

Advances in engineering CRISPR-Cas9 as a molecular Swiss Army knife

Grace A. Meaker^{1,2}, Emma J. Hair¹ and Thomas E. Gorochowski^{1,3,*}

¹ School of Biological Sciences, University of Bristol, Tyndall Avenue, Bristol BS8 1TQ, UK

² School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3AT, UK

³ BrisSynBio, University of Bristol, Tyndall Avenue, Bristol BS8 1TQ, UK

* Correspondence should be addressed to T.E.G. (thomas.gorochowski@bristol.ac.uk)

Keywords: synthetic biology; CRISPR; Cas9; biotechnology; biodesign; nickase; base editing; prime editing; genome editing; ethics; responsible innovation

Abstract

The RNA-guided endonuclease system CRISPR-Cas9 has been extensively modified since its discovery, allowing its capabilities to be extending far beyond double-stranded cleavage to high fidelity insertions, deletions, and single base edits. Such innovations have been possible due to the modular architecture of CRISPR-Cas9 and the robustness of its component parts to modifications and the fusion of new functional elements. Here, we review the broad toolkit of CRISPR-Cas9-based systems now available for diverse genome editing tasks. We provide an overview of their core molecular structure and mechanism and distil the design principles used to engineer their diverse functionalities. We end by looking beyond the biochemistry and towards the societal and ethical challenges that these CRISPR-Cas9 systems face if their transformative capabilities are to be deployed in a safe and acceptable manner.

Introduction

Defined originally as an array of DNA repeats in 1987¹, the exact function of the clustered regularly interspaced short palindromic repeats (CRISPR) remained a mystery until the further discovery of associated Cas proteins and RNA elements. This established their combined function as a prokaryotic immune system²⁻⁵, which had evolved to combat invading phages by cleaving and degrading their DNA. The core components are a Cas (CRISPR-associated) endonuclease, directed to a DNA target by a multi-component guide RNA (gRNA)^{6,7}, which has since been simplified into a single guide RNA (sgRNA)⁸.

The power of the CRISPR system comes from its highly programmable nature that allows it to be easily targeted to virtually any DNA locus by merely placing a complementary sequence within the gRNA. Whilst its built-in functionality has ushered in a new era of genome engineering, CRISPR's real merit lies in its robustness for significant modification. This has allowed the CRISPR system to be refined as well radically extended to broaden its capabilities. These developments have enabled CRISPR to be used for diverse applications covering gene regulation, large genomic insertions and deletions, accurate base editing, and precise sequence replacement⁹⁻¹³.

In this review, we explore the development of modified Cas9-based CRISPR systems for genome editing tasks, and the main approaches used to engineer these functionalities. This includes the mutagenesis of Cas9 domains, redesign of the gRNA, fusion of additional enzymatic domains to Cas9, and the screening of other organisms for naturally occurring CRISPR variants with more desirable features. Our aim is to provide a clear mechanistic overview of how the modular structure of the CRISPR-Cas9 system has facilitated engineering efforts and allowed for a 'plug-n-play' type approach to the development of new DNA-targeted functionalities. Whilst the potential benefits of such systems are already starting to be realized, we end by raising caution when considering their wider deployment and discuss some of the less widely acknowledged ethical and evolutionary challenges associated with this technology.

The native CRISPR-Cas9 system

The CRISPR-Cas9 system is a class II type II CRISPR system derived from *Streptococcus pyogenes*¹⁴. It consists of a Cas nuclease *SpCas9* and a gRNA⁸. The gRNA has two components – a trans-activating RNA (tracrRNA) and CRISPR RNA (crRNA)⁶. crRNA is responsible for recognition and binding of the target DNA region and tracrRNA for crRNA maturation and association with *SpCas9*. Once the gRNA binds the *SpCas9*, the *SpCas9* undergoes a conformational change which permits the *SpCas9*-crRNA-tracrRNA complex to relocate to the target region and cleave both DNA strands⁷. The target region is determined by a 20-nucleotide 'spacer' in the crRNA, complementary to the target 'protospacer' in the

DNA^{3,15}. For recognition, the protospacer must be superseded by several nucleotides called the protospacer adjacent motif (PAM). This varies for different Cas proteins; for *SpCas9* it is '5-NGG-3'^{8,16}. Providing there is the correct PAM present directly downstream of the target locus, engineering a gRNA with a different spacer region allows for targeting of a different genomic location.

When the target region is found, the bases upstream of the PAM are melted and bind to the complementary region of the gRNA^{17,18}. Once the complex is bound, the two nucleases produce a double-stranded break (DSB) 3–4 nucleotides (nt) upstream of the PAM¹⁹. The DSB induces the endogenous DNA repair machinery, commonly the non-homologous end-joining pathway (NHEJ). NHEJ is notoriously error-prone, so the break is often fixed incorrectly and the target sequence becomes mutated²⁰. Alternatively, the homology-directed repair pathway (HDR) can be used to fix the break using a homology template to accurately insert a desired sequence^{20,21}. Recognition of CRISPR's ability to perform gene knockdown/insertion was the beginning of a series of alterations which would highlight the diverse applications of this system and its derivatives.

Whilst CRISPR can perform efficient cleavage of a target genomic region, a common problem is the presence of non-target cleavage, or off-target effects, particularly in larger genomes²². The genomic target has 20 nt of complementarity to the spacer region of the gRNA, but mutations at the 5' end of the gRNA still permit efficient cleavage implying the first 12–13 nt are critical for specifying the target^{16,19,20}. These essential 13 nt of the spacer region have been dubbed the 'seed sequence'^{8,23}. Genomic regions with incomplete homology to the spacer region which contain all or most of the seed sequence could be targeted by the Cas9, resulting in off-target effects²⁴. Detection and prevention of this off-target activity is essential for CRISPR to be used as a therapeutic tool. Efforts utilising altered, higher-fidelity Cas9 proteins and truncated gRNA^{25–27} have been the focus of efforts to reduce such promiscuity and will be discussed later in this review.

To assist with the characterization of CRISPR, large-scale bioinformatic tools have been developed for genomic analysis and specifically the identification of potential editing sites. Complementary biological assays have also been developed to assess off-target cleavage²⁸. A widely used assay to investigate off-target binding is the T7 endonuclease 1 (T7E1) mismatch detection assay. Despite its widespread use, validations in the literature have exposed the poor accuracy and sensitivity of T7E1 assay²⁹. Cleavage by *SpCas9* has been observed at sites with up to 5 mismatches to the spacer region and even in sites without the 5'-NGG-3' PAM, for example at those containing 5'-NAG-3'^{30,31}.

Computational tools such as Cas-OFFinder and E-CRISP assume that sites with more homology to the spacer region are more likely to be targeted and vice versa, allowing the user to visualize potential off-target loci^{32,33}. These approaches, however, do not take into account

off-target sites which do not fit the model's parameters³⁴. Genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq) provides a more robust method for identifying off target effects, and has become widely used³⁵. A small oligo-nucleotide tag is integrated into DSB sites targeted by NHEJ, and sequencing analysis is used to pin-point the location of off-target sites. This permits the detection of sites neglected by other computational tools³². GUIDE-seq is a simple method to identify sites which have up to 6 mismatches to the protospacer sequence as well as noncanonical PAMs, giving a broad profile of off-target effects, but is limited by the use of an oligo tag^{34,35}. Another example of a genome-wide tool is digested genome sequencing (Digenome-seq) which involves the digestion of genomic DNA with Cas9-gRNA complexes and subsequent deep sequencing to identify identical Cas9 cleavage fragments³⁶. Analysis is performed on extracted DNA, eliminating the influence of cellular context (e.g. chromatin arrangements, methylation patterns and DNA accessibility). However, this method is time-consuming as many reads have to be analyzed to identify patterns, and fails to recognize identical fragments caused by chance³⁴. Overall, no single method is able to comprehensively analyse off-target effects and so the method employed must be carefully considered on a case-by-case basis. For example, in eukaryotic genomes, Digenome-seq is appropriate because it is not vulnerable to chromatin arrangements³⁶, but for large genomes GUIDE-seq is easier to use and more sensitive³⁵. For a truly comprehensive understanding of all off-target effects, a multi-system analysis involving both computational and biological approaches is necessary, but rarely performed.

