ANNUAL REVIEWS

Annual Review of Biochemistry

Evaluating and Enhancing Target Specificity of Gene-Editing Nucleases and Deaminases

Daesik Kim,[1,*] Kevin Luk,[2,*] Scot A. Wolfe,[2] and Jin-Soo Kim[1,3]

[1]Center for Genome Engineering, Institute for Basic Science, Daejeon 34126, Republic of Korea; email: jskim01@snu.ac.kr

[2]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA; email: scot.wolfe@umassmed.edu

[3]Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea

Annu. Rev. Biochem. 2019. 88:191–220

First published as a Review in Advance on March 18, 2019

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-013118-111730

*These authors contributed equally to this article

Keywords

CRISPR/Cas9, off-target, gene editing, base editors, Cas12a, Cpf1

Abstract

Programmable nucleases and deaminases, which include zinc-finger nucleases, transcription activator-like effector nucleases, CRISPR RNA-guided nucleases, and RNA-guided base editors, are now widely employed for the targeted modification of genomes in cells and organisms. These gene-editing tools hold tremendous promise for therapeutic applications. Importantly, these nucleases and deaminases may display off-target activity through the recognition of near-cognate DNA sequences to their target sites, resulting in collateral damage to the genome in the form of local mutagenesis or genomic rearrangements. For therapeutic genome-editing applications with these classes of programmable enzymes, it is essential to measure and limit genome-wide off-target activity. Herein, we discuss the key determinants of off-target activity for these systems. We describe various cell-based and cell-free methods for identifying genome-wide off-target sites and diverse strategies that have been developed for reducing the off-target activity of programmable gene-editing enzymes.

Contents

INTRODUCTION

Zinc-finger nuclease (ZFN): a programmable nuclease comprising a sequence-specific zinc-finger protein and a nuclease domain derived from the *FokI* restriction endonuclease

Transcription activator-like effector nuclease (TALEN): a programmable nuclease comprising a sequence-specific TALE protein and a nuclease domain derived from the *FokI* restriction endonuclease

The genome engineering field has transitioned from a discovery phase to wide-spread application over the past decade. In the early days of zinc-finger nucleases (ZFNs), the primary focus of the field was the discovery and development of reagents that would efficiently cleave a desired target site (1). With the advent of improved ZFNs (2) and development of transcription activator-like effector nucleases (TALENs) (3) and Class II CRISPR RNA programmable nucleases [Cas9 (4–7) and more recently Cpf1 (also known as Cas12a) (8)], researchers now have a palette of different reagents to choose from that are likely to have high activity at a target site of interest. Consequently, for therapeutic applications, the choice of nuclease-based reagent is driven by other factors, such as its ease of development, genome-wide specificity, and delivery method. A small number of therapeutic genome engineering tools have entered clinical trials for specific applications and many others are poised to enter clinical trials soon (9, 10).

In addition to the programmable nuclease systems, novel base-editing systems have recently been developed that hold tremendous promise for the correction of disease-causing point mutations within the genome. These base editors (BEs) have a Cas9 nickase (nCas9) fused to nucleobase deaminases that create targeted transitions (C→T or A→G) through strand-biased DNA repair. Cytidine BEs combine cytosine deaminases [e.g., APOBEC (11–13) or AID (14)] with nCas9 to convert C/G base pairs to T/A, whereas adenosine BEs combine a selected/engineered tRNA adenosine deaminase (TadA) (15) with nCas9 to convert A/T base pairs to G/C. These base-editing systems have also been applied *in vivo* (16, 17) to efficiently introduce directed base conversion at desired target sites within the genome in adult animals.

As these new genome-editing tools are progressing toward the clinic in the form of cell-based and *in vivo* therapies, one of the outstanding concerns is the specificity of these nucleases and BEs—in other words, what modifications, outside the intended target site, are they making within the genome? Undesired mutagenesis is a prominent concern for the therapeutic application of programmable nucleases and BEs. In addition to the production of unwanted local mutagenesis, off-target DNA cleavage can potentially give rise to chromosomal rearrangements between two double-strand breaks (DSBs) to produce aberrant gene transcripts and genomic instability (18–23). Consequently, it is important to measure the activity of nucleases and BEs in an unbiased genome-wide manner. With this knowledge, target site choice and nuclease functionality can be improved to achieve the desired goal: restricting editing to a single target site within the human genome.

Herein, we provide an overview of the comprehensive and unbiased methods that have been developed to identify and characterize genome-wide off-target sites for a programmable nuclease or BE and the various approaches that have been used to improve the specificity of these reagents. Owing to growing interest in Class II CRISPR effector systems and variants thereof, this review primarily focuses on these systems; however, many of the methodologies for assessing genome-wide off-target activity should also be applicable to other nuclease formats (e.g., ZFNs, TALENs, etc.). These new off-target detection methodologies in combination with the development of highly specific nucleases and BEs will provide a strong foundation for the development of a multitude of new therapeutic genome-editing modalities that should realize cures for many devastating human disorders.

CRISPR-associated protein 9 (Cas9): a single-effector RNA-guided endonuclease of the Class II, Type II CRISPR/Cas system

Cpf1 (also known as Cas12a): a single-effector RNA-guided endonuclease of the Class II, Type V CRISPR/Cas system

Base editors (BEs): chimeric proteins comprising a DNA-targeting module (Cas9 or Cpf1) and cytosine or adenine deaminase to create single-base substitutions

Off-target effect: DNA cleavage at an unintended genomic sequence by a programmable nuclease that produces an unwanted mutation or sequence variation

DETERMINANTS OF OFF-TARGET EFFECTS

Nuclease off-target effects are determined primarily by a combination of target site uniqueness, chromatin states, and the properties and concentration of the nuclease. First, the presence or absence of highly homologous sequences within the genome to the target site is one of the key determinants of off-target effects. Targeting a unique site with *Streptococcus pyogenes* Cas9 (hereafter referred to as SpCas9) that differs from any other site in the entire human genome by at least three mismatches can avoid many off-target effects (24, 25). Homologous sites that differ by only one or two mismatches from the target site sequence are likely to be mutated at some frequency in cells by SpCas9. The presence of suitable unique target sites within a genomic locus of interest can be found using a variety of web-based programs (26–36). To knock out a gene of interest, one can choose unique target sites anywhere within an open reading frame, although targeting constitutively incorporated upstream exons rather than terminal exons is typically more effective at disrupting a gene of interest (37). To improve knockout rates at the expense of target site uniqueness, conserved domains critical for gene function can be targeted by CRISPR systems (38).

Second, chromatin states can impact Cas9 or Cpf1 activity to varying degrees. Evidence has been found that DNA accessibility can modestly influence target site mutagenesis activity of SpCas9 (39) but that chromatin does not drastically influence on-target editing rates in proliferating cells. SpCas9 off-target activity is influenced more by DNA accessibility than on-target activity (40). However, other Cas9 orthologs and Cpf1 nucleases are influenced by target site accessibility, as nearby binding of a nuclease-dead SpCas9 can enhance nuclease activity for these factors (41). Similarly, an off-target site in an open-chromatin region is more likely to be mutated by Cas9 than a site with the same DNA sequence in a closed-chromatin region (28, 40). Because chromatin state is tissue and cell-type dependent, the off-target activity of a nuclease is also expected to be context dependent.

Insertion or deletion (InDel): common class of mutations referring to the insertion, deletion, or insertion and deletion of nucleotides in genomic DNA

Third, off-target effects are highly variable among different classes of nuclease proteins. Early generation ZFNs constructed via modular assembly are cytotoxic (42, 43), most likely because they can cleave multiple off-target sites in the genome, triggering p53 activation due to the large number of DSBs. Highly optimized ZFNs, TALENs, and CRISPR systems are more specific and are not generally toxic in cells, unless programmed to target a repetitive sequence within the genome (44, 45).

Fourth, on-target activity and off-target activity of a given nuclease are two sides of the same coin. A highly efficient nuclease [favorable k_{cat} relative to k_{off} (46)] is likely to suffer from more severe off-target activity than an inefficient nuclease (**Figure 1**). For a given nuclease, off-target editing is energetically more costly or slower than on-target editing, because off-target interaction is less stable (k_{off} is higher). In addition, allosteric regulation of nuclease activity (k_{cat}) reduces this rate in the presence of mismatches between the guide and DNA sequence (47–50). An optimal nuclease (**Figure 1**) is one retaining almost full on-target activity but with low or negligible off-target activity, which can be found or created by increasing the energetic cost required for off-target editing or slowing down the cleavage reaction (k_{cat}) up to a certain point, while minimally changing the energetic cost or the reaction rate for on-target editing.

Fifth, off-target effects are dependent on the intracellular concentration of the nuclease and the duration of nuclease activity. Decreasing the concentration of a given nuclease in cells, in general, reduces off-target activity more than on-target activity, leading to improvements in specificity (35). Likewise, limiting the duration of nuclease activity can decrease off-target activity (51–53).

QUANTIFICATION OF ON- AND OFF-TARGET MUTAGENESIS

Management guru Peter Drucker once said, "If you can't measure it, you can't improve it" (54, p. 135). To control or reduce off-target effects, it is critical to be able to measure and quantify them. Although high-throughput sequencing remains the gold standard for measuring mutagenesis frequencies at on- and off- target sites (55, 56), a number of alternative methods, such as mismatch-sensitive enzymes and tracking InDels (insertion or deletion of bases) by decomposition (TIDE), can be utilized. T7 endonuclease I (T7E1) and the less frequently employed CEL I (also known as the Surveyor nuclease) (57), which are mismatch-sensitive nucleases, are widely used to measure mutagenesis frequencies of programmable nucleases (43, 55). T7E1 cleaves mismatches or bulges in heteroduplexes, formed by hybridization of wild-type and mutant DNA sequences. Digested heteroduplexes are then analyzed by agarose gel electrophoresis, in which the fraction of mutations within the population can be calculated based on the intensity of the cleaved DNA bands (57). Although its low cost and technical simplicity make the T7E1 assay an attractive method for detecting nuclease-mediated gene editing, it is unable to accurately determine the fraction of modified alleles in some mixtures of edited genomes (55). In comparison with high-throughput sequencing, T7E1 has poor sensitivity (typically, mutation rates >1% are required), has a more limited dynamic range (typically, accuracy decreases when the fraction of mutant alleles within the population is greater than 30–40%), and it can underestimate the fraction of mutant sequences within a population when the InDel distribution has low complexity (55).

TIDE is an in silico program that deconvolutes Sanger sequencing traces produced by complex mixtures of genomic sequences (56, 58). This method estimates InDel frequencies within a population of edited alleles by decomposing a complex Sanger sequence chromatogram of the population of edited alleles into the estimated chromatograms for individual InDel alleles. Compared with T7E1, TIDE provides a more accurate estimate of the fraction of mutant genomes within a population as well as their identity (55). Inherently, TIDE is limited by quality of the sequence chromatograms that are being analyzed, which limits detection to InDel rates >1–2%.

