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

Grace A. Meaker[1,2], Emma J. Hair[1] 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 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*[19]. It consists of a Cas nuclease *Sp*Cas9 and a gRNA[8] (**Figure 1**). The gRNA has two components – a trans-activating RNA (tracrRNA) and a CRISPR RNA (crRNA)[6] (**Figure 1A**). crRNA is responsible for recognition and binding of the target DNA region and tracrRNA for crRNA maturation and association with *Sp*Cas9. Alternatively, a chimeric single-guide RNA (sgRNA) which performs both these functions can be used[6] (**Figure 1B**). Once the gRNA binds the *Sp*Cas9, the *Sp*Cas9 undergoes a conformational change which permits the *Sp*Cas9-crRNA-tracrRNA complex to relocate to the target region and cleave both DNA strands[7]. The target region is determined by a 20-nucleotide 'spacer' in the crRNA, complementary to the target 'protospacer' in the DNA[3,20]. 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 *Sp*Cas9 it is '5-NGG-3'[8,21]. 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[22,23]. Once the complex is bound, the two nucleases produce a double-stranded break (DSB) 3–4 nucleotides (nt) upstream of the PAM[24]. 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[25] (**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[25,26]. 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)[27]. 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[28]. 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[21,24,25]. These essential 13 nt have been dubbed the 'seed sequence'[8,29]. 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[30]. 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[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[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[35]. Cleavage by *Sp*Cas9 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'[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[38,39]. These approaches, however, do not consider off-target sites which do not fit the model's parameters[40]. To alleviate this issue, machine learning methods have recently been shown to offer improved performance[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[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[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[40,42]. 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 fragments[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[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, Digenome-seq is appropriate for *in vitro* applications because it is not vulnerable to chromatin arrangements[43], but for *in vivo* applications, GUIDE-seq or the new, multiplexing sister method Tagmentation-based tag integration site sequencing (TTISS) are more sensitive and easier to use[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[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[47]. Type I and III systems enlist multiple Cas proteins whereas type II uses a single, Cas9 protein for DNA cleavage[48]. Whilst *Sp*Cas9 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 (**Figure 2**). This includes *Fn*Cas9 from *Francisella novicida*[49], *Sa*Cas9 from *Staphylococcus aureus*[50,51] and recently the *Campylobacter jejuni* Cas9, the smallest to date[52,53].

*Sp*Cas9 is a multi-domain protein exhibiting a bilobed structure where the nuclease lobe and the recognition lobe[8,24] are linked by an arginine-rich bridge helix as well as a disordered linker[8] (**Figure 2A**). The overall shape of *Sp*Cas9 is oblong with two large grooves, to accommodate the DNA:RNA and RNA:RNA complexes. Adaptations of the two previously-recognized, adjacent nucleases (HNH[6], named for the three characteristic residues, and RuvC[54]) of the nuclease lobe facilitate much of the diversification of CRISPR's function[31,55]. Each nuclease cleaves one strand of DNA; RuvC cleaves the non-complementary and HNH the complementary strand[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[7]. Mutagenesis of these domains permits the evolution of CRISPR function.

*Sa*Cas9 has a longer PAM (5'-NNGRRT-3') than *Sp*Cas9 and is smaller at 1053 amino acids (aa) compared to 1368 aa[50] (**Figure 2B**). Due to its smaller size, *Sa*Cas9 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 *Sa*Cas9 has shown comparable on-target cleavage to *Sp*Cas9, whilst boasting a higher specificity and easier introduction into cells[56]. Both *Sp*Cas9 and *Sa*Cas9 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, *Sa*Cas9 only has 17% structural similarity to *Sp*Cas9; 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[56]. Despite these differences, it is apparent that *Sa*Cas9 and *Sp*Cas9 share important similarities, and that

*Sa*Cas9 is a useful case study for synthetic reduction of *Sp*Cas9 size and complexity, already attempted by the successful removal of the REC2 domain[57].

