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

Grace A. Meaker\textsuperscript{1,2}, Emma J. Hair\textsuperscript{1} and Thomas E. Gorochowski\textsuperscript{1,3,*}

\textsuperscript{1} School of Biological Sciences, University of Bristol, Tyndall Avenue, Bristol BS8 1TQ, UK
\textsuperscript{2} School of Biosciences, Cardiff University, Museum Avenue, Cardiff, CF10 3AT, UK
\textsuperscript{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 extend 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 1987\(^1\), the exact function of the clustered regularly interspaced short palindromic repeats (CRISPR) remained a mystery until the further discovery of CRISPR-associated (Cas) proteins and RNA elements. This established their combined function as a prokaryotic immune system\(^2\)–\(^5\), which had evolved to combat invading phages by cleaving and degrading their DNA. The core components are a Cas 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)\(^8\) (Figure 1).

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 replacement\(^9\)–\(^13\). This broad and significant utility has resulted in the term “CRISPR” becoming synonymous with CRISPR-Cas systems and their application.

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 scientific, ethical and evolutionary challenges associated with this technology.

It should be noted that other CRISPR systems employing alternative Cas proteins do exist and have begun to gain interest due to their unique and often complementary capabilities. For example, CRISPR-Cas12a based systems have been shown to simplify multiplexed editing and combinatorial screens due to their ability to process CRISPR arrays directly\(^14\)–\(^18\). However, Cas9 based systems are by far the most commonly used and modified to date, and so form the focus of this review.

The native CRISPR-Cas9 system
The CRISPR-Cas9 system is formally classified as a class 2, type II CRISPR system, which was originally derived from *Streptococcus pyogenes*. It consists of a Cas nuclease SpCas9 and a gRNA (Figure 1). The gRNA has two components – a trans-activating RNA (tracrRNA) and a CRISPR RNA (crRNA) (Figure 1A). crRNA is responsible for recognition and binding of the target DNA region and tracrRNA for crRNA maturation and association with SpCas9. Alternatively, a chimeric single-guide RNA (sgRNA) which performs both these functions can be used (Figure 1B). 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 strands. The target region is determined by a 20-nucleotide ‘spacer’ in the crRNA, complementary to the target ‘protospacer’ in the DNA. For recognition, the protospacer must be superseded at the 3’ end by several nucleotides called the protospacer adjacent motif (PAM). This varies for different Cas proteins; for SpCas9 it is ‘5-NGG-3′. Providing there is the correct PAM present at the 3’ end 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. Once the complex is bound, the two nucleases produce a double-stranded break (DSB) 3–4 nucleotides (nt) upstream of the PAM. 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 mutated (Figure 1C). Alternatively, the homology-directed repair pathway (HDR) can be used to fix the break using a homologous template to accurately insert a desired sequence. HDR is preferred to NHEJ in certain organisms (e.g. *S. cerevisiae*) as well as in cells containing a repair template (e.g. cells post S phase of the cell cycle). 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 genomes. The genomic target has 20 nt of complementarity to the spacer region of the gRNA, but mutations of the 5’ end of the gRNA still permit efficient cleavage implying only 12–13 nt at the 3’ end of the spacer region are critical for specifying the target. These essential 13 nt have been dubbed the ‘seed sequence’. 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 effects. 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\textsuperscript{31–33} have been the focus of efforts to reduce such promiscuity and will be discussed later.

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 cleavage\textsuperscript{34}. 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 the T7E1 assay\textsuperscript{35}. 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’\textsuperscript{36,37}.