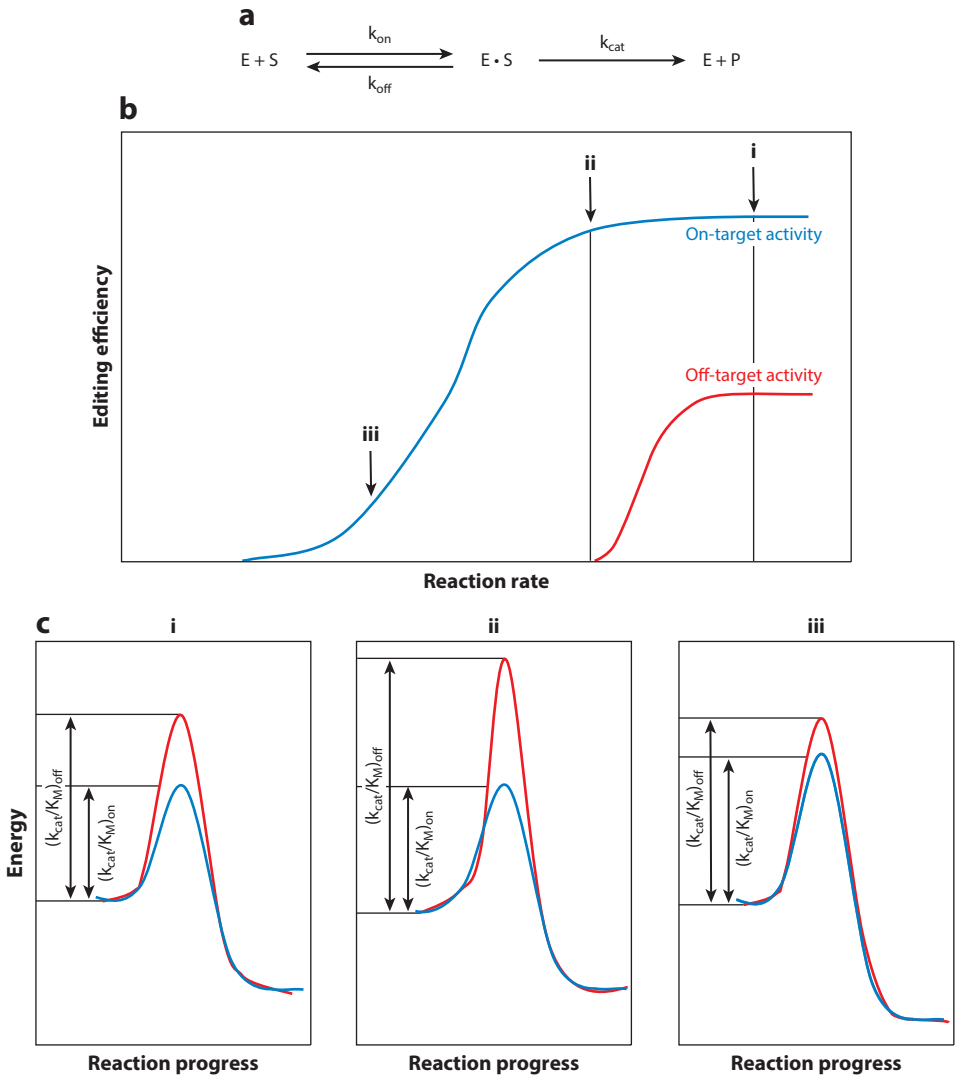


Figure 1

Editing efficiencies of various nucleases at on-target and off-target sites. (*a*) The simple Michaelis-Menten catalytic scheme for enzyme reactions, where E is the enzyme, S is the substrate, E•S is the enzyme–substrate complex, P is the product, k_{on} is the on rate, k_{off} is the off rate, and k_{cat} is the catalytic rate. (*b*) The optimal programmable nuclease catalysis rate for DNA cleavage at the target site should be chosen to maximize the discrimination between target-site cleavage rate and cleavage of near-cognate sequences within the genome. Maximizing the nuclease reaction rate for target-site cleavage (*i*) becomes counterproductive if it leads to promiscuous activity at off-target sites. Tuning the catalytic rate of the nuclease to maximize the discrimination between activity at the target site relative to the off-target site while maintaining high editing efficiency (*ii*) produces nucleases with improved specificity. Further reduction in the catalytic rate (*iii*) provides good discrimination but poor on-target activity. (*c*) Energy profiles of the transition state barrier for target-site (*blue line*) and off-target-site (*red line*) cleavage for the reaction rate profiles described in panel *b*. The optimal level of activity and discrimination (*ii*) is achieved by maximizing the difference in the transition state energy for target-site cleavage relative to off-target-site cleavage without dramatically increasing the barrier to target-site catalysis.

Although all of these methods are widely used to measure mutagenesis frequencies at a specific locus, they are dependent on unbiased genome-wide methods and computer algorithms to identify putative off-target sites.

Guide RNA (gRNA): crRNA or sgRNA that determines the specificity of an RNA-guided endonuclease in a complex with a nuclease protein component such as Cas9 or Cpf1

Protospacer: a CRISPR effector target sequence defined by the complexed gRNA; the seed region is a sub-element most sensitive to mismatches

Protospacer adjacent motif (PAM): a short CRISPR nuclease-specific 2–6 base pair DNA sequence adjacent to the protospacer sequence

R loop: a three-stranded nucleic acid structure comprising an RNA–DNA hybrid and an associated non-template single-stranded DNA

METHODS FOR DETECTING GENOME-WIDE OFF-TARGET SITES

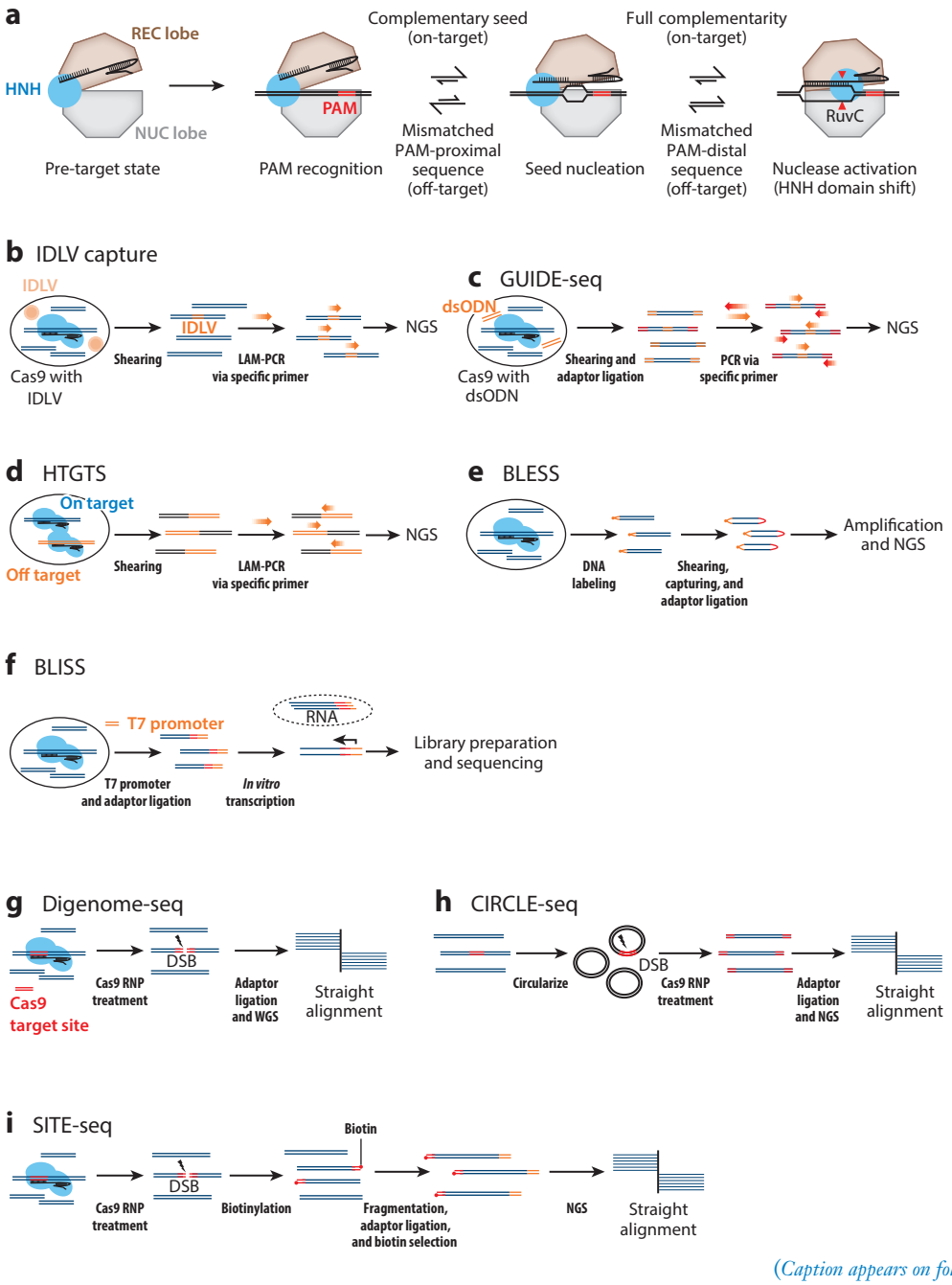
Methods for the genome-wide identification of potential nuclease off-target sites fall into three general categories: bioinformatic prediction based on sequence homology, *in cellulo* (cell-based) capture of editing events, and *in vitro* identification of cleavage sites using genomic DNA templates. These methods define a set of potential (or putative) off-target sites that require validation through targeted deep sequencing or another approach to verify the presence of nuclease-induced mutations. Typically, only a subset of identified potential off-target sites are validated as having InDels with frequencies above sequencing error rates (typically in the range of 0.01–0.1%). One metric for comparing the fidelity of editing for different nucleases is the off-target effect index (OTI), a measure of genome-wide off-target activity, which is defined as the sum of the mutation frequencies at validated off-target sites relative to the mutation frequency at the on-target site (25, 59). Only a few Cas9 and Cpf1 nuclease–guide RNA (gRNA) combinations have been characterized in sufficient depth to obtain OTIs, which range from 0 (no detectible off-target site activity) to 4.44 (more aggregate off-target activity than on-target activity) (59), demonstrating that off-target effects are highly variable among different gRNAs and different Cas9 or Cpf1 orthologs and variants.

Studies utilizing large-scale screens of gRNA variants for specific target sites have investigated the determinants of SpCas9 nuclease specificity (35, 60, 61). Not all mismatches between the gRNA and the target sequence have the same impact on SpCas9 nuclease activity: The number of mismatches, their positions within the target site, and the types of base mismatches all impact nuclease activity. In some cases, off-target sites have comparable or higher mutagenesis frequencies than the intended Cas9 target sites (61). Mismatches at the 5′ end of the protospacer sequence distal from the protospacer adjacent motif (PAM) are better tolerated than those in the seed region at the 3′ end of the protospacer proximal to the PAM, and consecutive mismatches are more disruptive than interspersed mismatches (35). PAM proximal mutations appear to negatively impact the rate of R-loop formation (62), whereas more distal mutations impact the allosteric regulation of the active conformation of the HNH domain, which permits DNA cleavage (48–50) (**Figure 2*a***).

For SpCas9, the type of base mismatches within the guide and PAM regions influence off-target activity. rG-dT is the most tolerated base mismatch within the RNA–DNA hybrid, whereas rC-dC mismatches are generally disfavored (35, 60). Likewise, SpCas9 can cleave at some suboptimal PAM sequences (e.g., 5′-NAG-3′, 5′-NGA-3′, and 5′-NCG-3′), albeit with lower efficiency (35, 60). Interestingly, a correlation exists between the number of active off-target sites for a SpCas9 guide combination and the resulting cellular toxicity of the nuclease. Guides with more near-cognate off-target matches within the genome typically produce higher nuclease toxicity (44, 45). Similar analyses have been undertaken to understand the sequence discrimination parameters for other Cas9 orthologs [from *Staphylococcus aureus* (SaCas9) (63) and *Neisseria meningitidis* (NmCas9) (64, 65)] and Cpf1 orthologs [from *Acidaminococcus* sp. BV3L6 (AsCpf1) and *Lachnospiraceae bacterium* ND2006 (LbCpf1) (59, 66, 67)], although the parameters influencing off-target discrimination for these nucleases are not as well defined.

Using these data sets, in silico prediction algorithms have been developed to identify and rank potential off-target sites for particular guides within the genome. Several web-based computer programs identify off-target sites by evaluating the similarity of a queried target sequence to all

possible sequences within a genome of interest, considering both the spacer sequence and canonical and noncanonical PAMs (**Table 1**). Some programs simply score off-target sites by their sequence similarity to the target sequence (relative orthogonality defined by the number of mismatches in the spacer and PAM) to provide a ranking of these sites. However, others use a form of



(*Caption appears on following page*)

Figure 2 (*Figure appears on preceding page*)

Influence of mismatches on Cas9 enzyme dynamics and unbiased genome-wide methods for identifying potential off-target sites. (*a*) Stages of target site engagement to yield a catalytically competent Cas9–gRNA–DNA complex. The Cas9–gRNA initially locates a compatible PAM element within the genome. R-loop formation between the DNA and gRNA is initiated neighboring the PAM. Mismatches within the seed region strongly disfavor progression of full R-loop formation with the entire guide sequence. Upon recognition of the cognate sequence, the HNH domain shifts to its active conformation, initiating DNA cleavage with the RuvC domain. Near-cognate (off-target) sequences are less likely to progress to the catalytically competent state. Mismatches at the PAM-proximal sequence of gRNA are less tolerated than mismatches at the PAM-distal portion. (*b–i*) Unbiased genome-wide off-target detection methods can be divided into two groups: cell-based methods and cell-free methods. Cell-based methods selectively tag genomic regions with DSBs to allow their selective amplification. These methods examine nuclease activity in the context of the epigenetic state and chromatin architecture within the nucleus of a cell, which can influence off-target nuclease activity: (*b*) Integrase-deficient lentiviral vector (IDLV) capture; (*c*) genome-wide, unbiased identification of DSBs evaluated by sequencing (GUIDE-seq); (*d*) high-throughput, genome-wide translocation sequencing (HTGTS); (*e*) direct in situ breaks labeling, enrichment on streptavidin, and next-generation sequencing (BLESS); and (*f*) breaks labeling in situ and sequencing (BLISS). Cell-free methods isolate purified genomic DNA or chromatin for *in vitro* reactions to determine sites of DNA modification as a function of purified nuclease (or deaminase) concentration. These methods may identify many potential off-target sites that are not cleaved in the cellular context: (*g*) digested genome sequencing (Digenome-seq); (*h*) circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq); and (*i*) selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-seq). Abbreviations: DSB, double-strand break; dsODN, double-stranded oligodeoxynucleotides; gRNA, guide RNA; LAM-PCR, linear amplification-mediated PCR; NGS, next-generation sequencing; PAM, protospacer adjacent motif; RNP, ribonucleoprotein; WGS, whole-genome sequencing.

specificity score calculation, which ranks guides on the basis of a weighted sum of the mismatches in their off-target sites, with weights experimentally determined to reflect the effect of the mismatch position on cleavage efficiency (35). More recently, Doench et al. (60) have developed a cutting frequency determination (CFD) score, which incorporates mismatch identity, along with position and number of mismatches, to calculate the off-target potential of single-guide RNA (sgRNA)–DNA interactions.