Structure of *SpCas9*

SpCas9 is a multi-domain protein exhibiting a bilobed structure where the nuclease lobe and the recognition lobe^{8,19} are linked by an arginine-rich bridge helix as well as a disordered linker⁸ (**Figure 1A**). The overall shape of *SpCas9* is oblong with two large grooves, to accommodate the DNA:RNA and RNA:RNA complexes. Adaptations of the two previously-recognized, adjacent nucleases (HNH⁶, named for the three characteristic residues, and RuvC³⁷) of the nuclease lobe facilitate much of the diversification of CRISPR's function^{25,38}. Each nuclease cleaves one strand of DNA; RuvC cleaves the non-complementary and HNH the complementary strand^{6,15}. Another key component of the nuclease lobe is the C-terminal domain, with a region essential for PAM recognition and binding often called the PAM-interacting (PI) domain⁷. Mutagenesis of these domains permits the evolution of CRISPR function.

Naturally occurring variants

CRISPR is a naturally occurring system in prokaryotes, thus different species possess different systems whose variations can be potentially exploited³⁹. Type I and III systems enlist

multiple Cas proteins whereas type II uses a single, Cas9 protein for DNA cleavage⁴⁰. Whilst *SpCas9* from *S. pyogenes* is the most heavily studied to date, Cas9 variants from different bacteria with distinct cleavage patterns and PAM requirements are becoming more widely used, such as *FnCas9* from *Francisella novicida*⁴¹, *SaCas9* from *Staphylococcus aureus*^{42,43} and recently the *Campylobacter jejuni* Cas9, the smallest to date^{44,45}.

SaCas9 has a longer PAM (5'-NNGRRT-3') than *SpCas9* and is smaller at 1053 amino acids (aa) compared to 1368 aa⁴² (**Figure 1B**). Because of its smaller size, *SaCas9* provides valuable information regarding the elements of Cas9 that are essential and those that can be removed or modified without impacting overall function. Characterization of *SaCas9* has shown comparable on-target cleavage to *SpCas9*, whilst boasting a higher specificity and easier introduction into cells⁴⁶. Both *SpCas9* and *SaCas9* are bilobed, with a nuclease (NUC) and recognition (REC) lobe linked by an arginine bridge and a linker region. They both contain two nuclease domains, HNH and RuvC and undergo a conformational change when gRNA is bound. However, *SaCas9* only has 17% structural similarity to *SpCas9*; key DNA/RNA binding domains such as the nucleases and PI domain have been conserved but others such as the REC2 domain are not, suggesting its presence is not crucial for Cas9 function. This demonstrates the flexibility of Cas9's structure whilst retaining efficacy⁴⁶. Despite these differences, it is apparent that *SaCas9* and *SpCas9* share important similarities, and that *SaCas9* is a useful case study for synthetic reduction of *SpCas9* size and complexity, already attempted by the successful removal of the REC2 domain⁴⁷.

Another *SpCas9* ortholog is *FnCas9* which produces staggered cleavage and binds less frequently to non-target regions^{41,48} (**Figure 1C**). The non-target strand is cleaved 3–8 bp upstream of the PAM (5'-NGG-3'), whereas the target strand is cleaved 3 bp upstream as by *SpCas9* and *SaCas9*, producing overhangs of up to 4 nt and more efficient recruitment of HDR⁴¹. *FnCas9* is considerably larger than *SpCas9* and *SaCas9*, comprised of 1629 aa⁴⁹. Whilst its larger size may be a hinderance for transfection, *FnCas9*'s markedly reduced tolerance of target mismatches makes it a valuable system for precise editing tasks. *SpCas9* tolerates several mismatches of the gRNA in the non-seed region, but just one mismatch at the 5' end of *FnCas9* gRNA is tolerated for successful cleavage⁴⁸. This increased specificity means *FnCas9* produces far less off-target cleavage as fewer sites are recognized as 'target'⁴¹. *FnCas9* is structurally dissimilar to *SpCas9* and *SaCas9*, lacking bilobed structure and containing distinct REC2 and REC3 domains (**Figure 1C**). REC3 domain mutations have generated high-fidelity Cas9 enzymes⁵⁰; these structural differences explain the striking differences in targeting specificity. Despite its increased specificity, it has much lower on-target recognition than *SpCas9* in eukaryotic genomes. As postulated in the literature⁴⁸, local chromatin conformations likely affect the access to DNA, a vulnerability not shared by *SpCas9*. To eliminate this problem *FnCas9* has been used alongside a catalytically dead *SpCas9*

(*SpdCas9*) to enable access and subsequent DNA cleavage⁴⁸. Such problems are not present when used in prokaryotes where *FnCas9* has been shown to function effectively⁵¹.

Finally, *CjCas9* is the smallest ortholog characterized to date at only 984 aa, which makes it suitable for size-restricted delivery methods such as those using adeno-associated viruses (AAV) (**Figure 1D**). It has a bilobed structure, as in *SaCas9* and *SpCas9*, with a simplified REC lobe and size-reduced NUC lobe⁴⁵ (**Figure 1D**). Initial studies showed recognition of a 5'-NNNNACA-3' PAM³⁹ or the more promiscuous 5'-NNNRYM-3'⁴⁵ providing an assortment of target sites. However, recent studies have found a requirement for an 8th C at the 3' end, suggesting 5'-NNNNRYAC-3'⁴⁴ and 5'-NNNNACAC-3' sequences⁵². Tested against *SaCas9* in human cells, *CjCas9* was found to be more specific with comparable efficiencies to other variants⁴⁴. However, due to discrepancies in the PAM recognition sequences and limited research into the structure and mechanism of *CjCas9*, care should be taken when placing confidence in this finding.

Comparisons of each Cas9 ortholog and their respective sgRNA has also revealed several structural and functional differences (**Figure 1**). The essential region of the sgRNA consists of a DNA binding region, the repeat:antirepeat duplex (R:AR) and at least 2 stem loops. Removal of stem loop 1, which has extensive interactions with Cas9, prevents cleavage, so its presence is essential^{6,42}. In contrast, removal of loops 2 or 3 decreases efficiency, without abolishing cleavage¹⁹. Stem loop 2 interacts with the PI and RuvC domains in *SaCas9* and *SpCas9*, and the REC domains in *FnCas9* and *CjCas9*^{7,42,45,47,49}. *SaCas9* and *SpCas9*'s sgRNAs exhibit the greatest similarity, particularly regarding cognate Cas9 interactions with the key difference being the lack of stem loop 3 in *SaCas9*⁴². This further highlights the simplicity of *SaCas9* compared to *SpCas9* because of the reduction of non-essential elements. *FnCas9* and *CjCas9*'s sgRNAs are structurally distinct to *SaCas9* and *SpCas9*, with the same core region but some unique features. For instance, *FnCas9* has a longer, U shaped linker, contrasting with the shorter, single-stranded linker present in *SaCas9* and *SpCas9*⁴⁹. The novel structural arrangement of *CjCas9*'s gRNA forms a triple helix between stem loops 1, 2, and 3⁴⁵. The relevance of this structure is unfortunately still unknown due to a lack of comprehensive structural studies of *CjCas9* complexes.

The domains of each Cas9 interact with their associated sgRNAs in a distinct way due to the slight differences in sgRNA structure⁴² (**Figure 1**). The stark differences between *SpCas9* and its orthologs demonstrate the diversity of naturally occurring Cas9 systems and their varying characteristics. Whilst the four orthologs discussed here have been characterized and established as potential genome-editing tools, their testing still pales in comparison to *SpCas9* and we expect that further characterisation experiments will be needed before their deployment. Even so, the differences in mechanism and function seen across these variants

clearly highlight the wealth of preexisting systems available that may be suitable for many applications.

Modifying CRISPR-Cas9 to enhance performance

Modification of guide RNAs

The CRISPR-Cas9 system requires a tracrRNA and a crRNA for target complementarity and complex maturation. To simplify use, a single chimeric guide RNA (sgRNA) is generally used to describe the dual-tracrRNA:crRNA structure (**Figure 1**, bottom row). As established by Jinek and colleagues, a seed region (13 nt of complementarity between the crRNA and the 3' end of the protospacer sequence) and a GG dinucleotide at the 3' end of the PAM are essential for sequence-specific recognition and cleavage⁶. By fusing the 3' end of the crRNA to the 5' end of tracrRNA this study simulated the tracrRNA:crRNA duplex formed in nature, inducing a Cas9 open conformation and directed DNA targeting. In this study, the chimeric gRNA produced cleaved all 5 expected targets *in vitro* and has since been widely used, confirming its efficacy⁶. Such mimicking of nature's gRNA design is a great example of how simple biotechnological approaches can yield more streamlined genetic engineering systems.