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 predict potential off-target loci\textsuperscript{38,39}. These approaches, however, do not consider off-target sites which do not fit the model’s parameters\textsuperscript{40}. To alleviate this issue, machine learning methods have recently been shown to offer improved performance\textsuperscript{41}. Experimentally, Genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq) provides a robust empirical method for identifying off-target effects and has become widely used\textsuperscript{42}. A small oligo-nucleotide tag is integrated into DSB sites targeted by NHEJ, and sequencing analysis is used to pinpoint the location of off-target sites. This permits the detection of sites difficult to capture with computational tools due to the complexity of the underlying rules and interactions\textsuperscript{38}. 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\textsuperscript{40,42}. Another example of a genome-wide tool is digested genome sequencing (Dig genome-seq) which involves the digestion of genomic DNA with Cas9-gRNA complexes and subsequent deep sequencing to identify identical Cas9 cleavage fragments\textsuperscript{43}. Analysis is performed on extracted DNA, eliminating the influence of cellular context (e.g. chromatin arrangements, methylation patterns and DNA accessibility). This method is time-consuming as many reads have to be analyzed to identify patterns, and it fails to recognize identical fragments caused by chance\textsuperscript{40}. Overall, no single method is able to comprehensively analyse off-target effects, therefore the method employed must be carefully considered on a case-by-case basis. For example, Dig genome-seq is appropriate for \textit{in vitro} applications because it is not vulnerable to chromatin arrangements\textsuperscript{43}, but for \textit{in vivo} applications, GUIDE-seq or the new, multiplexing sister method Tagmentation-based tag integration site sequencing (TTI ISS) are more sensitive and easier to use\textsuperscript{42,44}. 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. Whether the field of genome engineering can expect more accurate predictions will largely depend on the ability...
to combine versatile algorithms with ultrasensitive, genome-wide off-target detection methods and predictive modelling\textsuperscript{45,46}.

### Naturally occurring variants

CRISPR is a naturally occurring system in prokaryotes, thus different species possess different systems whose variations can be potentially exploited\textsuperscript{47}. Type I and III systems enlist multiple Cas proteins whereas type II uses a single, Cas9 protein for DNA cleavage\textsuperscript{48}. Whilst SpCas9 from \textit{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 (Figure 2). This includes \textit{FnCas9} from \textit{Francisella novicida}\textsuperscript{49}, \textit{SaCas9} from \textit{Staphylococcus aureus}\textsuperscript{50,51} and recently the \textit{Campylobacter jejuni} Cas9, the smallest to date\textsuperscript{52,53}.

\textit{SpCas9} is a multi-domain protein exhibiting a bilobed structure where the nuclease lobe and the recognition lobe\textsuperscript{8,24} are linked by an arginine-rich bridge helix as well as a disordered linker\textsuperscript{8} (Figure 2A). The overall shape of \textit{SpCas9} is oblong with two large grooves, to accommodate the DNA:RNA and RNA:RNA complexes. Adaptations of the two previously-recognized, adjacent nucleases (HNH\textsuperscript{6}, named for the three characteristic residues, and RuvC\textsuperscript{54}) of the nuclease lobe facilitate much of the diversification of CRISPR's function\textsuperscript{31,55}. Each nuclease cleaves one strand of DNA; RuvC cleaves the non-complementary and HNH the complementary strand\textsuperscript{6,20}. 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) domain\textsuperscript{7}. Mutagenesis of these domains permits the evolution of CRISPR function.

\textit{SaCas9} has a longer PAM (5'-NNGRRT-3') than \textit{SpCas9} and is smaller at 1053 amino acids (aa) compared to 1368 aa\textsuperscript{50} (Figure 2B). Due to its smaller size, \textit{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 \textit{SaCas9} has shown comparable on-target cleavage to \textit{SpCas9}, whilst boasting a higher specificity and easier introduction into cells\textsuperscript{56}. Both \textit{SpCas9} and \textit{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, \textit{SaCas9} only has 17% structural similarity to \textit{SpCas9}; key DNA/RNA binding domains such as the nucleases and the 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 efficacy\textsuperscript{56}. Despite these differences, it is apparent that \textit{SaCas9} and \textit{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 domain\textsuperscript{57}.