When compared with unbiased genome-wide off-target detection methods described below, the computational algorithms, although powerful and useful, often fail to identify the majority of the validated off-target sites. Tsai et al. (18) and Cameron et al. (51) demonstrated that in silico predictive algorithms, such as the E-CRISP software (**http://www.e-crisp.org/E-CRISP/**), CCTop (**https://crispr.cos.uni-heidelberg.de/**), and Cas-OFFinder (**http://www.rgenome.net/cas-offinder/**), were able to discover only a subset of putative off-target sites identified by unbiased genome-wide methods and could not effectively rank the active off-target sites in cells. These studies highlight the challenges faced by in silico predictive algorithms to comprehensively

Table 1 List of computational predictive algorithms for gRNA off-target analysis

Name	Website	Number of RGENs supported
E-CRISP	**http://www.e-crisp.org/E-CRISP/**	1
CCTop	**https://crispr.cos.uni-heidelberg.de/**	9
Cas-OFFinder	**http://www.rgenome.net/cas-offinder/**	11
CRISPRseek	**https://bioconductor.org/packages/release/bioc/html/CRISPRseek.html**	Unlimited
COSMID	**https://crispr.bme.gatech.edu/**	1
CROP-IT	**http://www.adlilab.org/CROP-IT/homepage.html**	1
CasFinder	**http://arep.med.harvard.edu/CasFinder/**	3
ChopChop	**http://chopchop.cbu.uib.no/**	4
CasOT	**http://casot.cbi.pku.edu.cn/**	1

Abbreviations: gRNA, guide RNA; RGEN, RNA-guided endonuclease (such as Cas9 or Cpf1).

identify and rank potential off-target sites given the multitude of factors that can influence nuclease activity throughout the genome.

UNBIASED GENOME-WIDE OFF-TARGET DETECTION METHODS

To circumvent the limitations of in silico predictive tools, several nuclease-based methods have been developed to identify potential off-target sites on a genome-wide level. These unbiased methods fall into two groups: *in cellulo* and *in vitro* techniques. *In cellulo* techniques typically utilize indirect readout of DSB formation through the capture of an exogenously supplied DNA tag or chromosomal translocation for the selective amplification of genomic regions with nuclease activity, whereas *in vitro* techniques capture nuclease-induced cleavage events directly on purified, cell-free genomic DNA. These unbiased methods vary in their comprehensiveness and sensitivity, which are defined by their ability to identify the full spectrum of off-target sites spanning high and low mutagenesis frequencies (characteristic of inefficiently cleaved DNA sequences) within the genome.

Cell-Based Methods

Current *in cellulo* unbiased methods include whole-genome sequencing (WGS); integrase-deficient lentiviral vector (IDLV) capture (68) (**Figure 2*b***); genome-wide, unbiased identification of DSBs evaluated by sequencing (GUIDE-seq) (18) (**Figure 2*c***); high-throughput, genome-wide translocation sequencing (HTGTS) (19) (**Figure 2*d***); direct in situ breaks labeling, enrichment on streptavidin; next-generation sequencing (BLESS) (69) (**Figure 2*e***); and Breaks Labeling In Situ and Sequencing (BLISS) (70) (**Figure 2*f***). As these methods require the introduction of nucleases directly into cells, they identify potential off-target sites within the genome in its native context (e.g., transcriptional status, chromatin architecture, epigenetic features, etc.). Among the *in cellulo* methods, GUIDE-seq and HTGTS are both similarly comprehensive (capturing a large number of putative off-target sites) and similarly sensitive with regard to the activity of the off-target sites that can be captured.

WGS. Whole-genome sequencing provides an unbiased method for measuring nuclease precision. WGS has been used for the analysis of nuclease-based mutagenesis for a small number of clonal edited cell lines (21, 71–73). These studies failed to identify any single-nucleotide polymorphisms (SNPs) or InDels that could be linked with SpCas9 off-target editing on the basis of sequence homology to the guide that was employed. WGS analyses of nuclease-treated mouse embryos have identified active off-target events in the offspring of treated founder animals, but these events were infrequent (74). Additionally, WGS analysis has been performed on SpCas9-treated human blastomeres, and whole-exome sequencing was performed on human embryonic stem cells derived from SpCas9-treated blastomeres (75). In both cases, no evidence of off-target mutagenesis was observed. Overall, these WGS studies provide the perspective that transient treatment with SpCas9 nuclease does not produce extensive collateral damage to the human genome. The primary drawback of WGS is the limited number of clones that can be analyzed, and consequently the data are not diagnostic of the mutagenesis rate over a large population of treated cells for sites with low mutagenesis frequencies, which is important when large numbers of cells (or a tissue) are treated in a therapeutic setting.

IDLV capture. IDLV capture is a tag-based genome-wide off-target detection method in which nuclease-mediated DSBs are identified by the propensity of IDLV to integrate in the vicinity of

DNA breaks (68) (**Figure 2*b***). IDLV integration sites within the genome are identified by selective amplification of genomic DNA flanking each site by linear amplification-mediated PCR (LAM-PCR), followed by library preparation and high-throughput sequencing (**Figure 2*b***). Importantly, IDLV capture works with a wide variety of different nuclease platforms (e.g., ZFNs, TALENs, and Cas9), allowing direct comparisons between different systems. For example, Wang et al. (76) found that TALENs were more specific than SpCas9 with regard to the number of validated off-target sites in a small sample set. The primary limitations of IDLV capture are its low sensitivity (InDel rates ≥0.5–1%) due to inefficient capture of IDLVs in DSBs (typically only a few tens of integration events occur at the target locus) and its high false positive rate due to IDLVs' ability to randomly integrate into a genome at a modest frequency even in the absence of nuclease-mediated DSBs.

Nonhomologous end joining (NHEJ): a double-strand break repair mechanism that ligates two ends without using a repair template that can produce InDels

GUIDE-seq. Developed by the Joung laboratory, GUIDE-seq is a tag-based method that relies on nonhomologous end joining (NHEJ)–mediated integration of exogenously supplied blunt, double-stranded oligodeoxynucleotides (dsODNs) of defined sequence into DSBs within the genome (18) (**Figure 2*c***). Following dsODN insertion, the tag and adjacent genomic sequences are selectively amplified and subjected to high-throughput sequencing (**Figure 2*c***). GUIDE-seq can detect off-target sites with mutation frequencies of 0.1% or lower in a population of cells. Consistent with other approaches for analyzing Cas9 activity at noncognate target sequences (35, 60), Tsai et al. (18) found that mismatch number, position, and type were the critical parameters that influence nuclease activity at off-target sequences. The primary limitation of GUIDE-seq is the requirement for high transfection efficiency of the dsODN. Thus, GUIDE-seq cannot be used in some cell types that are difficult to transfect or are sensitive to cellular dsODN levels. Nonetheless, GUIDE-seq is the most commonly employed unbiased *in cellulo* method for assessing nuclease off-target activity. It has been used to evaluate the specificity of many RNA-guided endonucleases (RGENs), such as Cas9 orthologs (SaCas9 and NmCas9) and Cpf1 orthologs (AsCpf1 and LbCpf1) (18, 63, 65, 67).

HTGTS. The presence of multiple DSBs within a genome can give rise to chromosomal rearrangement between these sites, such as chromosomal translocations. HTGTS utilizes chromosomal translocations between nuclease-induced "bait" DSBs and off-target "prey" DSBs to capture off-target sites within the genome (77). These translocations are detected by LAM-PCR as noncontiguous sequence junctions between the bait genomic locus and other genomic elements with active nuclease target sites (19) (**Figure 2*d***). Like IDLV capture, HTGTS is compatible with a variety of different nuclease platforms. A comparison of SpCas9 and TALENs by HTGTS revealed that SpCas9 appears to have fewer active off-target sites within the genome (19). Although HTGTS is useful owing to its high sensitivity, nuclease-induced translocations are inherently rare events, which require a high number of input genomes to sample deeply and a corresponding number of sequencing reads for detection.

BLESS/BLISS. BLESS directly captures DSBs genome wide via in situ ligation of biotinylated adaptors in fixed, permeable cells and enrichment of these genomic regions on streptavidin-coated beads (69) (**Figure 2*e***). BLESS, unlike other cellular off-target detection methods, does not depend on the endogenous cellular DNA damage repair machinery to tag a region where a DSB has occurred. However, BLESS can capture only DSBs present at a particular time interval (once repaired, the DSB site can no longer be captured). Furthermore, BLESS requires relatively high DNA input (number of cells) owing to the high background caused by spurious DSBs introduced during fixation and handling. This method has been used to evaluate

the specificity of genome editing for SpCas9 and SaCas9 at overlapping target sequences (78). This analysis suggested that SaCas9 may have fewer off-target sites than SpCas9: A similar conclusion was reached using GUIDE-seq for a different overlapping site for these two nucleases (63).

A more refined technique, breaks labeling in situ and sequencing (BLISS), has been developed to detect DSBs in low-input samples of cells and tissues (70). DSB ends are in situ blunted and ligated with dsODN adapters containing the T7 promoter sequence and linearly amplified via T7-mediated *in vitro* transcription; the subsequent RNA is used for Illumina library preparation and sequencing (**Figure 2*f***). BLISS was utilized to compare the editing precision of SpCas9, AsCpf1, and LbCpf1 (70), which revealed that both Cpf1 nucleases appear to have lower rates of off-target editing than SpCas9, consistent with prior studies using other genome-wide analysis methods (59, 67). In addition, on the basis of an aggregation of the active off-target sequences for the Cpf1 nucleases, Yan et al. (70) were able to define regions within the Cpf1 guide sequence that are more tolerant of mismatches.

Ribonucleoprotein (RNP): a complex consisting of gRNA and RNA-binding protein (Cas9/Cpf1); form used in the direct delivery of nuclease to cells

Next-generation sequencing (NGS): a sequencing technology, such as short-read Illumina sequencing, that typically produces millions of sequencing reads in a single experiment

Cell-Free Methods

In contrast to cell-based methods, *in vitro* methods are not restricted by nuclease or DNA delivery challenges, because these methods directly identify the position of DSB formation throughout the genome as a function of the Cas9–gRNA ribonucleoprotein (RNP) complex concentration. Three related, but distinct, methods—digested genome sequencing (Digenome-seq) (71) (**Figure 2*g***), circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq) (79) (**Figure 2*h***), and selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-seq) (51) (**Figure 2*i***)—permit in-depth definition of Cas9-induced cleavage events within isolated genomic DNA for a variety of different nuclease platforms.

Digenome-seq. Digenome-seq is highly sensitive, as it can detect off-target events with mutation frequencies as low as 0.1% (71). Genomic DNA purified from cells of interest is digested with purified Cas9 or Cpf1 RNP *in vitro* and then prepared for Illumina deep sequencing. Nuclease-digested fragments share the same 5′ ends at cleavage sites and thus can be identified in the aligned sequencing reads (**Figure 2*g***). Kim et al. (25) further improved Digenome-seq by refining the scoring system to more effectively capture sites where heterogenous nuclease-based DNA cleavage has occurred and by introducing multiple gRNAs for a single (multiplex) assay to assess the specificity of many SpCas9 RNPs simultaneously. Owing to background sequencing reads of non-nuclease–associated DSBs, Digenome-seq requires higher sequencing coverage (~400–500 million reads) to define the most comprehensive list of potential off-target sites. In addition, Digenome-seq is limited by the availability of a high-quality reference genome sequence.

In a comparative analysis with contemporary cell-based off-target detection methods (GUIDE-seq and HTGTS), all three approaches recovered an overlapping set of next-generation sequencing (NGS) validated off-target sites (71). However, all three methods also missed a subset of bona fide off-target sites captured by the other methods (25), suggesting that a comprehensive analysis of off-target sites for SpCas9 may require a combination of *in cellulo* and *in vitro* approaches.