Another modification involves truncating the gRNA such that it contains <20 nt of complementarity to a target locus. Truncated gRNAs or tru-gRNAs have been shown to have significantly lower off-target activity compared to full-length sgRNAs due to a reduction in binding affinity and greater mismatch intolerance^{33,53}. As demonstrated in two human cell lines, the specificity of tru-gRNAs as compared to wild-type was estimated to be >10,000-fold higher²⁷. Such estimates are supported by the finding that additional nucleotides added at the 5' end of gRNA increases binding affinity for off-target sites²². Using the same study systems, it has been shown that positive synergism between tru-gRNAs and paired Cas9 nickases permits a further reduction in off-target activity, demonstrating the promise of the additive effects when combining modifications.

Modification of Cas9

Another way to improve performance is through modification of the Cas9 enzyme itself (**Figure 2**). Analysis of CRISPR-Cas9 variants and their resultant cleavage products established RuvC and HNH nuclease-mediated cleavage of the non-complementary and complementary strand respectively^{6,15}. As double-stranded cleavage initiates the inaccurate NHEJ pathway, single-stranded cleavage (or 'nicking') is favorable for efficient genome editing. A deactivating mutation in the catalytic residues of one of the nucleases causes the Cas9 to cleave only one strand of the target DNA. Such nicking permits accurate HDR or base excision repair (BER)^{54,55}. Two nicking variants (henceforth nickases) were engineered by an aspartate to

alanine substitution in the active site of the RuvC domain to produce Cas9D10A and histidine to alanine substitution in the HNH domain to produce Cas9H840A^{15,20,25}. The benefits of these are twofold: they produce precise nicks in the DNA and exhibit decreased affinity to off-target loci²⁵. When a DSB is required, a nickase can be used with two different gRNAs that target each strand of the DNA. When both nicks are performed a staggered cleavage site is produced (**Figure 3**)⁵⁶. This dual nicking strategy has been shown to have comparable on-target cleavage to *SpCas9* whilst discriminating off-target sites more effectively⁵⁷. Continued editing of nickases forms the basis of many other CRISPR editing systems that will be explored in the next section.

As a mutation in one of the nuclease domains can alter Cas9 from a dsDNA endonuclease to a ssDNA nickase, mutation of both domains can remove all cleavage activity. An *SpCas9* enzyme containing the H840A and D10A mutations is catalytically dead (dCas9)^{6,58}, but is still able to target and bind DNA. dCas9 has been shown to be a versatile tool and can be tethered to other molecules such as other enzymes⁹ or used to visualize target affinity without cleavage³⁸. Furthermore, dCas9 has become widely used in regulating gene expression through CRISPR interference and activation (CRISPRi and CRISPRa, respectively)^{59,60} and become a valuable tool for knock-down screens where Cas9 is not suitable⁶¹.

An additional application of dCas9 is to fuse it to a FokI nuclease, an endonuclease which is strictly dependent on dimerization for cleavage activity⁶⁰. This fusion enlists a long, flexible linker with between 5-25 residues (eg. GGGGS)₅ fusing the FokI endonuclease to the Cas9 N-terminus^{62–64}. The RNA-guided FokI Nuclease (RFN) system consists of a dCas9-FokI fusion and two different gRNAs⁶⁵. These gRNAs must have specificity to the target region, and both must be bound to their respective loci to allow for a functional FokI dimer to form and cleavage to take place. When there is off-target binding by one gRNA:Cas9 complex, the FokI monomer remains inactive and cleavage does not occur⁶² (**Figure 4**). The use of these alternative, exogenous nucleases creates a highly specific system with significantly lower indel frequencies when compared to wild-type Cas9 nucleases and the use of single gRNAs⁶⁴. However, RFNs are limited for genome-wide application due to the required presence of PAM sequences either side of the protospacer regions (5'-CCNN₂₀-3' and 5'-N₂₀NGG-3') as well as 14–17 bp between these⁶³. This fusion system is also very large, limiting its application in AAV delivery methods⁶⁶. Efforts have been made to use the smaller SaCas9 based system instead of *SpCas9*, reducing the size and simplifying delivery⁶³.

Mutation of REC3 domain

Targeted mutagenesis of other Cas9 domains has also been performed to find additional useful modifications. For example, as DNA binds between the HNH and REC domains,

mutations of the positively charged residues of REC3 to alanine could reduce binding affinity making the Cas9 more strongly discriminate between target and off-target regions⁶⁷. Using this knowledge, a high fidelity Cas9, *SpCas9*-HF1 was produced via mutation of four DNA-interacting REC3 residues to alanine (N497A/R661A/Q695A/Q926A), with comparable on-target cleavage to *SpCas9*²⁶. Despite the reduction in off-target mutations as quantified by GUIDE-seq, this variant was incompatible with the optimized truncated gRNA demonstrating a case where independent enhancements could not be combined.

A failure to completely abolish off-target activity in *SpCas9*-HF1 led to further screening of REC3 mutants *in vivo* and the development of another highly-specific *SpCas9* variant, dubbed 'evoCas9'⁵⁰. This variant outperforms *SpCas9*-HF1 in distinguishing between on and off-target sites and has better compatibility with optimized gRNAs.

Directed evolution for altered PAM specificity

Alterations to the nuclease and recognition domains have been shown to improve target specificity and efficiency. However, *SpCas9* is still limited to targeting of genomic regions containing the 5'-NGG-3' PAM⁶, whose number may be further reduced by local chromatin or methylation patterns preventing Cas9 access to the site²⁰. PAM specificity is conferred by several residues of the PI domain, specifically *SpCas9* arginine residues 1333 and 1335 which interact with the two guanine nucleotides of the PAM⁷. Motivated by this, several studies have focused on mutagenizing this domain to change the PAM recognized by Cas9. An attempt in 2014 substituted the two critical guanine-recognizing residues which interact with adenine from arginine to glutamine in an attempt to modify *SpCas9* recognition to a 5'-NAA-3' PAM⁶⁸. This effort was unsuccessful and the R1333Q/R1335Q variant produced failed to cleave DNA *in vitro*. It was concluded that additional mutations were likely required for successful alteration of PAM recognition.

Building on this work, Nishimasu and colleagues employed a positive selection approach where survival of bacteria was only guaranteed by Cas9 cleavage of a toxic gene⁴³. This produced two main variants: VQR (D1135V/R1335Q/T1337R) which recognized 5'-NGAN-3' and 5'-NGCG-3' PAMs and VRER (D1135V/G1218R/R1335E/T1337R) which recognized the 5'-NGCG-3' PAM. The T1337R mutation was found to be a gain of function, contrasting with the loss of function mutations utilized by other domain mutagenesis studies. This specific gain of function permitted Cas9 recognition of a fourth PAM base which increased the stringency of binding and reduced off target effects compared to wild-type *SpCas9*⁴³. These evolved *SpCas9* variants with altered PAM specificities are still limited to one or two PAMs.

To expand PAM recognition, focus has shifted to generating *SpCas9* variants able to target multiple PAMs. So far, positive selection has been used to find useful mutagenized

SpCas9 variants using phage assisted continuous evolution (PACE)¹⁶. Such variants, dubbed 'xCas9' nucleases, had a different pattern of mutations than the rationally developed variants which covered the entire *cas9* gene^{7,43}. xCas9-3.7 showed the best cleavage efficiency, with a high indel formation of DNA adjacent to 5'-NG-3', 5'-GAA-3' and 5'-GAT-3' PAMs as well as comparable activity to 5'-NGG-3' with *SpCas9*¹⁶. Together with the broader on-target specificity, xCas9-3.7 produced less off-target cleavage than *SpCas9*, demonstrating the potential merits of using an engineered Cas9 rather than the native system.