Another *Sp*Cas9 ortholog is *Fn*Cas9 which produces staggered cleavage and binds less frequently to non-target regions[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 *Sp*Cas9 and *Sa*Cas9, producing overhangs of up to 4 nt and more efficient recruitment of HDR[49]. *Fn*Cas9 is considerably larger than *Sp*Cas9 and *Sa*Ca9, comprised of 1629 aa[59]. Whilst its larger size may be a hindrance for transfection due to the limited capacity of many delivery systems, *Fn*Cas9's markedly reduced tolerance of target mismatches makes it a valuable system for precise editing tasks. *Sp*Cas9 tolerates several mismatches of the gRNA in the non-seed region, but just one mismatch at the 5' end of *Fn*Cas9 gRNA is tolerated for successful cleavage[58]. This increased specificity means *Fn*Cas9 produces far less off-target cleavage as fewer sites are recognized as 'target'[49]. *Fn*Cas9 is structurally dissimilar to *Sp*Cas9 and *Sa*Cas9, lacking bilobed structure and containing distinct REC2 and REC3 domains (**Figure 2C**). REC3 domain mutations have generated high-fidelity Cas9 enzymes[60]; these structural differences explain the striking differences in targeting specificity. Despite its increased specificity, it has much lower on-target recognition than *Sp*Cas9 in eukaryotic genomes. As postulated in the literature[58], local chromatin conformations likely affect the access to DNA, a vulnerability not as significant for *Sp*Cas9. To eliminate this problem *Fn*Cas9 has been used alongside a catalytically dead *Sp*Cas9 (*Sp*dCas9) to enable access and subsequent DNA cleavage[58]. Such problems are not present when used in prokaryotes where *Fn*Cas9 has been shown to function effectively[61].

Finally, *Cj*Cas9 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 *Sa*Cas9 and *Sp*Cas9, with a simplified REC lobe and size-reduced NUC lobe[53] (**Figure 2D**). Initial studies showed recognition of a 5'-NNNNACA-3' PAM[47] or the more promiscuous 5'-NNNVRYM-3'[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'[52] and 5'-NNNNACAC-3' sequences[62]. Tested against *Sa*Cas9 in human cells, *Cj*Cas9 was found to be more specific with comparable efficiencies to some other variants, excluding *Fn*Cas9[52]. However, due to discrepancies in the PAM recognition sequences and limited research into the structure and mechanism of *Cj*Cas9, 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[6,50]. In contrast, removal of loops 2 or 3 decreases efficiency, without abolishing cleavage[24]. Stem loop 2 interacts with the PI and RuvC domains in *Sa*Cas9 and *Sp*Cas9, and the REC domains in *Fn*Cas9 and *Cj*Cas9[7,50,53,57,59]. *Sa*Cas9 and *Sp*Cas9's sgRNAs exhibit the greatest similarity, particularly regarding cognate Cas9 interactions with the lack of stem loop 3 in *Sa*Cas9 the defining key difference[50]. This further highlights the minimalism of *Sa*Cas9 compared to *Sp*Cas9 because of the reduction of non-essential elements like stem loop 3 and the REC2 domain[56]. *Fn*Cas9 and *Cj*Cas9's sgRNAs are structurally distinct to *Sa*Cas9 and *Sp*Cas9, with the same core region but some unique features. For instance, *Fn*Cas9 has a longer, U shaped linker, contrasting with the shorter, single-stranded linker present in *Sa*Cas9 and *Sp*Cas9[59]. The novel structural arrangement of *Cj*Cas9's gRNA forms a triple helix between stem loops 1, 2, and 3[53]. The relevance of this structure is still unknown due to a lack of comprehensive structural studies of *Cj*Cas9 complexes.