In addition, Digenome-seq has been adapted to the genome-wide assessment of the off-target activity of the APOBEC1–nCas9 base editor (BE3) (80). Purified genomic DNA is treated with BE3 RNPs, which can nick the complementary strand of the DNA and deaminate cytosine to uracil in the non-template strand. Uracils within the genomic DNA are processed with USER (uracil-specific excision reagent) to generate a gap opposite the nick, giving rise to a composite

DSB. These DSBs are then identified through the Digenome-seq pipeline (80). Comparison of BE3 and SpCas9 off-target events demonstrated that BE3 deaminases appear to be more specific than SpCas9 nucleases based on the number of potential off-target sites captured by Digenome-seq and the OTIs determined with seven sgRNAs in complex with BE3 or SpCas9. Notably, BE3 and SpCas9 off-target sites do not always coincide, calling for independent assessments of each type of editing tool.

CIRCLE-seq. CIRCLE-seq utilizes an alternate genomic substrate, DNA circles, for the *in vitro* genome-wide identification of nuclease off-target cleavage events (79). Genomic DNA is sheared into linear fragments and then circularized by intramolecular ligation prior to nuclease cleavage analysis (**Figure 2*b***). Any residual linear genomic DNA is degraded by exonuclease treatment prior to the nuclease assay. Following nuclease treatment, circularized DNA molecules containing a functional nuclease site are linearized, selectively amplified by adaptor ligation, and subjected to high-throughput sequencing to identify the genomic region that is cleaved (**Figure 2*b***). By employing DNA circles, instead of linear genomic DNA, CIRCLE-seq reduces the background of sequencing reads resulting from random DNA breaks within the input genomic DNA that occur during purification and processing, producing higher sensitivity.

CIRCLE-seq is able to provide a comprehensive list of genome-wide off-target sites while requiring lower sequencing coverage (~4–5 million reads). At common SpCas9 targets sites, CIRCLE-seq identified many more off-target sites than were previously found by both GUIDE-seq and HTGTS. However, standard high-depth targeted amplicon sequencing is unable to validate off-target sites with low mutagenesis frequencies (<0.1%), as any nuclease-produced InDels are obscured by the error rate of NGS. To validate these weak potential off-target sites identified by CIRCLE-seq, target amplicon sequencing was performed on genomic DNA obtained from cell-based GUIDE-seq experiments to measure the frequency of tag integration, which is easily distinguished from the sequencing error rate. This analysis revealed that approximately a quarter of these weak sites displayed evidence of nuclease activity (tag integration) in cells.

SITE-Seq. Similar to Digenome-seq, SITE-seq uses SpCas9–gRNA RNPs to cleave purified genomic DNA. In SITE-seq, high-molecular-weight genomic DNA is used to minimize the number of breaks that are present within the genome (51). Following nuclease treatment, cleaved DNA is selectively tagged with biotin prior to DNA fragmentation, enriched by streptavidin pull-down, and subjected to sequencing to reduce the background of non-nuclease associated sequencing reads (**Figure 2*i***). Enrichment of Cas9-induced cleavage sites enables specificity profiling with minimal read depth (~0.62–2.46 million reads). In comparison with other unbiased genome-wide off-target detection methods (GUIDE-seq, HTGTS, and Digenome-seq), SITE-seq recovered previously identified, validated off-target sites as well as additional potential off-target sites, a subset of which were subsequently validated. Like the other *in vitro* methods, the number of identified off-target sites is a function of the nuclease concentration that is employed in the *in vitro* digestion reaction. Many hundreds of potential off-targets can be recovered with high nuclease levels, but most of these sites are unlikely to be validated in NGS analysis of InDels in nuclease-treated cells.

DIG-seq. It remains impractical to validate all off-target sites in cells using NGS, especially when candidate lists reach the hundreds, as observed with CIRCLE-seq and SITE-seq. As noted above, many potential sites identified by the various *in vitro* methods are not edited in cells. This suggests the possibility that other genomic factors may play a role in modulating cleavage activity at putative off-target loci. ChIP-seq studies with catalytically deficient Cas9 (dCas9)–sgRNA in

Table 2 Comparison of several unbiased genome-wide methods at a single common target site, VEGFA TS1

Off-target method	Common target site	Number of identified potential off-target sites	Validation rate[a]
GUIDE-seq	VEGFA TS1	22	90.9%
Digenome-seq		69	53%
CIRCLE-seq		496	29%
SITE-seq		996	10%
DIG-seq		31	73%

Abbreviations: CIRCLE-seq, circularization for *in vitro* reporting of cleavage effects by sequencing; DIG-seq, digested genome sequencing using cell-free chromatin DNA; Digenome-seq, digested genome sequencing; GUIDE-seq, genome-wide, unbiased identification of double-strand breaks evaluated by sequencing; SITE-seq, selective enrichment and identification of tagged genomic DNA ends by sequencing; TS1, target site 1; VEGFA, vascular endothelial growth factor A.

[a]Number of sites validated by next-generation sequencing/number of potential off-target sites.

cells have established that the chromatin environment influences the binding of dCas9 within the genome (81, 82). Additionally, Kim & Kim (40) demonstrated, in a comparison of SpCas9 nuclease activity at a series of identical endogenous DNA sequences present in both open and closed chromatin regions, that chromatin accessibility influenced Cas9-mediated mutagenesis at cognate sites and near-cognate (mismatched) sites, where the latter were more strongly impacted by a closed chromatin state.

Together, these observations led to the development of DIG-seq (Digenome-seq using cell-free chromatin DNA) (40), an improved version of Digenome-seq. In contrast to previously described *in vitro* methods, DIG-seq accounts for chromatin environment in eukaryotic cells by using native chromatin for the nuclease-based *in vitro digestion* as opposed to purified genomic DNA. The resulting DNA is processed through the standard Digenome-seq pipeline. Compared with previous *in vitro* methods, DIG-seq produced shorter lists of potential off-target sites. Furthermore, the validation rate (number of bona fide off-target sites/number of predicted off-target sites) of DIG-seq is higher than for the other *in vitro* methods (**Table 2**) (40). These results suggest that DIG-seq can be widely adapted for the identification of nuclease-induced off-target sites in a comprehensive, yet succinct, manner.

Comparative Assessments of the Off-Target Detection Methods

To compare the various off-target detection methods for specificity and comprehensiveness, standard gRNAs have been used as benchmarks. For example, all of the unbiased genome-wide methods assessed at least one common gRNA from a group (18, 19, 25, 40, 51, 69, 70, 76, 79), although no gRNA appears in all of these studies. Although some degree of overlap in the number of validated off-target sites identified by each method was found, distinct off-target sites were also discovered by each method, suggesting that none of these methods may be completely comprehensive. However, this may also be a function of the depth of the analysis performed for each method. Most cell-based methods typically produce a list of ~5–40 potential off-target sites, occasionally these numbers reach into the 100s for promiscuous gRNAs in GUIDE-seq. Most of the *in vitro* off-target assessment methods produce a list of >100 potential off-target sites. In many cases, *in vitro* methods identified most, if not all, of the sites identified by *in cellulo* methods (**Table 2**), in addition to several new off-target sites. However, as the number of potential off-target sites that are identified by the *in vitro* methods increases, the validation of this large list of potential off-target sites can be both time consuming and cost inefficient, with only a fraction of these sites proving positive in the validation.

Despite these caveats, it is important to emphasize that an extraordinary amount of progress has been made in the development of methods to determine active nuclease off-target sites within the genome. A number of methods now exist that should provide a near-comprehensive list of potential off-target sites for any specific nuclease–gRNA combination. Factors such as sensitivity, cost, ease of implementation, and cellular context will influence the off-target detection method that is most suitable for a particular application. In addition, new methods, such as DIG-seq, are focused on improving both the sensitivity and the validation rate, which should further streamline the assessment of genome-wide off-target sites for both nucleases and BEs.

METHODS FOR REDUCING OFF-TARGET EFFECTS

Unlike ZFNs and TALENs, whose specificities must be enhanced through individual optimization via protein engineering, CRISPR systems can be improved in a predictable manner by modifying either gRNA scaffolds or nuclease proteins without reengineering them one by one. Here, we focus on various methods for reducing or avoiding off-target effects by CRISPR nuclease systems through modifications to one of these components while retaining robust on-target activity.

gRNA Modification

The majority of analysis on the impact of guide modifications on nuclease off-target activity has focused on SpCas9 gRNAs (**Figure 3*a***). The Joung laboratory demonstrated that truncated gRNAs (tru-gRNAs), with only 17 or 18 (instead of 20) nucleotides complementary to the target site (**Figure 3*b***), are more sensitive to mismatches than full-length gRNAs (83). These truncated guides decreased SpCas9-induced off-target mutations by up to 5,000-fold. GUIDE-seq experiments largely confirmed the improved specificity of tru-gRNAs (18) but also showed the presence of a small number of tru-gRNA–specific off-target sites with mismatches at the PAM distal end that no longer overlap with the truncated guide. Increased off-target activity for terminally mismatched sites has also been observed in other studies (80, 84). For SpCas9, truncated guides appear to operate by destabilizing cleavage complex formation, by reducing the lifetime of the R-loop complex (49), and by disfavoring the active conformation of the HNH domain (48). Interestingly, tru-gRNAs have not proven to be as potent in reducing off-target activity with other nuclease systems, such as Cpf1 (67).

Extended gRNAs with >21 complementary nucleotides, in general, do not enhance SpCas9 specificity or its efficiency at on-target sites. However, extended gRNAs that include two extra guanine nucleotides at the 5′ end (termed ggX_{20} gRNAs) can reduce off-target effects by a few orders of magnitude often without reducing on-target activities (21) (**Figure 3*c***). The molecular mechanism behind the improved specificity of ggX_{20} gRNAs remains unknown. The two extra guanines may impede Cas9 interaction with DNA by increasing the off rate, thereby leading to greater kinetic discrimination for cleavage of off-target sites than on-target sites.

Chemically modified gRNAs have been used to improve CRISPR specificity. For example, the substitution of deoxyribose nucleotides for ribose nucleotides within the 5′ end of the sgRNA can reduce or eliminate SpCas9 activity at off-target sites (85) (**Figure 3*d***). In one study, a chemical modification (2′-*O*-methyl-3′-phosphonoacetate, or MP) at a specific base in a gRNA reduced InDel frequency at an off-target site from 48.9% to 1.03% without sacrificing the on-target efficiency (86) (**Figure 3*e***). Likewise, incorporation of bridged nucleic acids (2′,4′-BNA^{NC}[N-Me]) in a CRISPR RNA (crRNA) drastically reduced Cas9 off-target activity by up to 25,000-fold (87) (**Figure 3*e***). Single-molecule fluorescence resonance energy transfer (FRET) experiments

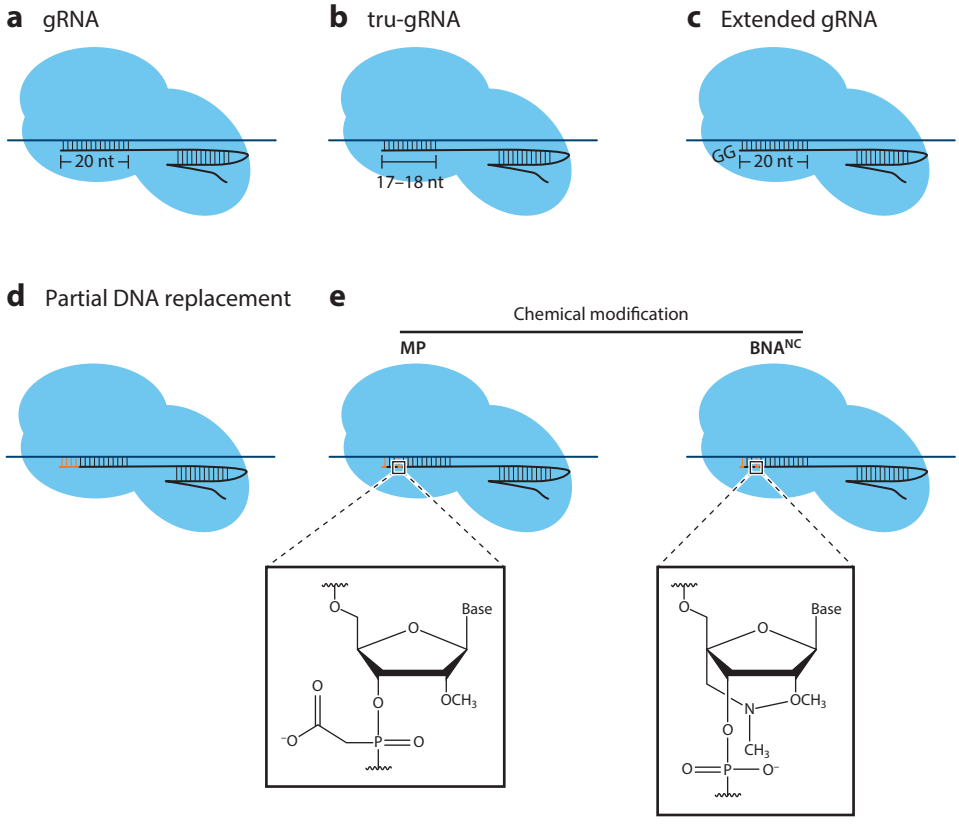


Figure 3

Schematic of the various guide RNA (gRNA) modifications used to improve nuclease specificity. (*a*) Cas9 guided by full-length gRNA (i.e., 20 nt for SpCas9). (*b*) Truncated gRNAs (tru-gRNAs) 17–18 nt in length improve nuclease specificity by reducing nuclease activity to increase discrimination. (*c*) Extended gRNAs add two guanines at the 5′ end of the guide, which improves discrimination between on-target and off-target sites. (*d*) Partial DNA (*orange*) substitution of RNA within the guide produces a chimeric RNA–DNA hybrid guide with improved discrimination. Non-natural chemical modification of the guide through the incorporation of (*e*) 2′-*O*-methyl-3′-phosphonoacetate (MP) or bridged nucleic acids (2′,4′-BNA^{NC}[N-Me]) at specific nucleotides within a gRNA sequence can also improve sequence discrimination.

showed that BNA^{NC} incorporation into crRNAs improves Cas9 specificity by slowing the kinetics of Cas9–DNA interaction.