Mutation of the PI domain in this way is not limited to *SpCas9* and has been performed in *SaCas9* to similar effect. Using an analogous bacterial selection approach, mutated *SaCas9* variants were tested for their efficiency for 5'-NNNRRT-3' PAM loci cleavage. Results showed that an E782K/N968K/R1015H variant called SaKKH was functional and that this variant disrupted 5'-NNGRRT-3' sites (and off-target loci) at a similar efficiency to wild-type *SaCas9* whilst also cleaving sites adjacent to 5'-NNARRT-3', 5'-NNTRRT-3' and 5'-NNCRRT-3'⁶⁹.

Plug-n-play CRISPR-Cas9 modules

Base editing

NHEJ-based methods are useful for the downregulation or knock-out of genes, but for more precise editing the less error-prone HDR is preferential. HDR has been shown to work alongside the CRISPR system and in theory can induce a range of genome edits, but is hard to employ *in vivo*⁷⁰. Additionally, both of these DNA repair pathways rely on the generation of DSBs, which can result in inadvertent genomic alterations, pathogenic lesions and deleterious p53 activation responses⁷¹. Single stranded nicks are repaired by the high-fidelity BER pathway, making this cleavage pattern preferable for specific base changes⁵⁵.

Studies of the mechanism of Cas9 cleavage have revealed that the displaced DNA strand is unbound, this finding coupled with the need to more accurately alter genetic sequences led to the development of base editors⁷² (**Figure 5**). A simple CRISPR base editor consists of a dCas9 protein, a sgRNA and a base editing enzyme (e.g. cytidine deaminase)⁷³. Cytidine deaminases catalyze the conversion of cytosine to uracil⁷⁴ and the rat cytidine deaminase (rAPOBEC1) has been selected in a number of systems due to its high activity. To localize rAPOBEC1 to a target site in DNA and create the first base editor (BE1), rAPOBEC1 was fused to dCas9 via an XTEN linker which is commonly used in FokI-dCas9 fusions^{64,75} (**Figure 5A**). BE1 is able to deaminate 5 bases at the 5' end of the protospacer and was found to have a 50–80% efficiency *in vitro*, but only 0.8–7.7% *in vivo* (human cells)⁵⁸. This discrepancy was attributed to the endogenous DNA repair machinery, specifically uracil DNA glycosylase (UDG), which reverses the UG pair to a CG pair⁵⁸. To combat this, a uracil DNA glycosylase inhibitor (UGI) was attached to the C-terminus of BE1, to create the second base

editor variant BE2 (**Figure 5B**). This alteration increased editing efficiencies in human cells 3-fold as UDG activity was drastically reduced⁵⁸. Both these editors are only active on the strand containing the cytosine so to broaden the editors' function dCas9 was modified to create variant BE3 that acted as a nickase targeting the non-edited strand (**Figure 5C**). BE3 was 2 to 6-fold more efficient in creating cytosine to thymine transitions than BE2. All 3 editors showed off target-binding, but no base editing was found to have occurred at these sites and indel formation was significantly less than that induced by Cas9-mediated DSBs. A further development produced an additional base editor variant BE4 which included three alterations to BE3 (**Figure 5D**). The linkers fusing the rAPOBEC1 and UGI proteins to Cas9 were extended to 32 and 9 aa, respectively, and an additional UGI was added to the C-terminus with a 9 aa linker⁷⁶. BE4 showed higher C to T editing efficiency as well as better product yield compared to BE3. The evolution of this base editor system highlights the robust nature of the Cas9 protein to the 'plug-n-play' for additional functional modules in a rational way.

Another study which used this combined approach employed a SaCas9 nickase instead of SpCas9 in a BE3 variant, SaBE3⁷⁷. As previously described, SaCas9 is much smaller than SpCas9⁴² and recognizes a 5'-NNGRRT-3' PAM. The creation of a base editing system with this different nickase allowed for targeting of not only 5'-NGG-3' but also 5'-NNGRRT-3' PAMs, increasing the number of potential editing sites. SaBE3 also possesses other benefits, such as an increased editing efficiency on target as well as base editing outside of the expected activity window compared to the SpCas9-based BE3^{58,77}. Furthermore, Kim and colleagues utilized SpCas9 variants with altered PAM specificities, specifically VQR and VRER (described previously) and EQR from the same study⁴³, as well as an engineered SaCas9 variant, SaKKH⁶⁹. All these variants had editing efficiencies of up to 50% for sites with relevant PAMs, with SaKKH-BE3 editing up to 62% of target sites. SaBE3 and SaKKH-BE3 had a similar off-target activity to SpCas9 whereas EQR-BE3 and VQR-BE3 showed markedly reduced levels⁷⁷. These data again highlight the merits of combining CRISPR-Cas9 modifications to extend functionalities.

Prime editing

A similar combinatorial approach was used to create another form of more complex editing machinery. So called, prime editing combines the functionalities of a Cas9 nickase, reverse transcriptase (RT) and unique prime editing gRNA (pegRNA) (**Figure 6**). By combining these elements more precise changes to DNA can be made that go beyond the capabilities of other base editors (e.g. transversion point mutations, insertions, deletions)¹¹. The pegRNA is novel, as it both guides the Cas9-gRNA complex to the target and provides the sequence substrate for the RT to rewrite into the genome. The first prime editor PE1 consisted of a wild-type M-MLV RT attached to the C-terminus of H840A nickase (**Figure 6A**). PE1 was able to generate

transversion mutations at efficiencies of up to 5.5% and insertions and deletions of up to 17%¹¹. To increase the efficiency of PE1, a second prime editor variant PE2 was produced by incorporating five RT mutations designed to enhance binding affinity (**Figure 6B**). PE2 had increased efficiency of insertions and deletions and up to 5.1-fold increases in efficiency of targeted point mutations as compared to PE1. The further prime editor PE3 used the PE2 protein machinery alongside an additional sgRNA targeting the non-edited strand (**Figure 6C**). This simple modification increased editing efficiency by 1.5–4.2-fold, which is thought to be due to the edited strand acting as a template for non-edited strand repair¹¹.

Challenges

Inconsistent off-target detection methods

Precise detection of off-target activity is crucial if CRISPR technology is to be used more widely and especially in a clinical setting⁷⁸. However, many existing methods have differing sensitivities⁷⁹ making comparisons between studies difficult (e.g. CIRCLE-seq has been shown to identify more off-target cleavage sites compared to GUIDE-seq and Digenome-seq, whilst Sanger sequencing identifies more compared to T7E1 assays). Furthermore, many of the original CRISPR-Cas9 results that the field has been built upon utilised suboptimal detection methods^{80,81}. A further complication are disagreements between *in vitro* and *in vivo* results, which have been reported even for some of the most robust methods developed⁵⁴. Together these problems make comparisons and decisions on use difficult. Therefore, moving forward it will be essential that more reliable off-target detection methods are developed, as well as revisiting historic results to verify their accuracy.

Biases in CRISPR research

Another factor hampering our understanding and comparison of CRISPR-Cas9 systems is the lack of standardised studies and benchmarking⁸². Most studies make use of different and limited genetic targets, environments (i.e. *in vivo/in vitro*) and experimental conditions. This makes comparisons impossible and further hinders effective reuse.

An additional bias when assessing CRISPR use is the relatively young age of the technology. Most studies to date have focused on demonstrating successful proofs-of-concept with little concern for the longer-term implications. Furthermore, those moderately longer-term studies that do exist have largely focused on ill-effects^{83,84}. Clearly these handful of examples do not paint a full picture and the reality is that we have a very limited and biased understanding as to the long-term consequences of CRISPR use⁸⁵. Ensuring we are aware of these biases will be crucial when considering possible future deployment into the clinic or the wider environment (e.g. through gene drives^{86,87}).

Ethical, societal and evolutionary concerns

Parallel to scientific advances, ethical and societal concerns have also grown around preclinical research, somatic cell editing, and germline alterations using CRISPR-Cas9. The main focus of these surround germline editing; the work of He Jianku in 2018 that led to the CRISPR-baby scandal re-emphasised the dangers of not regulating this technology⁸⁸. In Jianku's work, the CCR5 gene was largely disabled to confer protection from HIV infection. However, the pleiotropic role of CCR5 suggests likely undesirable long-term side effects⁸⁹. Understanding the full impact of any germline edit is incredibly difficult. It dictates the fate of individuals, forbids consent of future offspring and potentially exposes the lineage to off-target mutagenesis risks^{90,91}, making it ethically questionable in most cases. For those cases where it might be acceptable, it is essential that open and balanced discussions at a societal level are performed to ensure this technology is used in an understood and agreed manner.