The domains of each Cas9 distinctly interact with their associated sgRNAs due to the slight differences in sgRNA structure[50] (**Figure 2**). The stark differences between *Sp*Cas9 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 *Sp*Cas9 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[6]. 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[6]. 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[39,63]. 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[33]. Such estimates are supported by the finding that additional nucleotides added at the 5' end of gRNA increase binding affinity for off-target sites[28]. 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)[64]. 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[64]. 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[65].

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[6,20]. 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[27]. 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)[66,67]. 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[20,25,31]. The benefits of these are twofold: they produce precise nicks in the DNA and exhibit decreased affinity to off-target loci[31]. 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**)[68]. 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[69]. 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 decease mosaicism in monkey embryos [70]. 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[71].

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 *Sp*Cas9 enzyme containing the H840A and D10A mutations is catalytically dead (dCas9)[6,72], 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[9] or used to visualize target affinity without cleavage[55]. 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[73] 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[74].

Furthermore, dCas9 has become widely used in regulating gene expression through CRISPR interference and activation (CRISPRi and CRISPRa, respectively)[75,76]. Interference of gene expression is generally achieved by targeting the dCas9 protein to promoter regions and sterically blocking the initiation of RNA polymerase (RNAP)[77]. Additional, repression domains (e.g. KRAB) can also be fused to the dCas9 to enhance repression[78]. 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)[79]. 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[77,80], 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[81]. 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[82]. This fusion enlists a long, flexible linker with between 5-25 residues (e.g. $(GGGGS)_5$ fusing the FokI

endonuclease to the Cas9 N-terminus[82–84]. The RNA-guided FokI Nuclease (RFN) system consists of a dCas9-FokI fusion and two different gRNAs[85]. 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[82] (**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[84]. However, RFNs are limited for genome-wide application due to the required presence of PAM sequences either side of the protospacer regions (5'-$CCNN_{20}$-3' and 5'-N_{20}NGG-3') as well as 14–17 bp between these[83]. This fusion system is also very large, limiting its application in AAV delivery methods[86]. Efforts have been made to use the smaller *Sa*Cas9 based system instead of *Sp*Cas9, reducing the size and simplifying delivery[83]. Despite some documented successes[87,88], 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[89]. In contrast, Cas9 ribonucleoproteins are limited by transient expression and possible reduced on-target activity[90].

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[91]. Using this knowledge, a high fidelity Cas9, *Sp*Cas9-HF1 was produced via mutation of four DNA-interacting REC3 residues to alanine (N497A/R661A/Q695A/Q926A), with comparable on-target cleavage to *Sp*Cas9[32]. 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 *Sp*Cas9-HF1 led to further screening of REC3 mutants *in vivo* and the development of another highly-specific *Sp*Cas9 variant, dubbed 'evoCas9'[60]. This variant outperforms *Sp*Cas9-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, *Sp*Cas9 is still limited to targeting of genomic regions containing the 5'-NGG-3' PAM[6], whose number may be further reduced by local chromatin or

methylation patterns preventing Cas9 access to the site[25]. PAM specificity is conferred by several residues of the PI domain, specifically *Sp*Cas9 arginine residues 1333 and 1335 which interact with the two guanine nucleotides of the PAM[7]. 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[92]. 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[51]. 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 *Sp*Cas9[51]. These evolved *Sp*Cas9 variants with altered PAM specificities are still limited to one or two PAMs.

To expand PAM recognition, focus has shifted to generating *Sp*Cas9 variants able to target multiple PAMs. So far, positive selection has been used to find useful mutagenized *Sp*Cas9 variants using phage assisted continuous evolution (PACE)[21]. Such variants, dubbed 'xCas9' nucleases, had a different pattern of mutations than the rationally developed variants which covered the entire *cas9* gene[7,51]. 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 *Sp*Cas9[21]. Together with the broader on-target specificity, xCas9-3.7 produced less off-target cleavage than *Sp*Cas9, 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 *Sp*Cas9 and has been performed in *Sa*Cas9 to similar effect. Using an analogous bacterial selection approach, mutated *Sa*Cas9 variants were tested for their efficiency for 5'-NNNRRT-3' PAM loci cleavage. Results showed that an E782K/N968K/R1015H variant called *Sa*KKH was functional and that this variant disrupted 5'-NNGRRT-3' sites (and off-target loci) at a similar efficiency to wild-type *Sa*Cas9 whilst also cleaving sites adjacent to 5'-NNARRT-3', 5'-NNTRRT-3' and 5'-NNCRRT-3'[93].