Delivery of CRISPR Systems

The use of preassembled Cas9 RNPs rather than plasmid DNA or viral vector expression systems has been shown to reduce off-target effects of SpCas9 (52, 53), Cpf1 (59), and BEs (88) by an order of magnitude without sacrificing on-target editing rates. RNPs act on target DNA immediately after transfection and are rapidly degraded by proteases and ribonucleases in cells, usually within 24 hours (53). In contrast, proteins are expressed for several days from plasmid DNA and potentially expressed indefinitely from viral delivery systems, leading to the continued accumulation of off-target mutations after target-site editing is completed. RNPs can be delivered

into cells via electroporation (53), lipofection (89), nanoparticles (90, 91), and cell-permeable peptides (52).

Recombinant adeno-associated virus (rAAV): single-stranded DNA vectors used for delivery based on a class of viruses that commonly inhabit human hosts without causing any detectable pathology

Modified Cas9 Systems

Many different modifications of the Cas9 system have been employed to achieve improved specificity. The specificity of Cas9-based systems can be improved by requiring the assembly or DNA recognition of two components to achieve DNA cleavage or through the drug-dependent activation of Cas9 to restrict nuclease activity to a specific temporal window.

RNA-guided FokI–dCas9 nucleases. Similar to ZFNs and TALENs composed of DNA-binding domains (ZFP or TALE proteins) fused to the FokI nuclease moiety, RNA-guided FokI–dCas9 nucleases (RFNs) are composed of dCas9 fused to FokI (92, 93) (**Figure 4*a***). RFNs function as dimers, requiring two gRNAs to cleave target DNA (**Figure 4*a***). RFNs were able to increase specificity by up to 140-fold, compared with Cas9 nucleases, presumably through the recognition of more extended DNA sequences (92). In addition, RFNs complexed with tru-gRNAs further improved nuclease specificity (94). The genome-wide target specificity of RFNs has yet to be determined using unbiased genome-wide methods.

Paired Cas9 nickases. Cas9 contains two nuclease domains (RuvC and HNH), which cleave opposite strands of the DNA target to produce a blunt DSB (4). A SpCas9 nickase cleaves only one DNA strand by mutating one of the nuclease domains (D10A or H840A mutation) (4, 5, 95). When nCas9 are targeted to neighboring sites on the DNA to cleave opposite strands (known as paired nicking), a composite DSB can be generated (21, 96, 97) (**Figure 4*b***), although DSB efficiency depends on the type of nickase (D10A is better than H840A) and the relative orientation of the nickases (a DSB with 5′ overhangs is superior). Double nicking with D10A nCas9 increased specificity between 50- to 1,500-fold in one study (96) and up to 990-fold in an independent study (21) in cell lines compared with standard Cas9 nuclease without sacrificing on-target activity. Frock et al. (19) used HTGTS to show that double nicking via paired nCas9 reduced the number of detectable off-target sites to zero in the human genome. nCas9, however, can induce InDels at nicked sites, albeit at lower frequencies, compared with respective Cas9 nucleases (21, 83, 93). The repair of SSBs often leads to exposure of single-strand DNA for cytosine deamination and subsequent uracil removal, causing DSBs (98).

Split Cas9. SpCas9 can be split into two fragments, which can spontaneously be assembled into a functional protein with lower DNA cleavage efficiency but higher specificity, compared with the intact single-chain protein (99). To promote fragment assembly, the N-terminal domain and the C-terminal domain are fused to the rapamycin-binding dimerization domains, FRB and FKBP, to allow rapamycin-inducible genome editing (100) (**Figure 4*c***). The split SpCas9 domains have also been fused to split inteins, which can then be packaged into recombinant adeno-associated virus (rAAV) vectors to permit a larger fraction of the AAV packaging capacity to be used for other cargo in gene therapy applications (101, 102).

Cas9 Fusion and Disruption

Intein-disrupted SpCas9 and SpCas9 fused with DNA-binding proteins have been shown to improve the specificity of SpCas9.

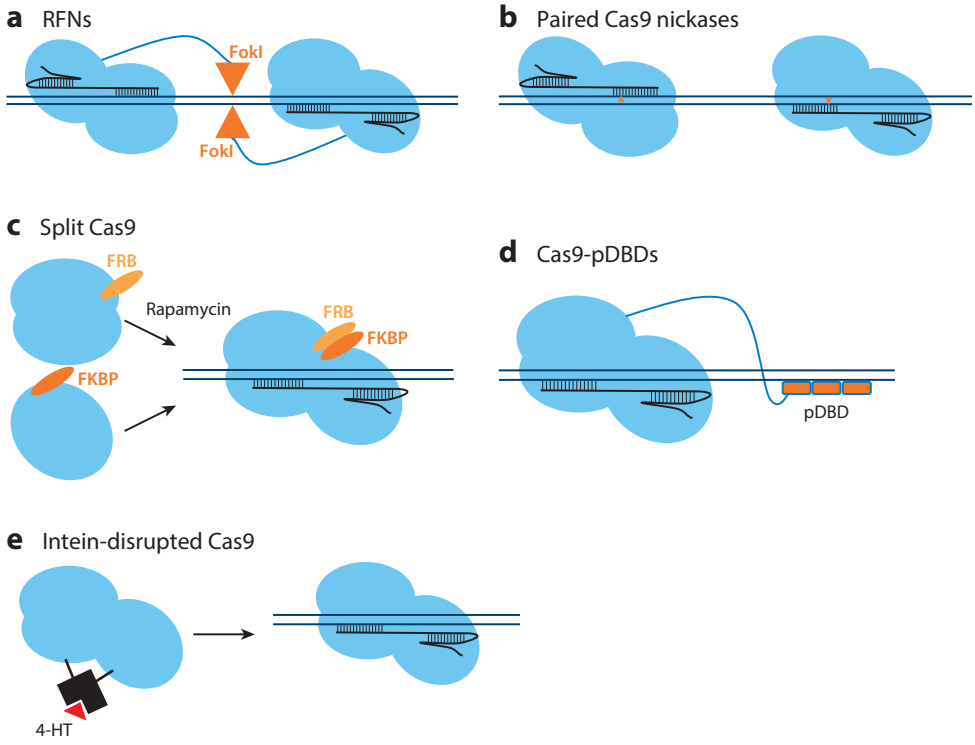


Figure 4

Dimeric or engineered Cas9 systems with improved precision. (*a*) RNA-guided FokI–dCas9 nucleases (RFNs) contain a catalytically deficient Cas9 (dCas9) fusion to the nuclease domain of FokI. RFNs function as a dimer to cleave DNA. (*b*) Cas9 nickases, in which one of the catalytic domains (either RuvC or HNH) is inactivated via loss-of-function mutation. These constructs are used in pairs to generate a DSB by the production of nicks on opposite DNA strands in close proximity. (*c*) In split Cas9, N-terminal and C-terminal fragments of Cas9 are fused to FRB and FKBP domains. These fragments can dimerize in the presence of rapamycin to assemble a functional nuclease. (*d*) Cas9-pDBDs (programmable DNA-binding domains) contain Cas9 with attenuated DNA-binding affinity fused to a pDBD. Nuclease activity is dependent on the recognition and binding of the pDBD to a sequence neighboring the target site. (*e*) Intein-disrupted Cas9 is a pharmacologically controlled nuclease system inactive in its native state owing to the insertion of a drug-dependent intein within the protein sequence. The nuclease is activated via treatment with 4-hydroxytamoxifen (4-HT) to remove the intein from the Cas9 protein.

Cas9-pDBDs. To improve the cleavage specificity of SpCas9, the Wolfe laboratory developed a Cas9-pDBD (programmable DNA-binding domain) system, which consists of a fusion of a Cys_2His_2 zinc-finger protein (ZFP)—a type of pDBD—to SpCas9, with an attenuating PAM-binding mutation ($SpCas9^{MT}$) that restricts its ability to engage its target site (103) (**Figure 4*d***). The ZFP fusion increases the recognition site length by recognizing a specific sequence neighboring the SpCas9 target site. When compared with SpCas9, the $SpCas9^{MT}$-ZFP nuclease has increased specificity up to 160-fold at off-target sites, and GUIDE-seq–mediated genome-wide off-target analysis with three different guides demonstrated that the number of off-target sites is decreased from 41 to 3, 6 to 0, and 4 to 0, respectively. This system has recently been extended to orthogonal Cas9-Cas9 fusion proteins, which provide a more flexible framework for programming the DNA-binding specificity of the pDBD component (104).

Intein-disrupted Cas9. The Liu laboratory inserted an evolved ligand-dependent intein at specific positions in SpCas9 to develop a pharmacologically controllable genome-editing system (105). The intein-inserted Cas9 is inactive; however, when treated with 4-hydroxytamoxifen (4-HT), which removes the intein from the Cas9 protein, the nuclease becomes activated (**Figure 4*e***). Although this system is less efficient at editing in human cells compared with wild-type SpCas9, pharmacologically activated Cas9 achieved editing with up to 25-fold higher specificity, compared with wild-type SpCas9, with minimal editing in the absence of a drug.

Protein Engineering via Structure-Based Design

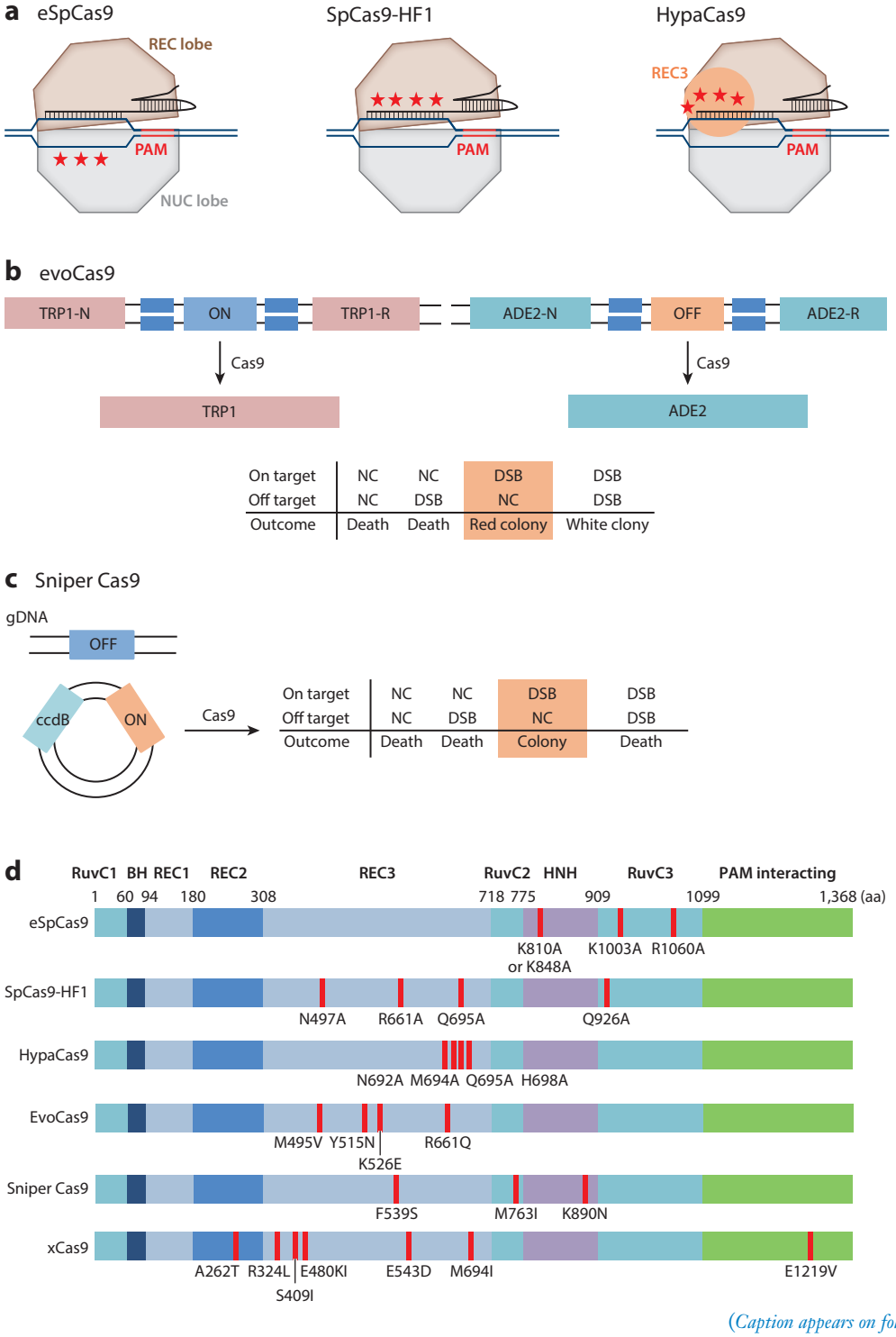
High-resolution 3D structures of SpCas9 complexed with gRNA and target DNA (106–108) provided detailed insights into its mode of target site recognition, which allowed protein engineering to improve its specificity. Three SpCas9 variants with high specificity have been developed via structure-based, rational design.