From a Darwinian perspective, CRISPR technologies are a powerful means by which individuals could eradicate genes they deem as deleterious from a population. Furthermore, the decision to remove one deleterious gene will likely make it easier to justify the removal of another⁹². This 'slippery slope' ultimately leads to removal of genes in a biased manner, moving from a situation where genome editing is used for medical necessity to one with a selfish purpose, such as enhancing one's offspring⁹³. The ability to select for and against traits would allow humans to act as mediators of natural selection, and bioethicists fear that such control tempts a backlash from nature⁹⁴. What form this might take has yet to be fully understood but has drawn recent attention^{89,95}. Longer-term, the ability to delete variation and distort heritability may eventually call for a revised theory of natural selection with ethical and societal implications that go far beyond clinical applications.

Conclusion

In this review we have shown how robust the CRISPR-Cas9 system is to modifications and extension, allowing its functionality to be tailored for a broad array of genome editing tasks. The rapid development of these systems was made possible by the highly modular structure of both the Cas9 protein and its associated gRNA that allowed in many cases for directed mutations to have a desired impact on the systems overall function. This bodes well for the engineering of other non-Cas9-based CRISPR systems that may better suited to other tasks such as multiplexed DNA editing (e.g. Cas12a^{96,97}) or the localization of enzymatic activities to RNAs (e.g. Cas13⁹⁸).

Whilst the studies explored in this review pave the way for making CRISPR-Cas9 an effective and safe tool, several hurdles spanning both science and society still remain.

Therefore, if maximum benefit is to be realized from this technology, it is vital that future studies widen their scope to consider the wider implications of use and the longer-term impacts they might have on society and the natural world.

Acknowledgements

This work was supported by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre grant BB/L01386X/1 (T.E.G.) and a Royal Society University Research Fellowship grant UF160357 (T.E.G.) This study did not involve any underlying data.

Author Contributions

All authors helped to write the manuscript. G.A.M. and T.E.G. created the figures. T.E.G. supervised the work.

Declaration of Interest

None.

References

1. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. *Nucleotide Sequence of the lap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia Coli, and Identification of the Gene Product*. Vol 169.; 1987:5429-5433. Accessed May 4, 2020. <http://j.b.asm.org/>
2. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. 2006;1. doi:10.1186/1745-6150-1-7
3. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*. 2005;60(2):174-182. doi:10.1007/s00239-004-0046-3
4. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*. 2005;151(3):653-663. doi:10.1099/mic.0.27437-0
5. Sapranaukas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res*. 2011;39(21):9275-9282. doi:10.1093/nar/gkr606
6. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-821. doi:10.1126/science.1225829
7. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*. 2015;348(6242):1477-1481. doi:10.1126/science.aab1452
8. Jinek M, Jiang F, Taylor DW, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014;343(6176). doi:10.1126/science.1247997
9. Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. 2018;19(12):770-788. doi:10.1038/s41576-018-0059-1
10. Barrangou R, Horvath P. A decade of discovery: CRISPR functions and applications. *Nat Microbiol*. 2017;2(7):1-9. doi:10.1038/nmicrobiol.2017.92

11. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019;576(7785):149-157. doi:10.1038/s41586-019-1711-4
12. He Z, Proudfoot C, Mileham AJ, McLaren DG, Whitelaw CBA, Lillico SG. Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol Bioeng*. 2015;112(5):1060-1064. doi:10.1002/bit.25490
13. Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun*. 2018;9(1):1-13. doi:10.1038/s41467-018-04252-2
14. Ishino Y, Krupovic M, Forterre P. History of CRISPR-Cas from encounter with a mysterious repeated sequence to genome editing technology. *J Bacteriol*. 2018;200(7). doi:10.1128/JB.00580-17
15. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*. 2012;109(39):E2579. doi:10.1073/pnas.1208507109
16. Hu JH, Miller SM, Geurts MH, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. 2018;556(7699):57-63. doi:10.1038/nature26155
17. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*. 2014;507(7490):62-67. doi:10.1038/nature13011
18. Szczelkun MD, Tikhomirova MS, Sinkunas T, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A*. 2014;111(27):9798-9803. doi:10.1073/pnas.1402597111
19. Doudna JA, Jiang F. CRISPR–Cas9 Structures and Mechanisms. *Annu Rev Biophys*. 2018;47(1):19-20. doi:10.1146/annurev-biophys
20. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819-823. doi:10.1126/science.1231143
21. Bibikova M, Carroll D, Segal DJ, et al. Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol Cell Biol*. 2001;21(1):289-297. doi:10.1128/MCB.21.1.289-297.2001

22. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol.* 2013;31(9):839-843. doi:10.1038/nbt.2673
23. Semenova E, Jore MM, Datsenko KA, et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A.* 2011;108(25):10098-10103. doi:10.1073/pnas.1104144108
24. Lin Y, Cradick TJ, Brown MT, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* 2014;42(11):7473-7485. doi:10.1093/nar/gku402
25. Satomura A, Nishioka R, Mori H, Sato K, Kuroda K, Ueda M. Precise genome-wide base editing by the CRISPR Nickase system in yeast. *Sci Rep.* 2017;7(1):1-10. doi:10.1038/s41598-017-02013-7
26. Kleinstiver BP, Pattanayak V, Prew MS, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* 2016;529(7587):490-495. doi:10.1038/nature16526
27. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol.* 2014;32(3):279-284. doi:10.1038/nbt.2808
28. Gkazi SA. Quantifying CRISPR off-target effects. *Emerg Top Life Sci.* 2019;3(3):327-334. doi:10.1042/ETLS20180146
29. Sentmanat MF, Peters ST, Florian CP, Connelly JP, Pruett-Miller SM. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep.* 2018;8(1):888. doi:10.1038/s41598-018-19441-8
30. Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 2013;31(9):827-832. doi:10.1038/nbt.2647
31. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* 2013;31(9):822-826. doi:10.1038/nbt.2623
32. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. *Nat Methods.* 2014;11(2):122-123. doi:10.1038/nmeth.2812

33. Bae S, Park J, Kim J-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*. 2014;30(10):1473-1475. doi:10.1093/bioinformatics/btu048
34. Tsai SQ, Joung JK. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet*. 2016;17(5):300-312. doi:10.1038/nrg.2016.28
35. Tsai SQ, Zheng Z, Nguyen NT, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*. 2015;33(2):187-197. doi:10.1038/nbt.3117
36. Kim D, Bae S, Park J, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*. 2015;12(3):237-243. doi:10.1038/nmeth.3284
37. Connolly B, Parsons CA, Benson FE, et al. Resolution of Holliday junctions in vitro requires the Escherichia coli ruvC gene product. *Proc Natl Acad Sci U S A*. 1991;88(14):6063-6067.
38. Martens KJA, van Beljouw SPB, van der Els S, et al. Visualisation of dCas9 target search in vivo using an open-microscopy framework. *Nat Commun*. 2019;10(1):3552. doi:10.1038/s41467-019-11514-0
39. Fonfara I, Le Rhun A, Chylinski K, et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res*. 2014;42(4):2577-2590. doi:10.1093/nar/gkt1074
40. Makarova KS, Koonin EV. Annotation and Classification of CRISPR-Cas Systems. *Methods Mol Biol Clifton NJ*. 2015;1311:47-75. doi:10.1007/978-1-4939-2687-9_4
41. Acharya S, Mishra A, Paul D, et al. Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proc Natl Acad Sci U S A*. 2019;116(42):20959-20968. doi:10.1073/pnas.1818461116
42. Nishimasu H, Cong L, Yan WX, et al. Crystal Structure of Staphylococcus aureus Cas9. *Cell*. 2015;162(5):1113-1126. doi:10.1016/j.cell.2015.08.007
43. Kleinstiver BP, Prew MS, Tsai SQ, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015;523(7561):481-485. doi:10.1038/nature14592