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[27]. 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[94]. Single-stranded nicks are repaired by the high-fidelity BER pathway, making this cleavage pattern preferable for specific base changes[67].

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[95] (**Figure 6**). A simple CRISPR base editor consists of a dCas9 protein, a sgRNA and a base editing enzyme (e.g. cytidine deaminase)[96]. Cytidine deaminases catalyze the conversion of cytosine to uracil[97] 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[84,98] (**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[72]. This discrepancy was attributed to the endogenous DNA repair machinery, specifically uracil DNA glycosylase (UDG), which reverses the UG pair to a CG pair[72]. 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[72]. 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[99]. 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 *Sa*Cas9 nickase instead of *Sp*Cas9 in a BE3 variant, *Sa*BE3[100]. As previously described, *Sa*Cas9 is much smaller than *Sp*Cas9[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. *Sa*BE3 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 *Sp*Cas9-based BE3[72,100]. Furthermore, Kim and colleagues utilized *Sp*Cas9 variants with altered PAM specificities, specifically VQR and VRER (described previously) and EQR from the same study[51], as well as an engineered *Sa*Cas9 variant, *Sa*KKH[93]. All these variants had editing efficiencies of up to 50% for sites with relevant PAMs, with *Sa*KKH-BE3 editing up to 62% of target sites. *Sa*BE3 and *Sa*KKH-BE3 had a similar off-target activity to *Sp*Cas9 whereas EQR-BE3 and VQR-BE3 showed markedly reduced levels[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)[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%[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[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[101]. However, many existing methods have differing sensitivities[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[103,104]. A further complication concerns the disagreements between *in vitro* and *in vivo* results, which have been reported even for some of the most robust methods developed[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.

Limitations in CRISPR research

Another factor hampering our understanding and comparison of CRISPR-Cas9 systems is the lack of standardised studies and benchmarking[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. *in vivo/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[106] and defined microbial communities to test metagenomic inference from mixed pools of organisms[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, *p53*[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[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[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[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[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[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[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[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[119]. What form this might take has yet to be fully understood but has drawn recent attention[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 suited to other tasks such as multiplexed DNA editing (e.g. Cas12a[14,18]) or the localization of enzymatic activities to RNAs (e.g. Cas13[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

1. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. *Nucleotide Sequence of the Iap 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://jb.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. Sapranauskas 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. 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

15. Bin Moon S, Lee JM, Kang JG, et al. Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3′-overhang. *Nat Commun*. 2018;9(1):3651. doi:10.1038/s41467-018-06129-w

16. Sanson KR, DeWeirdt PC, Sangree AK, et al. Optimization of AsCas12a for combinatorial genetic screens in human cells. *bioRxiv*. Published online August 28, 2019:747170. doi:10.1101/747170

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

18. 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

19. 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

20. 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

21. 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

22. 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

23. 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

24. Doudna JA, Jiang F. CRISPR–Cas9 Structures and Mechanisms. *Annu Rev Biophys*. 2018;47(1):19-20. doi:10.1146/annurev-biophys

25. 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

26. 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

27. 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

28. 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

29. 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

30. 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

31. 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

32. 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

33. 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

34. Gkazi SA. Quantifying CRISPR off-target effects. *Emerg Top Life Sci*. 2019;3(3):327-334. doi:10.1042/ETLS20180146

35. 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

36. 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

37. 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

38. 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

39. 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

40. 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

41. Listgarten J, Weinstein M, Kleinstiver BP, et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng*. 2018;2(1):38-47. doi:10.1038/s41551-017-0178-6