eSpCas9. The Zhang laboratory hypothesized that weakening the affinity of Cas9 for the nontarget DNA strand (**Figure 5*a***), through substitution of several positively charged residues with alanine residues, could destabilize Cas9 binding at off-target sites containing mismatches and thereby reduce their mutation rates while retaining on-target activity. They created enhanced SpCas9 (eSpCas9) variants with alanine substitutions at four positions [eSpCas9 1.0 (K810A, K1003A, and R1060A) and eSpCas9 1.1 (K848A, K1003A, and R1060A)] (84) (**Figure 5*d***). These engineered SpCas9 variants displayed high sensitivity to single and double mismatches between sgRNA and target DNA. They also compared the genome-wide specificities of SpCas9 and eSpCas9 using BLESS and targeted deep sequencing, showing that eSpCas9 reduced the number of active off-target sites within the human genome without generating any new off-target sites.

SpCas9-HF1. The Joung laboratory used a similar strategy in their development of a high-fidelity Cas9 variant. However, they focused mutagenesis on four SpCas9 residues (N497, R661, Q695, Q926) that form hydrogen bonds with the phosphate backbone of the target DNA strand (109) (**Figure 5*a*,*d***). Accordingly, SpCas9-HF1 (high fidelity 1) with alanine mutations at these four residues was constructed. GUIDE-seq analysis showed that SpCas9-HF1 had far fewer off-target sites (0.1 on average) than wild-type SpCas9 (8.1 on average) in human cells.

The SpCas9-HF1 and eSpCas9 variants in complex with full-length gRNAs retain high on-target activity at many sites but are poorly active when combined with tru-gRNAs (109, 110). Consistent with this observation, both SpCas9-HF1 and eSpCas9 retain high on-target activity when utilizing a GX_{19} gRNA ("G" is a matched guanosine) to target a site with a complementary 5′ G nucleotide at position 20 distal from the PAM, but these nucleases often have weak activity when utilizing a gX_{19} gRNA ("g" is a mismatched guanosine) to target a site with a mismatch at the 5′ terminus (111, 112). Unlike wild-type SpCas9, these two attenuated variants are sensitive to a terminal mismatch at the 5′ end of the guide sequence, revealing unequivocal contribution of the 5′ terminal sgRNA–DNA interaction to the specificity of Cas9–DNA recognition. By fusing an sgRNA to a self-cleaving ribozyme or tRNA to produce 5′-matched (non-G) sgRNAs, the activity of SpCas9-HF1 and eSpCas9 can be at least partially rescued at target sites with a 5′ adenine, cytosine, or thymine nucleotide (111, 112).

HypaCas9. FRET-based biophysical studies of the conformational change in the structure of the SpCas9–sgRNA–DNA complex indicate that the HNH domain is more likely to spend appreciable time in an active conformation with the cognate target sequence, or a near-cognate DNA



(Caption appears on following page)

Figure 5 (*Figure appears on preceding page*)

Protein engineering strategies to improve precision of the Cas9 system. (*a*) Structure-based protein engineering of Cas9 to improve nuclease specificity. Mutations are introduced to increase the threshold of gDNA complementarity necessary for nuclease activity. Schematic overview of eSpCas9 (1.0 or 1.1), SpCas9-HF1, and HypaCas9 nuclease systems with position of mutations (*red stars*). eSpCas9 and SpCas9-HF1 attenuate contacts with the phosphodiester backbone. HypaCas9 mutations increase the stringency of the allosteric regulation of HNH domain activity (see **Figure 2*a***). (*b*) In the evoCas9 system, a yeast-based screening method is used to identify nucleases with improved discrimination between cleavage of the target site (ON) and the off-target site (OFF) from a library with random mutations within the recognition (REC3) domain of SpCas9 protein. Cleavage at the target site leads to reconstitution of a positive selectable marker (TRP1) allowing cell survival, whereas cleavage at the off-target site reconstitutes ADE2 function, which produces white colonies instead of red (default). The desired SpCas9 clones (*orange box*, chart) produce primarily red colonies. (*c*) In the Sniper Cas9 system, positive and negative selection systems (sniper screening) in *Escherichia coli* are used to identify clones with improved specificity. The Cas9 target sequence was placed in a plasmid encoding *ccdB*, a lethal gene, and an off-target sequence with two-nucleotide mismatches was placed in the *E. coli* genome. Cells containing SpCas9 clones with the desired activity produce a colony by destroying the ccdB plasmid without cleavage of the off-target site within the genome (*orange box*, chart). (*d*) Position and identity of mutations within engineered SpCas9 variants with improved precision. Mutations are marked (*red bars*). Abbreviations: DSB, double-strand break; eSpCas9, enhanced SpCas9; evoCas9, evolved Cas9; gDNA, guide DNA; HF1, high fidelity 1; HypaCas9, hyper-accurate Cas9 variant; NC, no cleavage; NUC lobe, nuclease lobe; PAM, protospacer adjacent motif; REC lobe, recognition lobe; xCas9, expanded PAM Cas9 variant.

sequence when the mutations are distal from the PAM element (113). SpCas9-HF1 and eSpCas9 1.1 reduce off-target cleavage by increasing the threshold of guide complementarity to the DNA sequence, which is necessary to achieve HNH conformational activation. That is, when binding with mismatched targets, they remain trapped in an inactive state that reduces DNA cleavage rates (113). A domain within the Cas9 recognition (REC) lobe (REC3) (amino acids 308–718) interacts with the RNA–DNA heteroduplex, acting as an effector that allosterically regulates the activation of the HNH nuclease domain (48). Based on this mechanistic hypothesis, the Doudna laboratory and its collaborators developed the hyper-accurate Cas9 variant (HypaCas9) through mutations in the REC3 domain (N692, M694, Q695, H698) (**Figure 5*a*,*d***), which raises the threshold of gRNA complementarity to the DNA necessary for allosteric activation of the HNH nuclease domain beyond SpCas9-HF1 or eSpCas9 1.1. GUIDE-seq analysis showed that HypaCas9 is more specific than wild-type SpCas9, reducing the number of off-target sites from 33 to 3.5 on average. It is not known whether HypaCas9 can be combined with tru-gRNAs or extended gRNAs to further reduce its residual off-target activity.

Protein Engineering via Cell-Based Selection Methods

As discussed above, rationally designed SpCas9 proteins have low activity at certain target sites that are accessible to wild-type SpCas9. Additionally, these proteins cannot be combined with tru-gRNAs or extended gRNAs to further reduce their residual off-target activity. As an alternative to the structure-based rational design approach, selection methods employing randomized libraries can be employed to isolate Cas9 variants with improved specificity and activity.

evoCas9. Casini et al. (114) developed a yeast-based screening system to isolate SpCas9 variants that actively cleave a target sequence in a reporter gene while discriminating against a near-cognate off-target sequence in another reporter gene. They created a library of random mutants within the REC3 domain of Cas9. After isolating and characterizing several mutant clones with improved off-target discrimination, they created SpCas9 variants with different combinations of these mutations and tested them in a human cell line using an enhanced green fluorescent protein disruption assay (**Figure 5*b***). One of the most promising mutants [evolved Cas9 (evoCas9)] contained four mutations (M495V + Y515N + K526E + R661Q) (**Figure 5*d***). They used GUIDE-seq to show that evoCas9 was more specific than SpCas9-HF1 or eSpCas9 1.1: The number of genome-wide

off-target sites identified using GUIDE-seq was, on average, 1.6 ($n = 8$) with evoCas9, 5.9 with SpCas9-HF1, 7.6 with eSpCas9 1.1, and 129 with wild-type SpCas9. However, evoCas9 was not compatible with tru-gRNAs or extended gRNAs to further reduce its off-target activity.

Sniper Cas9. Lee et al. (110) performed so-called sniper screening in *Escherichia coli*, a directed evolution approach, to create SpCas9 variants with high precision (110). In this positive and negative selection system, a single copy of a Cas9 target sequence was placed in a plasmid encoding *ccdB*, a lethal gene, and an off-target sequence with two-nucleotide mismatches was placed in the *E. coli* genome (**Figure 5*c***). Among millions of SpCas9 mutants produced by random mutagenesis, only those that efficiently cleaved the on-target site in the *ccdB* plasmid and did not appreciably cleave the off-target site in the genome survived (**Figure 5*c***). The resulting SpCas9 mutants were tested in human cells. The best-performing SpCas9 variant (Sniper Cas9) contained three mutations (F539S, M763I, and K890N) (**Figure 5*d***). Digenome-seq analysis showed that Sniper Cas9 cleaved far fewer sites in the human genome than wild-type Cas9. Unlike the other engineered Cas9 variants discussed above, Sniper Cas9 was fully compatible with extended or truncated gRNAs to reduce its remaining off-target activity and performed well in a preassembled RNP format to enable DNA-free genome editing.

xCas9. The Liu laboratory used phage-assisted continuous evolution (PACE) to create SpCas9 variants with broad PAM compatibility (115). In this system, the selection of dCas9 with suboptimal PAM sequences was linked to phage propagation. dCas9 variants that recognized a broad range of PAM sequences were selected after three rounds of continuous evolution. The best-performing expanded PAM Cas9 (xCas9) nuclease variant in an *E. coli*–based PAM depletion assay contained seven mutations (xCas9 3.7, containing A262T, R324L, S409I, E480K, E543D, M694I, and E1219V) and demonstrated an expanded PAM-targeting range in human cells, recognizing NG, GAA, and GAT PAM sequences (**Figure 5*d***). Furthermore, cytosine or adenine BEs composed of xCas9 variants and nucleobase deaminases allowed base editing at sites with non-NGG PAM sequences, expanding the range of targetable sequences for the dSpCas9-type BE. Despite its expanded PAM tolerance, GUIDE-seq analysis showed that xCas9 3.7 outperformed wild-type Cas9, displaying off-target activity at far fewer sites and with lower mutagenesis frequencies.

Cas9 Orthologs and Cpf1

Several Cas9 orthologs other than SpCas9 have also been successfully repurposed for genome editing in eukaryotic cells, including cultured human cells (63–65, 116–118). Some of these Cas9 orthologs are smaller than SpCas9 and, thereby, can be delivered *in vivo* via rAAV vectors, which have a limited cargo size (~4.7 kbp). For example, SaCas9 consists of 1,053 amino acid residues (encoded by a 3.16-kbp gene), which is smaller than SpCas9 (1,368 amino acid residues, 4.10 kbp), and SaCas9 displays better specificity. The gene encoding SaCas9, its regulatory elements, and an sgRNA sequence can be readily packaged in rAAV vectors, enabling efficient gene knockout in the mouse liver (78). Likewise NmCas9 is a compact Type IIC effector that consists of 1,081 amino acids (3.24 kb) that can edit human cells with high specificity (64, 65, 119, 120). NmCas9 nuclease can also be packaged in a single AAV with its guide to achieve editing within the mouse liver with minimal off-target effects (121). Cas9 derived from *Campylobacter jejuni* (CjCas9), composed of 984 amino acid residues (2.95 kb), is the smallest Cas9 ortholog characterized to date (118) and can be coexpressed with a fluorescent marker protein such as green fluorescent protein for ease of tracking *in vivo* via an all-in-one rAAV vector (116). Kim et al. (116) used Digenome-seq to demonstrate that CjCas9 is more specific than SpCas9 and SaCas9 in human or mouse cells,

cleaving far fewer or no off-target sites within the genome. The high specificity of CjCas9 can be attributed in part to its extended, 4-nucleotide PAM sequences (5′-NNNNRYAC-3′, where R is A or G, and Y is C or T) and 22-nucleotide protospacer sequence, compared with 2-nucleotide PAM sequences (5′-NGG-3′) and the 20-nucleotide protospacer sequence recognized by SpCas9.