44. Kim E, Koo T, Park SW, et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat Commun*. 2017;8(1):14500. doi:10.1038/ncomms14500
45. Yamada M, Watanabe Y, Gootenberg JS, et al. Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Mol Cell*. 2017;65(6):1109-1121.e3. doi:10.1016/j.molcel.2017.02.007
46. Friedland AE, Baral R, Singhal P, et al. Characterization of *Staphylococcus aureus* Cas9: A smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol*. 2015;16(1):257. doi:10.1186/s13059-015-0817-8
47. Nishimasu H, Ran FA, Hsu PD, et al. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*. 2014;156(5):935-949. doi:10.1016/j.cell.2014.02.001
48. Chen F, Ding X, Feng Y, Seebeck T, Jiang Y, Davis GD. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun*. 2017;8(1):14958. doi:10.1038/ncomms14958
49. Hirano H, Gootenberg JS, Horii T, et al. Structure and Engineering of *Francisella novicida* Cas9. *Cell*. 2016;164(5):950-961. doi:10.1016/j.cell.2016.01.039
50. Casini A, Olivieri M, Petris G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol*. 2018;36(3):265-271. doi:10.1038/nbt.4066
51. Struhl K. Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell*. 1999;98(1):1-4. doi:10.1016/S0092-8674(00)80599-1
52. Dugar G, Leenay RT, Eisenbart SK, et al. CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the *Campylobacter jejuni* Cas9. *Mol Cell*. 2018;69(5):893-905.e7. doi:10.1016/j.molcel.2018.01.032
53. Hsu PD, Lander ES, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*. 2014;157(6):1262-1278. doi:10.1016/j.cell.2014.05.010
54. Cullot G, Boutin J, Toutain J, et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat Commun*. 2019;10(1):1-14. doi:10.1038/s41467-019-09006-2
55. Dianov GL, Hübscher U. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res*. 2013;41(6):3483-3490. doi:10.1093/nar/gkt076

56. Ran FA, Hsu PD, Lin C-Y, et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*. 2013;154(6):1380-1389. doi:10.1016/j.cell.2013.08.021
57. Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 2014;24(1):132-141. doi:10.1101/gr.162339.113
58. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533(7603):420-424. doi:10.1038/nature17946
59. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. 2013;8(11):2180-2196. doi:10.1038/nprot.2013.132
60. Qi LS, Larson MH, Gilbert LA, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*. 2013;152(5):1173-1183. doi:10.1016/j.cell.2013.02.022
61. McNally SG, Hagen KD, Nosala C, et al. Robust and stable transcriptional repression in *Giardia* using CRISPRi. *Mol Biol Cell*. 2019;30(1):119-130. doi:10.1091/mbc.E18-09-0605
62. Tsai SQ, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*. 2014;32(6):569-576. doi:10.1038/nbt.2908
63. Havlicek S, Shen Y, Alpagu Y, et al. Re-engineered RNA-Guided FokI-Nucleases for Improved Genome Editing in Human Cells. *Mol Ther*. 2017;25(2):342-355. doi:10.1016/j.ymthe.2016.11.007
64. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol*. 2014;32(6):577-582. doi:10.1038/nbt.2909
65. Kim E, Kim S, Kim DH, Choi B-S, Choi I-Y, Kim J-S. Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res*. 2012;22(7):1327-1333. doi:10.1101/gr.138792.112

66. Dong B, Nakai H, Xiao W. Characterization of Genome Integrity for Oversized Recombinant AAV Vector. *Mol Ther*. 2010;18(1):87-92. doi:10.1038/mt.2009.258
67. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016;351(6268):84-88. doi:10.1126/science.aad5227
68. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014;513(7519):569-573. doi:10.1038/nature13579
69. Kleinstiver BP, Prew MS, Tsai SQ, et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol*. 2015;33(12):1293-1298. doi:10.1038/nbt.3404
70. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice: Molecular Cell. Accessed May 23, 2020. [https://www.cell.com/molecular-cell/fulltext/S1097-2765\(12\)00656-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276512006569%3Fshowall%3Dtrue](https://www.cell.com/molecular-cell/fulltext/S1097-2765(12)00656-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276512006569%3Fshowall%3Dtrue)
71. Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol*. Published online June 22, 2020. doi:10.1038/s41587-020-0561-9
72. Jore MM, Lundgren M, van Duijn E, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*. 2011;18(5):529-536. doi:10.1038/nsmb.2019
73. Nishida K, Arazoe T, Yachie N, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science*. 2016;353(6305). doi:10.1126/science.aaf8729
74. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol*. 2008;9(6):229. doi:10.1186/gb-2008-9-6-229
75. Schellenberger V, Wang C, Geething NC, et al. A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol*. 2009;27(12):1186-1190. doi:10.1038/nbt.1588

76. Komor AC, Zhao KT, Packer MS, et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv.* 2017;3(8):eaao4774. doi:10.1126/sciadv.aao4774
77. Kim YB, Komor AC, Levy JM, Packer MS, Zhao KT, Liu DR. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol.* 2017;35(4):371-376. doi:10.1038/nbt.3803
78. Hirakawa MP, Krishnakumar R, Timlin JA, Carney JP, Butler KS. Gene editing and CRISPR in the clinic: current and future perspectives. *Biosci Rep.* 2020;40(4). doi:10.1042/BSR20200127
79. Zischewski J, Fischer R, Bortesi L. Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv.* 2017;35(1):95-104. doi:10.1016/j.biotechadv.2016.12.003
80. Cho SW, Kim S, Kim JM, Kim J-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol.* 2013;31(3):230-232. doi:10.1038/nbt.2507
81. Duan J, Lu G, Xie Z, et al. Genome-wide identification of CRISPR/Cas9 off-targets in human genome. *Cell Res.* 2014;24(8):1009-1012. doi:10.1038/cr.2014.87
82. Raitskin O, Schudoma C, West A, Patron NJ. Comparison of efficiency and specificity of CRISPR-associated (Cas) nucleases in plants: An expanded toolkit for precision genome engineering. *PLoS ONE.* 2019;14(2). doi:10.1371/journal.pone.0211598
83. Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med.* 2018;24(7):927-930. doi:10.1038/s41591-018-0049-z
84. Ihry RJ, Worringer KA, Salick MR, et al. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nat Med.* 2018;24(7):939-946. doi:10.1038/s41591-018-0050-6
85. Baylis F, McLeod M. First-in-human Phase 1 CRISPR Gene Editing Cancer Trials:Are We Ready? *Curr Gene Ther.* 2017;17(4):309-319. doi:10.2174/1566523217666171121165935

86. Brossard D, Belluck P, Gould F, Wirz CD. Promises and perils of gene drives: Navigating the communication of complex, post-normal science. *Proc Natl Acad Sci*. 2019;116(16):7692-7697. doi:10.1073/pnas.1805874115
87. Esvelt KM, Smidler AL, Catteruccia F, Church GM. Concerning RNA-guided gene drives for the alteration of wild populations. Tautz D, ed. *eLife*. 2014;3:e03401. doi:10.7554/eLife.03401
88. Cyranoski D, Ledford H. Genome-edited baby claim provokes international outcry. *Nature*. 2018;563(7733):607-608. doi:10.1038/d41586-018-07545-0
89. Li T, Shen X. Pleiotropy Complicates Human Gene Editing: CCR5 Δ 32 and Beyond. *Front Genet*. 2019;10. doi:10.3389/fgene.2019.00669
90. Manghwar H, Li B, Ding X, et al. CRISPR/Cas Systems in Genome Editing: Methodologies and Tools for sgRNA Design, Off-Target Evaluation, and Strategies to Mitigate Off-Target Effects. *Adv Sci*. 2020;7(6). doi:10.1002/advs.201902312
91. Kim HJ, Lee HJ, Kim H, Cho SW, Kim J-S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res*. 2009;19(7):1279-1288. doi:10.1101/gr.089417.108
92. HUMAN GENE THERAPY: DOWN THE SLIPPERY SLOPE? - HOLTUG - 1993 - Bioethics - Wiley Online Library. Accessed July 13, 2020. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8519.1993.tb00231.x?sid=nlm%3Apubmed>
93. Omodamilola OI. CRISPR Technology; Advantages, Limitations and Future Direction. 2018;1(2):5.
94. Peters T. Should CRISPR Scientists Play God? *Religions*. 2017;8(4):61. doi:10.3390/rel8040061
95. Quammen D. *The Tangled Tree*. Simon & Schuster; 2019. Accessed June 18, 2020. <https://www.simonandschuster.com/books/The-Tangled-Tree/David-Quammen/9781476776637>
96. Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. 2015;163(3):759-771. doi:10.1016/j.cell.2015.09.038