42. 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

43. 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

44. Schmid-Burgk JL, Gao L, Li D, et al. Highly Parallel Profiling of Cas9 Variant Specificity. *Mol Cell*. 2020;78(4):794-800.e8. doi:10.1016/j.molcel.2020.02.023

45. Listgarten J, Weinstein M, Kleinstiver BP, et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat Biomed Eng*. 2018;2(1):38-47. doi:10.1038/s41551-017-0178-6

46. Lee SH, Yu J, Hwang G-H, et al. CUT-PCR: CRISPR-mediated, ultrasensitive detection of target DNA using PCR. *Oncogene*. 2017;36(49):6823-6829. doi:10.1038/onc.2017.281

47. 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

48. 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

49. 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

50. 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

51. 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

52. 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

53. 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

54. 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.

55. 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

56. 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

57. 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

58. 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

59. 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

60. 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

61. 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

62. 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

63. 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

64. Hendel A, Bak RO, Clark JT, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol*. 2015;33(9):985-989. doi:10.1038/nbt.3290

65. Khvorova A, Watts JK. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol*. 2017;35(3):238-248. doi:10.1038/nbt.3765

66. 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

67. Dianov GL, Hübscher U. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res*. 2013;41(6):3483-3490. doi:10.1093/nar/gkt076

68. 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

69. 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

70. Tu Z, Yang W, Yan S, et al. Promoting Cas9 degradation reduces mosaic mutations in non-human primate embryos. *Sci Rep*. 2017;7(1):42081. doi:10.1038/srep42081

71. Yang S, Li S, Li X-J. Shortening the Half-Life of Cas9 Maintains Its Gene Editing Ability and Reduces Neuronal Toxicity. *Cell Rep*. 2018;25(10):2653-2659.e3. doi:10.1016/j.celrep.2018.11.019

72. 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

73. McDonald JI, Celik H, Rois LE, et al. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. *Biol Open*. 2016;5(6):866-874. doi:10.1242/bio.019067

74. Chen B, Gilbert LA, Cimini BA, et al. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*. 2013;155(7):1479-1491. doi:10.1016/j.cell.2013.12.001

75. 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

76. 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

77. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res*. 2013;41(15):7429-7437. doi:10.1093/nar/gkt520

78. Yeo NC, Chavez A, Lance-Byrne A, et al. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat Methods*. 2018;15(8):611-616. doi:10.1038/s41592-018-0048-5

79. McInally 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

80. Perez-Pinera P, Kocak DD, Vockley CM, et al. RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat Methods*. 2013;10(10):973-976. doi:10.1038/nmeth.2600

81. Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nat Commun*. 2018;9(1):2489. doi:10.1038/s41467-018-04901-6

82. 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

83. 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

84. 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

85. 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

86. 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

87. Liang Z, Chen K, Li T, et al. Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat Commun*. 2017;8. doi:10.1038/ncomms14261

88. Chen S, Lee B, Lee AY-F, Modzelewski AJ, He L. Highly Efficient Mouse Genome Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *J Biol Chem*. 2016;291(28):14457-14467. doi:10.1074/jbc.M116.733154

89. Lyu P, Javidi-Parsijani P, Atala A, Lu B. Delivering Cas9/sgRNA ribonucleoprotein (RNP) by lentiviral capsid-based bionanoparticles for efficient “hit-and-run” genome editing. *Nucleic Acids Res*. 2019;47(17):e99. doi:10.1093/nar/gkz605

90. Vakulskas CA, Dever DP, Rettig GR, et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med*. 2018;24(8):1216-1224. doi:10.1038/s41591-018-0137-0