Another SpCas9 ortholog is FnCas9 which produces staggered cleavage and binds less frequently to non-target regions\textsuperscript{49,58} (Figure 2C). 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 HDR\textsuperscript{49}. FnCas9 is considerably larger than SpCas9 and SaCa9, comprised of 1629 aa\textsuperscript{69}. Whilst its larger size may be a hindrance for transfection due to the limited capacity of many delivery systems, 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 cleavage\textsuperscript{58}. This increased specificity means FnCas9 produces far less off-target cleavage as fewer sites are recognized as ‘target’\textsuperscript{49}. FnCas9 is structurally dissimilar to SpCas9 and SaCas9, lacking bilobed structure and containing distinct REC2 and REC3 domains (Figure 2C). REC3 domain mutations have generated high-fidelity Cas9 enzymes\textsuperscript{80}; 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 literature\textsuperscript{58}, local chromatin conformations likely affect the access to DNA, a vulnerability not as significant for SpCas9. To eliminate this problem FnCas9 has been used alongside a catalytically dead SpCas9 (SpdCas9) to enable access and subsequent DNA cleavage\textsuperscript{58}. Such problems are not present when used in prokaryotes where FnCas9 has been shown to function effectively\textsuperscript{61}.

Finally, CjCas9 is the smallest ortholog characterized to date at only 984 aa, making it suitable for size-restricted delivery methods such as those using adeno-associated viruses (AAV) (Figure 2D). It has a bilobed structure, akin to SaCas9 and SpCas9, with a simplified REC lobe and size-reduced NUC lobe\textsuperscript{53} (Figure 2D). Initial studies showed recognition of a 5'-NNNACA-3' PAM\textsuperscript{47} or the more promiscuous 5'-NNNVRYM-3\textsuperscript{53} providing an assortment of target sites. However, recent studies have found a requirement for an 8th cytosine at the 3’ end, suggesting 5'-NNNNRYAC-3\textsuperscript{52} and 5'-NNNNACAC-3' sequences\textsuperscript{62}. Tested against SaCas9 in human cells, CjCas9 was found to be more specific with comparable efficiencies to some other variants, excluding FnCas9\textsuperscript{52}. 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 2). The essential region of the sgRNA consists of a DNA binding region, the repeat:anti-repeat 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. In contrast, removal of loops 2 or 3 decreases efficiency, without abolishing cleavage. Stem loop 2 interacts with the PI and RuvC domains in SaCas9 and SpCas9, and the REC domains in FnCas9 and CjCas9. SaCas9 and SpCas9’s sgRNAs exhibit the greatest similarity, particularly regarding cognate Cas9 interactions with the lack of stem loop 3 in SaCas9 the defining key difference. This further highlights the minimalism of SaCas9 compared to SpCas9 because of the reduction of non-essential elements like stem loop 3 and the REC2 domain. 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. The novel structural arrangement of CjCas9’s gRNA forms a triple helix between stem loops 1, 2, and 3. The relevance of this structure is still unknown due to a lack of comprehensive structural studies of CjCas9 complexes.

The domains of each Cas9 distinctly interact with their associated sgRNAs due to the slight differences in sgRNA structure (Figure 2). 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 2, 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 cleavage. 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 efficacy. 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 demonstrated significantly lower off-target activity compared to full-length sgRNAs due to a reduction in binding affinity and greater mismatch intolerance\(^3^9,6^3\). As demonstrated in two human cell lines, the specificity of tru-gRNAs as compared to wild-type was estimated to be >5,000-fold higher\(^3^3\). Such estimates are supported by the finding that additional nucleotides added at the 5’ end of gRNA increase binding affinity for off-target sites\(^2^8\). 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.

Beyond sequence changes to gRNAs, another method that has been used to improve editing efficiency is the chemical modification of key nucleotides. Chemically synthesized and modified sgRNAs have shown significantly improved editing efficiencies in human primary T cells and CD34+ hematopoietic stem and progenitor cells (HSPCs)\(^6^4\). The ability for Cas9 to handle significant modifications has enabled the effective use of gRNAs with >80% ribose substitutions and at least one chemical modification (e.g. 2’-O-methyl, 2’-Fluoro, phosphorothioate) at every nucleotide position\(^6^4\). Such modifications are useful as they can help ensure metabolic stability and reduce the chance of nanoparticle formation, which can elicit an immune response. Furthermore, such modifications offer the ability to use chemical conjugates as a means to target the cell-surface and improve uptake\(^6^5\).