Cpf1 (also known as Cas12a) is an RGEN of the Class II Type V CRISPR-Cas family (8, 122). Unlike Cas9, which is guided by two small RNAs, namely the crRNA and trans-activating crRNA (tracrRNA), or their fusion into a single-chain sgRNA (103 nucleotides in length) (4), Cpf1 is guided by a single, ~43-nucleotide crRNA. Some Cpf1 nucleases (1,200 to ~1,500 amino acid residues, 3.6 to ~4.5 kbp) are smaller in size than SpCas9 and consequently are amenable to co-expressing with their crRNA *in vivo* via a single rAAV vector (123, 124). Unlike SpCas9, which recognizes G-rich PAMs, Cpf1 recognizes T-rich PAMs, allowing Cpf1 to target sequences that cannot be cleaved by SpCas9. Genome-wide target specificities of Cpf1 nucleases have extensively been examined using Digenome-seq (59), GUIDE-seq (67), and BLISS (70). These studies independently showed that Cpf1 is highly specific in human cells, often cleaving no off-target sites in the human genome. Thus, Cpf1 is another promising platform for the development of highly specific genome-editing tools.

CRISPR SPECIFICITY PARADOX

CRISPR nucleases, which originated from prokaryotic adaptive immune systems, are expected to be somewhat promiscuous, tolerating a few mismatches between a gRNA and a target DNA sequence to fend off rapidly evolving foreign genetic elements (125). Highly specific CRISPR nucleases that cannot tolerate a few mismatches should be easily circumvented by mutations to the phage genome or plasmid sequence. Yet, when introduced into human cells, Cas9 and Cpf1 nucleases are highly specific, often distinguishing a few nucleotide mismatches efficiently without inducing unwanted off-target mutations at sites with high sequence homology in the human genome, the size of which is ~1,000× larger than that of a typical microbial genome.

The high specificities of CRISPR nucleases, revealed by various genome-wide methods, make them promising tools for genome editing in eukaryotic cells. But how such highly specific nucleases remain functional as adaptive immune systems in prokaryotes is paradoxical. The success of the CRISPR-Cas defense systems appears to stem from two properties (126, 127): (*a*) multiple spacers acquired from a parasitic DNA element within the CRISPR loci dramatically reduce the likelihood that a phage or plasmid can accumulate multiple mutations necessary to escape recognition and degradation by CRISPR systems, and (*b*) spacer diversity within the prokaryotic population further increases the fitness of the population against invading foreign genetic elements. Notably, Cas9 systems likely operate at a higher cellular concentration in bacteria, because of both the smaller cell volume and the lower amount of competing genomic DNA. Consequently, CRISPR nucleases in bacterial systems should have a higher inherent chemical potential than in mammalian systems. Thus, in bacterial systems, the CRISPR nuclease may be more likely to occupy and cleave suboptimal target sites (near-cognate sequences within the host genome). Unlike eukaryotic cells, prokaryotic cells lack efficient NHEJ mechanisms to repair DSBs and cannot survive repeated enzymatic DSBs. Thus, CRISPR nucleases might have evolved high specificity to avoid lethal off-target DNA cleavage within the host genome.

GENETIC VARIATIONS AND PERSONAL OFF-TARGET SITES

Genetic variation in human populations can cause differential on-target or off-target activity among individuals. Large-scale genome projects have revealed genetic diversity in the form of

SNPs, InDels, and structural or copy number variations (SVs or CNVs) (128, 129). These variations can change the sequence of the target site and the landscape of off-target sequences present in any given individual. In fact, Yang et al. (130) and Tsai et al. (79) experimentally confirmed the presence of cell line–specific off-target sites in induced pluripotent stem cells and human cell lines, respectively. Lessard et al. (131) used 7,444 whole-genome sequences to examine the effect of SNPs and InDels on the genome-wide specificity of ~3,000 sgRNAs targeted to 30 loci of therapeutic relevance in the human genome. They found that genetic variation could produce changes in the target site sequence, which was predicted to disrupt cleavage activity as well as creating and destroying potential off-target sites for each guide within the genome. Importantly, the frequency at which genetic variation dramatically impacted predicted on-target or off-target activity was low, but their analysis highlights the importance of considering genetic variation in nuclease target site design destined for therapeutic applications to minimize risk of treatment failure and adverse side effects.

Homology-directed repair (HDR): DNA repair mechanism that utilizes a DNA template (either endogenous or exogenously supplied) with sequence homology to repair a double-strand break

DO OFF-TARGET EFFECTS MATTER IN THE CLINIC?

No small-molecule drugs or antibodies are free from off-target effects. A single chemical or a monoclonal antibody can potentially bind and inhibit many proteins or antigens with a similar structure. Identifying all of these off-target interactions is impossible with currently available tools. The ability to define and understand these off-target interactions would allow the creation of modified drugs that avoid side effects. Unlike other types of drugs or biologics, we now have efficient methods to enable the identification and quantification of mutation frequencies for the majority of likely off-target sites within the genome for gene-editing nucleases and deaminases. This knowledge will allow researchers to tailor these systems to minimize collateral damage to the genome. However, any collateral damage by gene-editing systems, unlike most off-target effects by small molecules and antibodies, is permanent and irreversible.

One key question facing the field for therapeutic genome editing is how important is a minor amount of collateral damage to the genome? The degree of concern for the off-target effects of programmable nucleases will depend on the target cell or tissue that is being modified. Minor collateral damage to the genome of highly differentiated or postmitotic cells is unlikely to have oncogenic consequences. For example, the US Food and Drug Administration has approved Phase I/II clinical tests using a *CCR5*-targeted ZFN developed by Sangamo Therapeutics, Inc., for the treatment of HIV infection through the *ex vivo* modification of patient T cells. To date, more than 200 patients with HIV infection have been treated. This nuclease has been proven safe so far because of the targeted cell type, although it cleaves many off-target sites in the human genome (68, 132), including a site in the highly homologous *CCR2* gene, potentially causing segmental deletions, duplications, and inversions (22, 133). Thus, it is the biological consequences of mutations that matter but not the mutations per se in the clinic. In this regard, etoposide, a small-molecule inhibitor of topoisomerase, is of note. Inhibition of topoisomerase by etoposide gives rise to random chromosomal DSBs, causing mutations across the genome. Because cancer cells often lack efficient DSB repair systems and, thereby, are vulnerable to DSBs, etoposide is used widely as an anticancer drug. In contrast, programmable nucleases and BEs operate with precision, cleaving DNA in a targeted manner and inducing genetic modifications site specifically.

Cas9 has been shown to be an efficient genome-editing tool in several systems, but the creation of DSBs within the genome results in the activation of DNA damage response pathways that may negatively impact the growth of primary cells. While attempting to understand homology-directed repair (HDR) and editing efficiency in challenging cell types—immortalized human retinal pigment epithelial cells (RPE1) and human pluripotent stem cells (hPSCs)—two groups from

Novartis and the Karolinska Institute demonstrated that DSBs induced by Cas9 lead to p53-dependent toxicity and cell cycle arrest (134, 135). This DSB-associated toxicity was even observed with the delivery of Cas9 RNPs, which have a limited lifetime in cells. Inhibition of p53 resulted in dramatically increased HDR efficiency and a decrease in damage response, providing a potential pathway to enhance HDR rates in primary cell types. Importantly, these results also imply that clones subjected to genome editing, and then isolated on the basis of the desired gene conversion, are under selective pressure to lose p53 function. If these cells are destined for therapeutic application, they may have an enhanced potential to become tumorigenic. However, this is likely to be a concern only for therapeutic applications in which a high degree of cellular proliferation following genome modification is required to generate the cell population of interest (e.g., induced pluripotent stem cell–based therapeutics). Nonetheless, future work on nuclease-based therapeutics will need to examine the impact of genome modification on DNA damage response pathways to more fully understand the safety implications of these approaches.

CONCLUSIONS AND PROSPECTS

Therapeutic genome editing with programmable nucleases and deaminases posits that the entire human genome is a potential drug target. Unlike monoclonal antibodies or small-molecule drugs, which are limited to extracellular or druggable targets such as certain enzymes or receptors, gene-editing enzymes can be used for genetic modifications anywhere within the genome, including traditionally undruggable protein-coding genes, noncoding RNA genes, regulatory elements, and so forth. Programmable nucleases and BEs are now so sufficiently mature that the generation of directed edits for therapeutic purposes within the genome of a desired tissue is now primarily limited by the challenge of their delivery, as opposed to the efficacy of these tools. Furthermore, programmable nucleases and deaminases will enable germline genome editing for the treatment of fatal genetic diseases—with all of the ethical caveats that are associated with directed germline modification. However, utilizing the human genome as a drug target presents new challenges and obstacles.

One of the key issues in therapeutic genome editing has been the genome-wide specificity of gene-editing tools. As described in this review, many laboratories have contributed to the development of methods for profiling genome-wide CRISPR off-target sites in cells or *in vitro* and for reducing or avoiding CRISPR off-target activity. Importantly, some nucleases such as Cpf1 systems or engineered SpCas9 variants have been shown to be remarkably specific, cleaving no off-target sites in the entire human genome. Yet, the sensitivity of standard methods of off-target detection within a population of nuclease-treated cells is limited by the high-throughput sequencing error rate, typically in the range of 0.01–1%: Off-target mutations that are induced below this limit cannot be easily detected. Note that millions or billions of cells are edited in a typical therapeutic genome-editing approach performed *ex vivo* or *in vivo*. Thus, the current sensitivity of off-target rate measurement is insufficient to identify off-target sites with extremely low mutation frequencies that are relevant to these therapeutic methods, so further improvement in off-target detection methods is warranted.

Another challenge is how to estimate (and ultimately avoid) the risk of oncogenesis caused by off-target activity. To minimize the risk, any off-target activity in tumor-suppressor genes within the stem and progenitor cell population should be avoided. It will be necessary to evaluate potential off-target sites broadly—not only within protein-coding exons but also at introns, because intronic mutations can lead to gene-disrupting alternative splicing. Off-target mutations in the vicinity of tumor-suppressor genes may also inactivate regulatory elements, suppressing expression of the genes. Off-target activity in proto-oncogenes is less of a concern because oncogenic

mutations are typically confined to specific sites in the genes. Nevertheless, off-target mutations creating novel regulatory elements near the protein-coding region of proto-oncogenes must be carefully monitored to avoid overexpression of the genes. The improvement of computer algorithms that can anticipate nuclease off-target activity in the vicinity of tumor-suppressor genes and proto-oncogenes will be helpful to avoid any unwanted, potential oncogenic mutations in the genome. In addition, the generation of unwanted breaks within the genome can lead to rearrangements that can generate aberrant transcripts and genome instability. All of these concerns can be alleviated by the continued development of genome-editing tools with improved specificity as well as off-target cleavage and mutagenesis detection methods with higher sensitivity. The realization of these technologies, in conjunction with whole genome sequencing data for each individual patient, will allow personalized genetic treatments for a multitude of monogenic and complex genetic or nongenetic disorders, which will revolutionize human health care.

DISCLOSURE STATEMENT

The authors have filed patent applications related to genome engineering technologies. D.K. and J.-S.K. have filed patent applications on Digenome-seq. J.-S.K. is a cofounder and stockholder of ToolGen, Inc., which is focused on genome editing. D.K. was supported by grants from ToolGen, Inc. (0409-20160107). S.A.W. is a consultant for Beam Therapeutics.

ACKNOWLEDGMENTS

This work was supported by grants from the US National Institutes of Health (R01AI117839, 1R01GM115911, and R01HL093766 to S.A.W.) and the Institute for Basic Science (IBS-R021-D1 to J.-S.K.).