91. 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

92. 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

93. 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

94. 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

95. 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

96. 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

97. Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol*. 2008;9(6):229. doi:10.1186/gb-2008-9-6-229

98. 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

99. 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

100. 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

101. 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

102. 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

103. 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

104. 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

105. 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

106. Baker SC, Bauer SR, Beyer RP, et al. The External RNA Controls Consortium: a progress report. *Nat Methods*. 2005;2(10):731-734. doi:10.1038/nmeth1005-731

107. Greenwald WW, Klitgord N, Seguritan V, et al. Utilization of defined microbial communities enables effective evaluation of meta-genomic assemblies. *BMC Genomics*. 2017;18(1):296. doi:10.1186/s12864-017-3679-5

108. 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

109. 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

110. 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

111. 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

112. 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

113. Cyranoski D, Ledford H. Genome-edited baby claim provokes international outcry. *Nature*. 2018;563(7733):607-608. doi:10.1038/d41586-018-07545-0

114. Li T, Shen X. Pleiotropy Complicates Human Gene Editing: CCR5Δ32 and Beyond. *Front Genet*. 2019;10. doi:10.3389/fgene.2019.00669

115. 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

116. 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

117. 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

118. Omodamilola OI. CRISPR Technology; Advantages, Limitations and Future Direction. 2018;1(2):5.

119. Peters T. Should CRISPR Scientists Play God? *Religions*. 2017;8(4):61. doi:10.3390/rel8040061

120. Quammen D. *The Tangled Tree*. Simon & Schuster; 2019. Accessed June 18, 2020. https://www.simonandschuster.com/books/The-Tangled-Tree/David-Quammen/9781476776637

121. Cox DBT, Gootenberg JS, Abudayyeh OO, et al. RNA editing with CRISPR-Cas13. *Science*. 2017;358(6366):1019-1027. doi:10.1126/science.aaq0180

122. Price AA, Sampson TR, Ratner HK, Grakoui A, Weiss DS. Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proc Natl Acad Sci U S A*. 2015;112(19):6164-6169. doi:10.1073/pnas.1422340112

123. LeBlanc C, Zhang F, Mendez J, et al. Increased efficiency of targeted mutagenesis by CRISPR/Cas9 in plants using heat stress. *Plant J*. 2018;93(2):377-386. doi:10.1111/tpj.13782

124. Wang H, Yang H, Shivalila CS, et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013;153(4):910-918. doi:10.1016/j.cell.2013.04.025

125. Hwang WY, Fu Y, Reyon D, et al. Efficient In Vivo Genome Editing Using RNA-Guided Nucleases. *Nat Biotechnol*. 2013;31(3):227-229. doi:10.1038/nbt.2501

126. Ma Y, Shen B, Zhang X, et al. Heritable Multiplex Genetic Engineering in Rats Using CRISPR/Cas9. *PLOS ONE*. 2014;9(3):e89413. doi:10.1371/journal.pone.0089413

127. Wang Z-P, Xing H-L, Dong L, et al. Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. *Genome Biol*. 2015;16(1):144. doi:10.1186/s13059-015-0715-0

128. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. RNA-programmed genome editing in human cells. *eLife*. 2013;2. doi:10.7554/eLife.00471

129. Yu Z, Chen H, Liu J, et al. Various applications of TALEN- and CRISPR/Cas9-mediated homologous recombination to modify the Drosophila genome. *Biol Open*. 2014;3(4):271-280. doi:10.1242/bio.20147682

130. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res*. 2013;41(7):4336-4343. doi:10.1093/nar/gkt135

131. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. CRISPR-assisted editing of bacterial genomes. *Nat Biotechnol*. 2013;31(3):233-239. doi:10.1038/nbt.2508

132. Sugano SS, Shirakawa M, Takagi J, et al. CRISPR/Cas9-Mediated Targeted Mutagenesis in the Liverwort Marchantia polymorpha L. *Plant Cell Physiol*. 2014;55(3):475-481. doi:10.1093/pcp/pcu014

