyer, A.I., Wang, J., Muench, M.O., and Kan, Y.W. (2014). Seamless gene correction of β -thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. *Genome Res.* 24, 1526–1533. <http://dx.doi.org/10.1101/gr.173427.114>
- Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., and Koonin, E.V. (2016). Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 165, 949–962. <http://dx.doi.org/10.1016/j.cell.2016.04.003>
- Yang, D., Xu, J., Zhu, T., Fan, J., Lai, L., Zhang, J., and Chen, Y.E. (2014). Effective gene targeting in rabbits using RNA-guided Cas9 nucleases. *J. Mol. Cell. Biol.* 6, 97–99.
- Yin, H., Xue, W., Chen, S., Bogorad, R.L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P.A., Jacks, T., and Anderson, D.G. (2014). Corrigendum: Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* 32, 952–952.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., and Regev, A. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771. <http://dx.doi.org/10.1016/j.cell.2015.09.038>
- Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.-Y., and Liu, D.R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* 33, 73–80. <http://dx.doi.org/10.1038/nbt.3081>