LITERATURE CITED

1. Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *PNAS* 93:1156–60
2. Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, et al. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435:646–51
3. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, et al. 2011. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29:143–48
4. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–21
5. Cong L, Ran FA, Cox D, Lin S, Barretto R, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–23
6. Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31:230–32
7. Mali P, Yang L, Esvelt KM, Aach J, Guell M, et al. 2013. RNA-guided human genome engineering via Cas9. *Science* 339:823–26
8. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, et al. 2015. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:759–71
9. Bak RO, Gomez-Ospina N, Porteus MH. 2018. Gene editing on center stage. *Trends Genet.* 34:600–11
10. Urnov FD. 2018. Ctrl-Alt-inDel: genome editing to reprogram a cell in the clinic. *Curr. Opin. Genet. Dev.* 52:48–56
11. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533:420–24

12. Gehrke JM, Cervantes O, Clement MK, Wu Y, Zeng J, et al. 2018. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* 36:977–82
13. Wang X, Li J, Wang Y, Yang B, Wei J, et al. 2018. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* 36:946–49
14. Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, et al. 2016. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:aaf8729
15. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, et al. 2017. Programmable base editing of A* T to G* C in genomic DNA without DNA cleavage. *Nature* 551:464–71
16. Ryu SM, Koo T, Kim K, Lim K, Baek G, et al. 2018. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* 36:536–39
17. Zafra MP, Schatoff EM, Katti A, Foronda M, Breinig M, et al. 2018. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* 36:888–93
18. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, et al. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33:187–97
19. Frock RL, Hu J, Meyers RM, Ho YJ, Kii E, Alt FW. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* 33:179–86
20. Maddalo D, Manchado E, Concepcion CP, Bonetti C, Vidigal JA, et al. 2014. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 516:423–27
21. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, et al. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 24:132–41
22. Lee HJ, Kweon J, Kim E, Kim S, Kim JS. 2012. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* 22:539–48
23. Brunet E, Simsek D, Tomishima M, DeKelver R, Choi VM, et al. 2009. Chromosomal translocations induced at specified loci in human stem cells. *PNAS* 106:10620–25
24. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, et al. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33:187–97
25. Kim D, Kim S, Kim S, Park J, Kim JS. 2016. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* 26:406–15
26. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. 2016. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* 44:W272–76
27. Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. 2015. CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLOS ONE* 10:e0124633
28. Singh R, Kuscu C, Quinlan A, Qi Y, Adli M. 2015. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res.* 43:e118
29. Zhu LJ, Holmes BR, Aronin N, Brodsky MH. 2014. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLOS ONE* 9:e108424
30. Xiao A, Cheng Z, Kong L, Zhu Z, Lin S, et al. 2014. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* 30:1180–82
31. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* 42:W401–7
32. Heigwer F, Kerr G, Boutros M. 2014. E-CRISP: fast CRISPR target site identification. *Nat. Methods* 11:122–23
33. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. 2014. COSMID: a web-based tool for identifying and validating CRISPR/Cas off-target sites. *Mol. Ther. Nucleic Acids* 3:e214
34. Bae S, Park J, Kim JS. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30:1473–75
35. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31:827–32
36. Aach J, Mali P, Church GM. 2014. CasFinder: flexible algorithm for identifying specific Cas9 targets in genomes. bioRxiv. **https://doi.org/10.1101/005074**
37. Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343:80–84

38. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. 2015. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* 33:661–67
39. Chari R, Mali P, Moosburner M, Church GM. 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* 12:823–26
40. Kim D, Kim J-S. 2018. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Res*. 28:1894–1900
41. Chen F, Ding X, Feng Y, Seebeck T, Jiang Y, Davis GD. 2017. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.* 8:14958
42. Cornu TI, Thibodeau-Beganny S, Guhl E, Alwin S, Eichtinger M, et al. 2008. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther.* 16:352–58
43. Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res*. 19:1279–88
44. Morgens DW, Wainberg M, Boyle EA, Ursu O, Araya CL, et al. 2017. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun*. 8:15178
45. Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang CZ, et al. 2016. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov*. 6:914–29
46. Bisaria N, Jarmoskaite I, Herschlag D. 2017. Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell Syst*. 4:21–29
47. Yang M, Peng S, Sun R, Lin J, Wang N, Chen C. 2018. The conformational dynamics of Cas9 governing DNA cleavage are revealed by single-molecule FRET. *Cell Rep*. 22:372–82
48. Dagdas YS, Chen JS, Sternberg SH, Doudna JA, Yildiz A. 2017. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* 3:eaao0027
49. Lim Y, Bak SY, Sung K, Jeong E, Lee SH, et al. 2016. Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nat. Commun.* 7:13350
50. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. 2015. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527:110–13
51. Cameron P, Fuller CK, Donohoue PD, Jones BN, Thompson MS, et al. 2017. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods* 14:600–6
52. Ramakrishna S, Kwaku Dad AB, Beloor J, Gopalappa R, Lee SK, Kim H. 2014. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res*. 24:1020–27
53. Kim S, Kim D, Cho SW, Kim J, Kim JS. 2014. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res*. 24:1012–19
54. Katzman S. 2016. *Operational Assessment of IT*. Boca Raton, FL: CRC Press
55. Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM. 2018. A survey of validation strategies for CRISPR-Cas9 editing. *Sci. Rep.* 8:888
56. Brinkman EK, Chen T, Amendola M, van Steensel B. 2014. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res*. 42:e168
57. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. 2010. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* 649:247–56
58. Brinkman EK, Kousholt AN, Harmsen T, Leemans C, Chen T, et al. 2018. Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res*. 46:e58
59. Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. 2016. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* 34:863–68
60. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34:184–91
61. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, et al. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 31:822–26
62. Boyle EA, Andreasson JOL, Chircus LM, Sternberg SH, Wu MJ, et al. 2017. High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *PNAS* 114:5461–66

63. Friedland AE, Baral R, Singhal P, Loveluck K, Shen S, et al. 2015. Characterization of *Staphylococcus aureus* Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol.* 16:257
64. Lee CM, Cradick TJ, Bao G. 2016. The *Neisseria meningitidis* CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol. Ther.* 24:645–54
65. Amrani N, Gao XD, Liu P, Edraki A, Mir A, et al. 2018. NmeCas9 is an intrinsically high-fidelity genome editing platform. *Genome Biol.* 19:214
66. Kim HK, Song M, Lee J, Menon AV, Jung S, et al. 2017. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods* 14:153–59
67. Kleinstiver BP, Tsai SQ, Prew MS, Nguyen NT, Welch MM, et al. 2016. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* 34:869–74
68. Gabriel R, Lombardo A, Arens A, Miller JC, Genovese P, et al. 2011. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 29:816–23
69. Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, et al. 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* 10:361–65
70. Yan WX, Mirzazadeh R, Garnerone S, Scott D, Schneider MW, et al. 2017. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* 8:15058
71. Kim D, Bae S, Park J, Kim E, Kim S, et al. 2015. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 12:237–43
72. Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, et al. 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
73. Smith C, Gore A, Yan W, Abalde-Atristain L, Li Z, et al. 2014. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* 15:12–13
74. Iyer V, Shen B, Zhang W, Hodgkins A, Keane T, et al. 2015. Off-target mutations are rare in Cas9-modified mice. *Nat. Methods* 12:479
75. Ma H, Marti-Gutierrez N, Park SW, Wu J, Lee Y, et al. 2017. Correction of a pathogenic gene mutation in human embryos. *Nature* 548:413–19
76. Wang X, Wang Y, Wu X, Wang J, Wang Y, et al. 2015. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* 33:175–78
77. Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, et al. 2011. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147:107–19
78. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, et al. 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* 520:186–91
79. Tsai SQ, Nguyen NT, Malagon-Lopez J, Topkar VV, Aryee MJ, Joung JK. 2017. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* 14:607–14
80. Kim D, Lim K, Kim ST, Yoon SH, Kim K, et al. 2017. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* 35:475–80
81. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 32:677–83
82. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, et al. 2014. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 32:670–76
83. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 32:279–84
84. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science* 351:84–88
85. Yin H, Song CQ, Suresh S, Kwan SY, Wu Q, et al. 2018. Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat. Chem. Biol.* 14:311–16
86. Ryan DE, Taussig D, Steinfeld I, Phadnis SM, Lunstad BD, et al. 2018. Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.* 46:792–803

87. Cromwell CR, Sung K, Park J, Krysler AR, Jovel J, et al. 2018. Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat. Commun.* 9:1448
88. Rees HA, Komor AC, Yeh WH, Caetano-Lopes J, Warman M, et al. 2017. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* 8:15790
89. Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, et al. 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* 33:73–80
90. Choi JG, Dang Y, Abraham S, Ma H, Zhang J, et al. 2016. Lentivirus pre-packed with Cas9 protein for safer gene editing. *Gene Ther.* 23:627–33
91. Lee K, Conboy M, Park HM, Jiang F, Kim HJ, et al. 2017. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* 1:889–901
92. Guilinger JP, Thompson DB, Liu DR. 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 32:577–82
93. Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, et al. 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 32:569–76
94. Wyvekens N, Topkar VV, Khayter C, Joung JK, Tsai SQ. 2015. Dimeric CRISPR RNA-guided FokI-dCas9 nucleases directed by truncated gRNAs for highly specific genome editing. *Hum. Gene Ther.* 26:425–31
95. Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS* 109:E2579–86
96. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, et al. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154:1380–89
97. Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, et al. 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 31:833–38
98. Lei L, Chen H, Xue W, Yang B, Hu B, et al. 2018. APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks. *Nat. Struct. Mol. Biol.* 25:45–52
99. Wright AV, Sternberg SH, Taylor DW, Staahl BT, Bardales JA, et al. 2015. Rational design of a split-Cas9 enzyme complex. *PNAS* 112:2984–89
100. Zetsche B, Volz SE, Zhang F. 2015. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* 33:139–42
101. Truong DJ, Kuhner K, Kuhn R, Werfel S, Engelhardt S, et al. 2015. Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* 43:6450–58
102. Chew WL, Tabebordbar M, Cheng JK, Mali P, Wu EY, et al. 2016. A multifunctional AAV–CRISPR–Cas9 and its host response. *Nat. Methods* 13:868–74
103. Bolukbasi MF, Gupta A, Oikemus S, Derr AG, Garber M, et al. 2015. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* 12:1150–56
104. Bolukbasi MF, Liu P, Luk K, Kwok SF, Gupta A, et al. 2018. Orthogonal Cas9-Cas9 chimeras provide a versatile platform for genome editing. *Nature Commun.* 9(1):4856
105. Davis KM, Pattanayak V, Thompson DB, Zuris JA, Liu DR. 2015. Small molecule–triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* 11:316–18
106. Anders C, Niewoehner O, Duerst A, Jinek M. 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513:569–73
107. Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, et al. 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351:867–71
108. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, et al. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156:935–49
109. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, et al. 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529:490–95
110. Lee JK, Jeong E, Lee J, Jung M, Shin E, et al. 2018. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* 9:3048

111. Kim S, Bae T, Hwang J, Kim JS. 2017. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5′ nucleotides. *Genome Biol*. 18:218
112. Zhang D, Zhang H, Li T, Chen K, Qiu JL, Gao C. 2017. Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol*. 18:191
113. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, et al. 2017. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550:407–10
114. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, et al. 2018. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* 36:265–71
115. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, et al. 2018. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556:57–63
116. Kim E, Koo T, Park SW, Kim D, Kim K, et al. 2017. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun*. 8:14500
117. Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, et al. 2016. Structure and engineering of *Francisella novicida* Cas9. *Cell* 164:950–61
118. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, et al. 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res*. 42:2577–90
119. Hou Z, Zhang Y, Propson NE, Howden SE, Chu LF, et al. 2013. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *PNAS* 110:15644–49
120. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* 10:1116–21
121. Ibraheim R, Song C, Amrani N, Xue W, Sontheimer EJ. 2018. All-in-one adeno-associated virus delivery and genome editing by *Neisseria meningitidis* Cas9 in vivo. *Genome Biol*. 19:137
122. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, et al. 2015. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* 13:722–36
123. Koo T, Park SW, Jo DH, Kim D, Kim JH, et al. 2018. CRISPR-LbCpf1 prevents choroidal neovascularization in a mouse model of age-related macular degeneration. *Nat. Commun.* 9:1855
124. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, et al. 2017. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* 35:31–34
125. Carroll D. 2014. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* 83:409–39
126. Levin BR, Moineau S, Bushman M, Barrangou R. 2013. The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLOS Genet*. 9:e1003312
127. van Houte S, Ekroth AK, Broniewski JM, Chabas H, Ashby B, et al. 2016. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532:385–88
128. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, et al. 2015. A global reference for human genetic variation. *Nature* 526:68–74
129. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285–91
130. Yang L, Grishin D, Wang G, Aach J, Zhang CZ, et al. 2014. Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat. Commun.* 5:5507
131. Lessard S, Francioli L, Alfoldi J, Tardif JC, Ellinor PT, et al. 2017. Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *PNAS* 114:E11257–66
132. Pattanayak V, Ramirez CL, Joung JK, Liu DR. 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* 8:765–70
133. Lee HJ, Kim E, Kim JS. 2010. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res*. 20:81–89
134. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. 2018. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24:927–30
135. Ihry RJ, Worringer KA, Salick MR, Frias E, Ho D, et al. 2018. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* 24:939–46