97. Ciurkot K, Vonk B, Gorochowski TE, Roubos JA, Verwaal R. CRISPR/Cas12a Multiplex Genome Editing of *Saccharomyces cerevisiae* and the Creation of Yeast Pixel Art. *JoVE J Vis Exp*. 2019;(147):e59350. doi:10.3791/59350
98. Cox DBT, Gootenberg JS, Abudayyeh OO, et al. RNA editing with CRISPR-Cas13. *Science*. 2017;358(6366):1019-1027. doi:10.1126/science.aag0180

Figures and captions

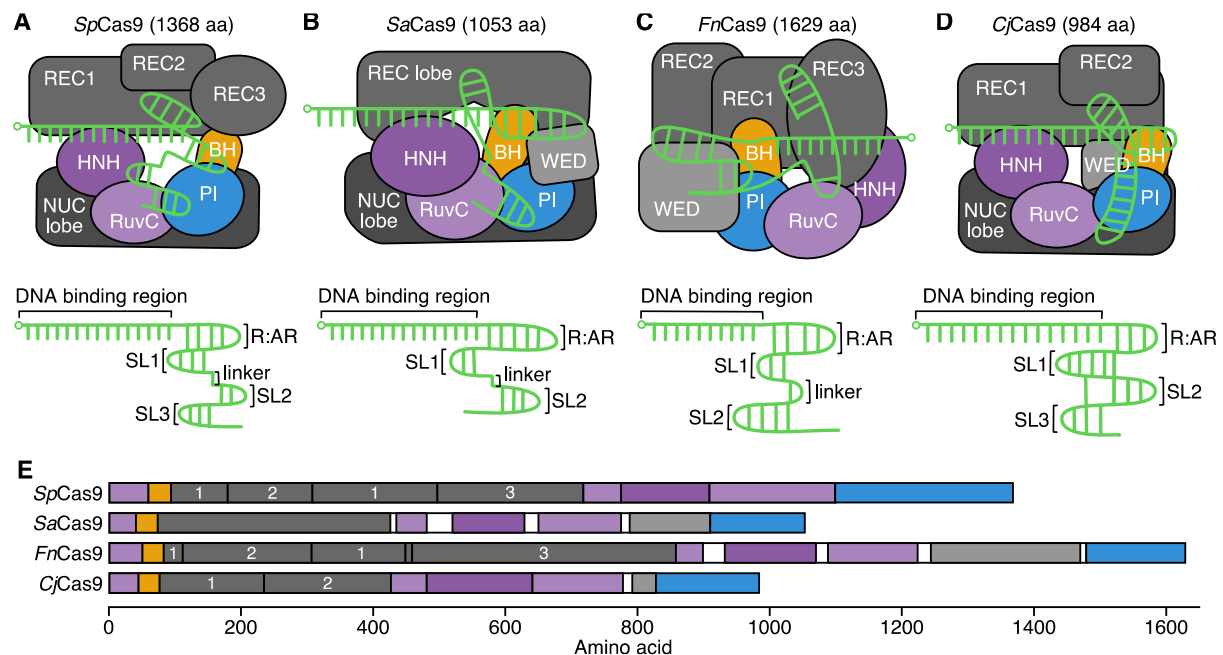


Figure 1: Naturally occurring variants of Cas9 and their respective gRNA structures.

Top diagrams show the Cas9:gRNA complex and interactions of the gRNA with core Cas9 domains (labelled in white). Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting, CTD = C-terminal domain, WED = wedge. Bottom diagrams show the gRNA structure with the DNA binding region, major stem loops (SLs) and repeat:antirepeat (R:AR) duplex highlighted. The 5' end of each gRNA is denoted by a small circle. (A) *Streptococcus pyogenes* Cas9 (*SpCas9*). (B) *Staphylococcus aureus* Cas9 (*SaCas9*). (C) *Francisella novicida* Cas9 (*FnCas9*). (D) *Campylobacter jejuni* Cas9 (*CjCas9*). (E) Domain structure of the Cas9 variants. Linkers are denoted by white regions and REC domains are numbered in white text where relevant.

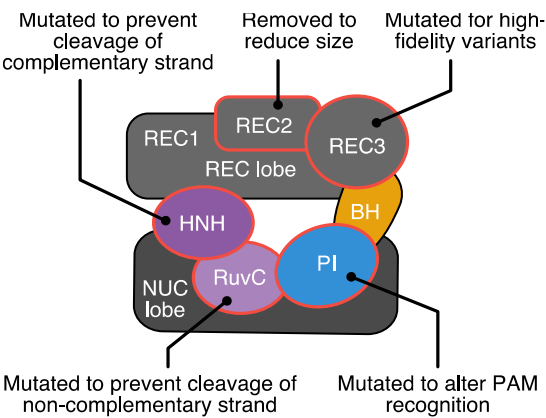


Figure 2: Key domains of Cas9 and the effect of modifications of each on phenotype. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting.

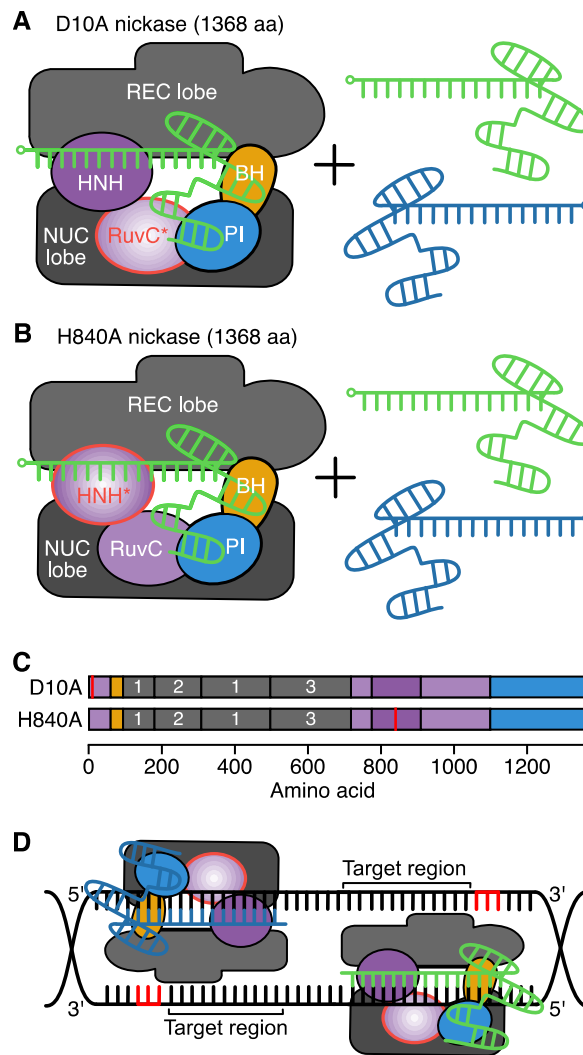


Figure 3: Cas9D10A and Cas9H840A nickase systems. (A) The Cas9D10A nickase system which nicks the complementary strand. This Cas9D10A is used in conjunction with a pair of guides to target each strand independently. The 5' end of each gRNA is denoted by a small circle and inactive domains are outlined in red. (B) A complementary Cas9H840A nickase system is able to nick the non-target strand. Again, this system is normally used with two complementary guides to target each strand of DNA. (C) Domain structure of the nickase system. Mutations are shown by bright red lines and three REC domains numbered in white text. (D) Example of the Cas9D10A nickase system targeting two regions to create complementary nicks on opposite strands. The PAM is shown in red.