**Modification of Cas9**

Another method to improve performance is through modification of the Cas9 enzyme itself (Figure 3). 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\(^6,2^0\). As double-stranded cleavage often favors the inaccurate NHEJ pathway (depending on the organism, cell type and stage in the cell cycle), single-stranded cleavage (or ‘nicking’) is favorable for efficient targeted replacement\(^2^7\). 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)\(^6^6,6^7\). 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\(^2^0,2^5,3^1\). The benefits of these are twofold: they produce precise nicks in the DNA and exhibit decreased affinity to off-target loci\(^3^1\). 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 4)\(^6^8\). This dual nicking strategy has been shown to have comparable on-
target cleavage to SpCas9 whilst discriminating off-target sites more effectively, however, requires the presence of two neighboring PAM sites which limits the number of potential editing sites. Continued editing of nickases forms the basis of many other CRISPR editing systems that will be explored in the next section. Additional reductions in off-target effects have also been achieved by controlling the expression and stability of the Cas9 protein. For example, increasing the degradation rate of Cas9 by adding a ubiquitin-targeting signal added to the N-terminus has been shown to decrease mosaicism in monkey embryos. Furthermore, the addition of an N-terminus geminin tag to Cas9 has been used to regulate Cas9 concentration in response to the cell cycle allowing the editing capacity to be maintained while greatly reducing neurotoxicity.

As a mutation in one of the nuclease domains can alter Cas9 from a dsDNA endonuclease to a ssDNA nickase, mutation of both domains will remove all cleavage activity. An SpCas9 enzyme containing the H840A and D10A mutations is catalytically dead (dCas9), 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 enzymes or used to visualize target affinity without cleavage. Such an approach has enabled the development of programmable DNA methylation systems formed from a dCas9 protein fused to a DNA (cytosine-5)-methyltransferase 3A. This particular system permitted up to 50% methylation for targeted CpG dinucleotides in HEK293T cells and a better understanding of the influence chromatin organization and dynamics plays has on gene expression. Particularly in human cells, programmable DNA methylation systems allow for the visualization of specific genetic loci via a dCas-eGFP fusion and fluorescence microscopy.

Furthermore, dCas9 has become widely used in regulating gene expression through CRISPR interference and activation (CRISPRi and CRISPRa, respectively). Interference of gene expression is generally achieved by targeting the dCas9 protein to promoter regions and sterically blocking the initiation of RNA polymerase (RNAP). Additional, repression domains (e.g. KRAB) can also be fused to the dCas9 to enhance repression. This ability to inhibit but not completely turn off gene expression has made CRISPRi a valuable tool for knock-down screens where Cas9 is not suitable (e.g. due to genotoxicity). Activation of gene expression has been similarly achieved by fusing transcription activating domains (e.g. VP64 for human cells or SoxS for Escherichia coli) to dCas9, or by modifying the sgRNA and using an RNA binding protein (e.g. MS2 coat protein) fused to an activator domain that can then be targeted to this sgRNA. In both cases, targeting these systems to regions upstream of a promoter without blocking transcription initiation enables activation of the downstream gene.

An additional application of dCas9 concerns fusion to a FokI nuclease, an endonuclease which is strictly dependent on dimerization for cleavage activity. This fusion enlists a long, flexible linker with between 5-25 residues (e.g. GGGGS) fusing the FokI
endonuclease to the Cas9 N-terminus. The RNA-guided FokI Nuclease (RFN) system consists of a dCas9-FokI fusion and two different gRNAs. 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 occur (Figure 5). 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 gRNAs. 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 these. This fusion system is also very large, limiting its application in AAV delivery methods. Efforts have been made to use the smaller SaCas9 based system instead of SpCas9, reducing the size and simplifying delivery. Despite some documented successes, it is worth noting the range of confounding effects associated with the different delivery methods. For example, a complication when employing lentivirus vectors concerns long-term Cas9 expression which promotes the likelihood of off-target effects. In contrast, Cas9 ribonucleoproteins are limited by transient expression and possible reduced on-target activity.

**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 regions. 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. 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’. 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' PAM, whose number may be further reduced by local chromatin or
methylation patterns preventing Cas9 access to the site. 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 PAM. 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' PAM. 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 gene. 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. 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). Such variants, dubbed ‘xCas9’ nuclease, had a different pattern of mutations than the rationally developed variants which covered the entire cas9 gene. 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. 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'-NGRRT-3' sites (and off-target loci) at a similar efficiency to wild-type SaCas9 whilst also cleaving sites adjacent to 5'-NNARRT-3', 5'-NNTRTT-3' and 5'-NCCRTT-3'.

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* due to the difficulties associated with successful delivery of both the editing machinery and template DNA. Additionally, both of these DNA repair pathways rely on the generation of DSBs, which can result in inadvertent genomic alterations, pathogenic lesions and deleterious tumor suppressor p53 activation responses. Single-stranded nicks are repaired by the high-fidelity BER pathway, making this cleavage pattern preferable for specific base changes.

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 editors (Figure 6). A simple CRISPR base editor consists of a dCas9 protein, a sgRNA and a base editing enzyme (e.g. cytidine deaminase). Cytidine deaminases catalyze the conversion of cytosine to uracil and the rat cytidine deaminase (rAPOBEC1) has been selected in several 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 (Figure 6A). 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 human cells. This discrepancy was attributed to the endogenous DNA repair machinery, specifically uracil DNA glycosylase (UDG), which reverses the UG pair to a CG pair. 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 6B). This alteration increased editing efficiencies in human cells 3-fold as UDG activity was drastically reduced. 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 6C). 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 6D). 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 linker. BE4 showed higher C to T editing efficiency and 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, SaBE3\textsuperscript{100}. As previously described, SaCas9 is much smaller than SpCas9\textsuperscript{50} 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\textsuperscript{72,100}. Furthermore, Kim and colleagues utilized SpCas9 variants with altered PAM specificities, specifically VQR and VRER (described previously) and EQR from the same study\textsuperscript{51}, as well as an engineered SaCas9 variant, SaKKH\textsuperscript{93}. 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 levels\textsuperscript{100}. 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 7). 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)\textsuperscript{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 7A). PE1 was able to generate transversion mutations at efficiencies of up to 5.5% and insertions and deletions of up to 17%\textsuperscript{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 7B). 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 7C). 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 repair\textsuperscript{11}.

**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 setting\textsuperscript{101}. However, many existing methods have differing sensitivities\textsuperscript{102} 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\textsuperscript{103,104}. A further complication concerns the disagreements between \textit{in vitro} and \textit{in vivo} results, which have been reported even for some of the most robust methods developed\textsuperscript{66}. 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.

\textbf{Limitations in CRISPR research}

Another factor hampering our understanding and comparison of CRISPR-Cas9 systems is the lack of standardised studies and benchmarking\textsuperscript{105}. Most studies to date have made use of different genetic targets of a limited number, with experiments performed under a variety of environments (i.e. \textit{in vivo}/\textit{in vitro}) and conditions. While this is understandable given the often-applied focus of research to a particular disease, it does however make clear comparisons between methods impossible and further hinders effective reuse of data. In other areas like sequencing, standardised materials have been developed to allow for the robust benchmarking of methods (e.g. synthetic RNA libraries to assess the accuracy of read counts\textsuperscript{106} and defined microbial communities to test metagenomic inference from mixed pools of organisms\textsuperscript{107}). Although difficult given the broad potential applications of CRISPR, having a set of standardised organisms, cell lines, targets and conditions that cover a wide variety of possibilities would greatly aid in the unbiased assessment of new methods and ensure results can be directly compared. It should be noted that such issues with standardisation do not only affect CRISPR research but are a challenge across the whole of the synthetic biology and bioengineering fields.