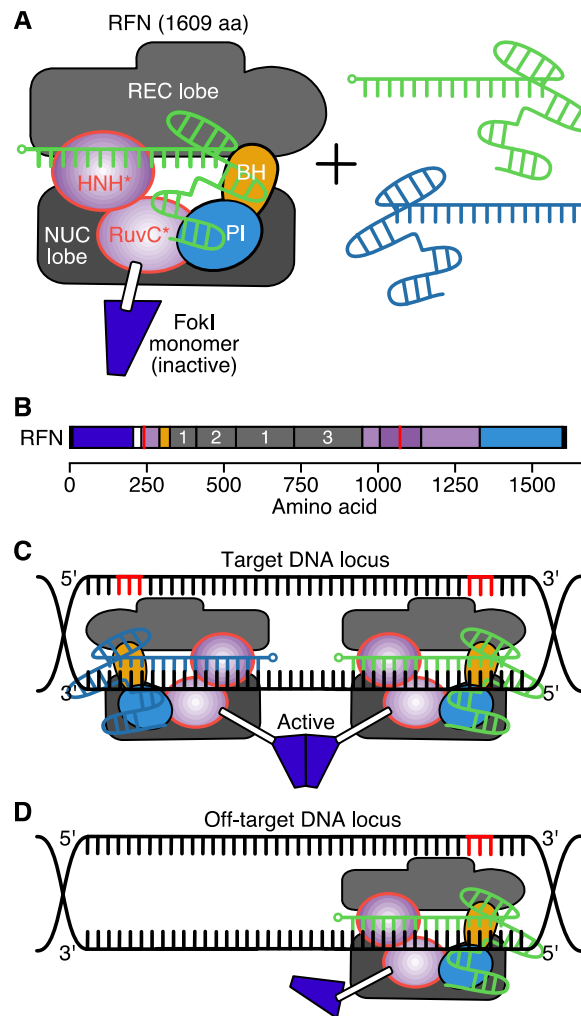


Figure 4: An RNA-guided FokI Nuclease (RFN) system. (A) An RFN system consists of a dCas9-FokI fusion and two gRNA (green and blue) with targets ~15 bp apart. Two FokI monomers (blue) are required for the active dimer (purple) to cleave DNA, so off-target binding of a single RFN does not (usually) result in cleavage. Domains outlined in red are inactive. The 5' end of each gRNA is denoted by a small circle. The PAM is shown in red. Linkers are denoted by white rectangles. (B) Domain structure of the RFN. Linkers and nuclear localization signals (NLSs) are denoted by white and black regions, respectively, and mutations are shown by bright red lines. Three REC domains numbered in white text. (C) Two RFNs bound in an active conformation to a target DNA locus. (D) Single inactive RFN bound to an off-target DNA locus.

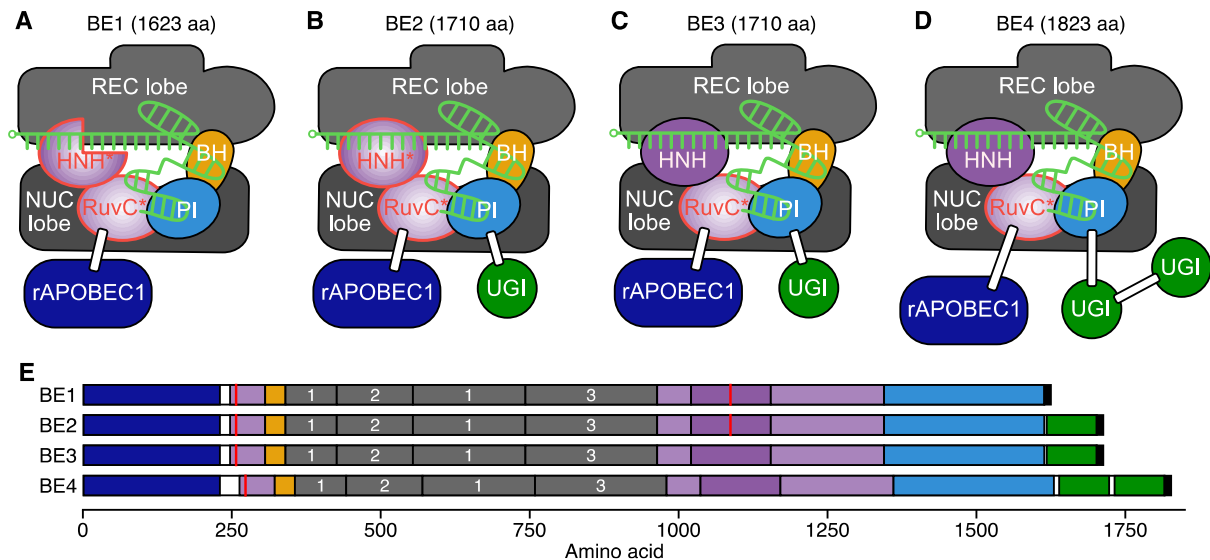


Figure 5: Base editing systems. (A) Base editor 1 (BE1) consists of a SpdCas9 with a cytidine deaminase (rAPOBEC1) fused to its N-terminus. (B) Base editor 2 (BE2) is similar to BE1 but includes an additional uracil glycosylase inhibitor (UGI) fused to the C-terminus. (C) Base editor 3 (BE3) is similar to BE2 but includes the catalytic activity of the HNH nuclease domain restored, to allow target strand nicking. (D) Base editor 4 is as BE3 but with longer linker proteins and an additional UGI fused to the C terminus. The 5' end of each gRNA is denoted by a small circle. Linkers are denoted by white rectangles. Mutated domains are outlined in red. (E) Domain structure of the base editors. Linkers and nuclear localization signals (NLSs) are denoted by white and black regions, respectively, and mutations are shown by bright red lines. Three REC domains numbered in white text.

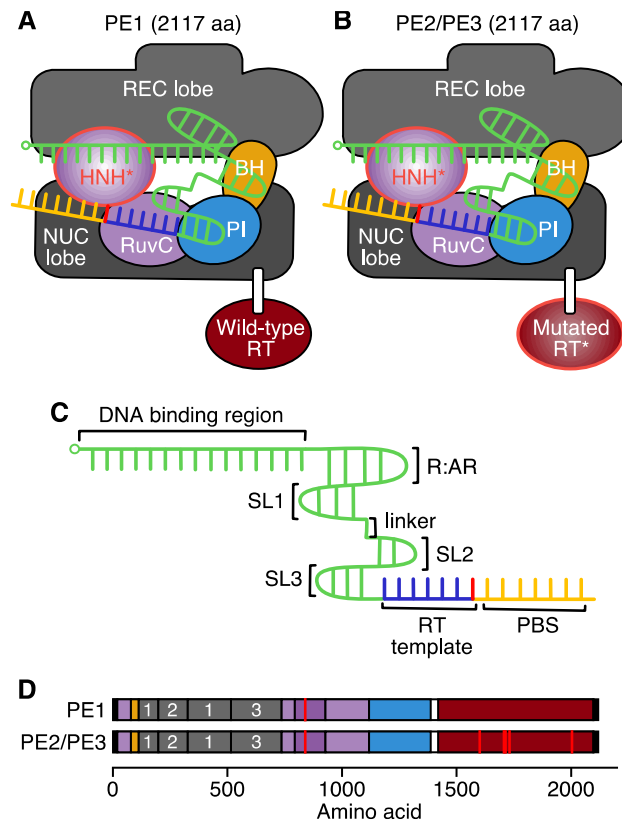


Figure 6: Prime editing systems and pegRNA. (A) Prime editor 1 (PE1) consists of a H840A nickase with a flexible linker fusing an M-MLV wild-type (WT) reverse transcriptase (RT; red) to the C-terminus. Linkers are denoted by white rectangles. Mutated domains are outlined in red. (B) Prime editor 2 (PE2) is similar to PE1 but contains a mutated/engineered RT rather than the WT variant. Prime editor 3 is identical to PE2 but makes use of an additional gRNA targeting the unedited strand, allowing for increased editing efficiency. (C) The pegRNA consists of a seed region and sgRNA (green) with a primer binding site (PBS; dark yellow) and repair template (RT template, blue) containing a base edit (red). The PBS binds to the nicked strand for initiation of repair via RT, using the repair template. The 5' end of each gRNA is denoted by a small circle. (D) Domain structure of the prime editors. Linkers and nuclear localization signals (NLSs) are denoted by white and black regions, respectively, and mutations are shown by bright red lines. Three REC domains numbered in white text.