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 e.g. effects on the tumour suppressor gene, \textit{p53}\textsuperscript{108,109}. Clearly, this 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 use\textsuperscript{110}. 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\textsuperscript{111,112}).
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 technology\textsuperscript{113}. 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 effects\textsuperscript{114}. 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\textsuperscript{115,116}, making it ethically questionable in most cases. For those cases where it might be acceptable, open and balanced discussions at a societal level must be performed to ensure this technology is used in an understood and agreed manner. Such ethical considerations should also extend to that of the manufacturing sectors (e.g. agriculture, pharmaceutical and chemical). Although there is promise for CRISPR technologies here, genetically modified food controversies, arguments concerning human health and environmental implications threaten such uses.

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 another\textsuperscript{117}. 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 offspring\textsuperscript{118}. 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 nature\textsuperscript{119}. What form this might take has yet to be fully understood but has drawn recent attention\textsuperscript{114,120}. Longer-term, the ability to delete variation and distort heritability, two factors influential of selection, 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 in virtually any organism (Table 1). 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 suit other tasks such as multiplexed DNA editing (e.g. Cas12a\textsuperscript{14,18}) or the localization of enzymatic activities to RNAs (e.g. Cas13\textsuperscript{121}).

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 remain. Therefore, if maximum benefit is to be realized from this technology, future studies must 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**

G.A.M. and E.H. wrote the initial draft, which T.E.G. then edited. G.A.M. and T.E.G. created the figures. T.E.G. supervised the work.

**Declaration of Interest**

None.
References


17. Multiplex enCas12a screens show functional buffering by paralogs is systematically absent from genome-wide CRISPR/Cas9 knockout screens | *bioRxiv*. Accessed September 20, 2020. https://www.biorxiv.org/content/10.1101/2020.05.18.102764v1


Table 1: Organisms and key cell types targeted using CRISPR-Cas9 systems.

<table>
<thead>
<tr>
<th>System</th>
<th>Target</th>
<th>Notes</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FnCas9</td>
<td>- Mouse (kidney cells*)</td>
<td>- Comparable indel formation to SpCas9 with little to no off-target activity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Human (PBMCs*, kidney cells*, liver carcinoma cells*)</td>
<td>- Hirano et al. found it only worked when provided as an RNP complex</td>
<td>49,59,122</td>
</tr>
<tr>
<td>SpCas9</td>
<td>- Human (embryonic kidney cells*) - Mouse (embryonic stem cells)</td>
<td>- First CRISPR-Cas9 system to be used in vivo</td>
<td>123–135</td>
</tr>
<tr>
<td></td>
<td>- Rat (one-cell stage embryos) - Zebrafish (one-cell stage embryos)</td>
<td>- Has been used to edit genomes of a broad variety of organisms across most kingdoms of life</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Drosophila melanogaster (embryonic cells)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Arabidopsis thaliana (one-cell stage embryos)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Liverwort (gametophytes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Caenorhabditis elegans (germline syncytia)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Yeast (Saccharomyces cerevisiae, Pichia pastoris, …)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Gram-positive/negative bacteria (Escherichia coli, Streptomyces lividans, Streptococcus pneumoniae, Bacillus subtilis, …)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CjCas9</td>
<td>- Mouse (retinal cells, muscle cells, pancreatic cells)</td>
<td>- Comparable indel formation to SpCas9 and no off-target activity</td>
<td>52,136–138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- No signs of toxicity 14 months after editing</td>
<td></td>
</tr>
<tr>
<td>SaCas9</td>
<td>- Mouse (hepatic cells, embryo fibroblasts)</td>
<td>- No observable off-target activity at candidate sites (mouse and human)</td>
<td>139–141</td>
</tr>
<tr>
<td></td>
<td>- Human (embryonic kidney cells*) - Arabidopsis thaliana</td>
<td>- No signs of toxicity 1-month post manipulation (mouse and human)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- SaCas9 gave more DSB induction than SpCas9 in Arabidopsis</td>
<td></td>
</tr>
<tr>
<td>SpCas9 Nickase</td>
<td>- Human (HeLa cells*)</td>
<td>- Can be used in a 'paired nickase' approach for increased targeting specificity</td>
<td>69,142,143</td>
</tr>
<tr>
<td></td>
<td>- Brown Norway rat (midbrain neurons) - Arabidopsis thaliana</td>
<td>- Used in many more studies in more complex systems e.g. base editors</td>
<td></td>
</tr>
<tr>
<td>SpCas9 RFN</td>
<td>- Human (osteosarcoma cells*, embryonic kidney cells*)</td>
<td>- Less off-target cleavage than wild-type (WT) SpCas9</td>
<td>82,84,144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Has limited target sites due to extra requirements</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Greater specificity than the paired nickase approach</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Targets</td>
<td>( \text{Off-target sites}^* )</td>
<td>( \text{Efficiency} )</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>SaCas9 RFN</td>
<td>- Human (embryonic kidney-GFP cells, embryonic stem cells)</td>
<td>- More restrictive requirements than SpCas9 RFNs, but different PAM required so different target sites available</td>
<td>- Can be paired with a SpCas9 RFN monomer for a heterodimer, higher efficiency than SaCas9 RFN dimer</td>
</tr>
<tr>
<td>SpCas9-HF1</td>
<td>- Human (osteosarcoma cells, embryonic kidney cells) - Potato (protoplasts) - Chicken (embryo fibroblasts)</td>
<td>- 70% of WT SpCas9's target sites were targeted by SpCas9-HF1</td>
<td>- No activity at the off-target sites where WT SpCas9 was active</td>
</tr>
<tr>
<td>evoCas9</td>
<td>- Saccharomyces cerevisiae - Human (embryonic kidney cells)</td>
<td>- Higher targeting efficiency than WT SpCas9</td>
<td>- Significantly more on-target cleavage than SpCas9-HF1</td>
</tr>
<tr>
<td>VQR/VRE R SpCas9</td>
<td>- Zebrafish (one-cell stage embryos) - Human (osteosarcoma cells) - E. coli - C. elegans - Rice - Arabidopsis thaliana</td>
<td>- VQR targets 5'-NGAN-3' and 5'-NGCG-3' PAMs, VRER 5'-NGCG-3'</td>
<td>- Both variants could target sites which WT SpCas9 cannot</td>
</tr>
<tr>
<td>xCas9(3.7)</td>
<td>- Human (embryonic kidney cells) - Rice</td>
<td>- Targets 5'-NG-3', 5'-NNG-3', 5'- GAA-3', 5'-GAT -3' and 5'-CAA-3' PAMs</td>
<td>- Targets 5'-NGG-3' PAMs with higher efficiency than WT SpCas9</td>
</tr>
<tr>
<td>SpCas9-BE1, BE2, BE3, BE4</td>
<td>- Human (embryonic kidney cells) - Rabbit (blastocysts) - Sheep (one-cell stage embryos) - Xenopus laevis (one-cell stage embryos) - Xenopus tropicalis (one-cell stage embryos) - Silkworm (embryonic cells)</td>
<td>- 300-900 human genetic diseases are potential targets for correction via base editing</td>
<td>- BE3 had the best editing yield of BE1, BE2 and BE3</td>
</tr>
<tr>
<td>SaCas9-BE3</td>
<td>- Human (embryonic kidney cells, osteosarcoma cells)</td>
<td>- Can target sites not accessible to SpCas9-BE3</td>
<td></td>
</tr>
<tr>
<td>SaKKH-BE3</td>
<td>- Human (embryonic kidney cells, osteosarcoma cells)</td>
<td>- Targets 5'-NNNRRT-3' PAMs</td>
<td>- Higher efficiency of on-target editing than EQR-BE3 and VQR-BE3</td>
</tr>
<tr>
<td>EQR-BE3</td>
<td>- Human (embryonic kidney cells, osteosarcoma cells)</td>
<td>- Targets 5'-NGAG-3' PAMs</td>
<td>- Less off-target activity than SaBE3 and SaKKH-BE3</td>
</tr>
<tr>
<td>VQR-BE3</td>
<td>- Human (embryonic kidney cells, osteosarcoma cells)</td>
<td>- Targets 5'-NGAN-3' PAMs</td>
<td>- Less off-target activity than SaBE3 and SaKKH-BE3</td>
</tr>
</tbody>
</table>
| SpCas9 PE1, PE2, PE3 | - Human (embryonic kidney cells, osteosarcoma cells, leukemic bone marrow cells, HeLa cells, iPSCs)  
- Mouse (neuro-2a cells)  
- Rice (protoplasts)  
- Wheat (protoplasts) | - 75,000 pathogenic genetic variants diseases are potential targets for correction via prime editing  
- Can perform insertions, deletions, all base conversions and combinations of these | 11,158–162 |

* Application of CRISPR-Cas9 system only shown in vitro.
Figure 1: Core components of the CRISPR-Cas9 system. (A) In the native system a CRISPR RNA (crRNA; pink) and trans-activating crRNA (tracrRNA; blue), bind together (yellow interactions) to form a guide RNA (gRNA) that then complexes with the Cas9 protein (grey). The 5’ end of each RNA is denoted by a small circle. (B) A single guide RNA (sgRNA; green) is produced by fusing a crRNA and tracrRNA using a short linker. This creates a CRISPR-Cas9 system requiring only two components: Cas9 and an sgRNA. (C) Function of the CRISPR-Cas9 system. The Cas9:gRNA complex is able to bind DNA and upon recognition of a complementary sequence to the DNA binding region of the gRNA, double-strand cleavage occurs. Where possible, the cell will attempt to repair this break, which can potentially introduce mutations (cyan bases). Red DNA bases show the protospacer adjacent motif (PAM) and red stars denote DNA cleavage.
Figure 2: 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). Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting, CTD = C-terminal domain, WED = wedge. HNH and RuvC are nuclease domains. Bottom diagrams show the gRNA structure with the DNA binding region, major stem loops (SLs) and repeat:anti-repeat (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 shown by black regions and REC domains are numbered.
Figure 3: 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. HNH and RuvC are nuclease domains. Thick red outlines indicate domains which have been modified.
Figure 4: 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. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. HNH and RuvC are nuclease domains. (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 red lines and the three REC domains are numbered. (D) Example of the Cas9D10A nickase system targeting two regions to create complementary nicks on opposite strands. The PAM is shown in red and red stars denote DNA cleavage.
Figure 5: 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 abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. HNH and RuvC are nuclease domains. 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 black and white regions, respectively, and mutations are shown by red lines. The three REC domains are numbered. (C) Two RFNs bound in an active conformation to a target DNA locus. Red stars denote DNA cleavage. (D) Single inactive RFN bound to an off-target DNA locus.
**Figure 6: Base editing systems.** (A) Base editor 1 (BE1) consists of a SpdCas9 with a cytidine deaminase (rAPOBEC1) fused to its N-terminus. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. HNH and RuvC are nuclease domains. (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 black and white regions, respectively, and mutations are shown by red lines. The three REC domains are numbered.
Figure 7: 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. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. HNH and RuvC are nuclease domains. Linkers are denoted by black 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. This second gRNA for PE3 is not a pegRNA and does not contain any modification. (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). Major stem loops (SLs), linker and repeat:anti-repeat (R:AR) duplex are also labeled. 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 black and white regions, respectively, and mutations are shown by red lines. The three REC domains are numbered.