133. Chiu H, Schwartz HT, Antoshechkin I, Sternberg PW. Transgene-Free Genome Editing in Caenorhabditis elegans Using CRISPR-Cas. *Genetics*. 2013;195(3):1167-1171. doi:10.1534/genetics.113.155879

134. Cobb RE, Wang Y, Zhao H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth Biol*. 2015;4(6):723-728. doi:10.1021/sb500351f

135. Westbrook AW, Moo-Young M, Chou CP. Development of a CRISPR-Cas9 Tool Kit for Comprehensive Engineering of Bacillus subtilis. *Appl Environ Microbiol*. 2016;82(16):4876-4895. doi:10.1128/AEM.01159-16

136. Koo T, Lu-Nguyen NB, Malerba A, et al. Functional Rescue of Dystrophin Deficiency in Mice Caused by Frameshift Mutations Using Campylobacter jejuni Cas9. *Mol Ther*. 2018;26(6):1529-1538. doi:10.1016/j.ymthe.2018.03.018

137. Jo DH, Koo T, Cho CS, Kim JH, Kim J-S, Kim JH. Long-Term Effects of In Vivo Genome Editing in the Mouse Retina Using Campylobacter jejuni Cas9 Expressed via Adeno-Associated Virus. *Mol Ther*. 2019;27(1):130-136. doi:10.1016/j.ymthe.2018.10.009

138. Chang YJ, Bae J, Zhao Y, et al. In vivo multiplex gene targeting with Streptococcus pyogens and Campylobacter jejuni Cas9 for pancreatic cancer modeling in wild-type animal. *J Vet Sci*. 2020;21(2). doi:10.4142/jvs.2020.21.e26

139. Ran FA, Cong L, Yan WX, et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature*. 2015;520(7546):186-191. doi:10.1038/nature14299

140. Wolter F, Klemm J, Puchta H. Efficient in planta gene targeting in Arabidopsis using egg cell-specific expression of the Cas9 nuclease of Staphylococcus aureus. *Plant J*. 2018;94(4):735-746. doi:10.1111/tpj.13893

141. Kaminski R, Bella R, Yin C, et al. Excision of HIV-1 DNA by Gene Editing: A Proof-of-Concept In Vivo Study. *Gene Ther*. 2016;23(8-9):690-695. doi:10.1038/gt.2016.41

142. Bäck S, Necarsulmer J, Whitaker LR, et al. Neuron-Specific Genome Modification in the Adult Rat Brain Using CRISPR-Cas9 Transgenic Rats. *Neuron*. 2019;102(1):105-119.e8. doi:10.1016/j.neuron.2019.01.035

143. Schiml S, Fauser F, Puchta H. The CRISPR/Cas system can be used as nuclease for in planta gene targeting and as paired nickases for directed mutagenesis in Arabidopsis resulting in heritable progeny. *Plant J*. 2014;80(6):1139-1150. doi:10.1111/tpj.12704

144. Wyvekens N, Topkar VV, Khayter C, Joung JK, Tsai SQ. Dimeric CRISPR RNA-Guided FokI-dCas9 Nucleases Directed by Truncated gRNAs for Highly Specific Genome Editing. *Hum Gene Ther*. 2015;26(7):425-431. doi:10.1089/hum.2015.084

145. Nadakuduti SS, Starker CG, Ko DK, et al. Evaluation of Methods to Assess in vivo Activity of Engineered Genome-Editing Nucleases in Protoplasts. *Front Plant Sci*. 2019;10. doi:10.3389/fpls.2019.00110

146. Challagulla A, Jenkins KA, O'Neil TE, et al. Germline engineering of the chicken genome using CRISPR/Cas9 by in vivo transfection of PGCs. *Anim Biotechnol*. 2020;0(0):1-10. doi:10.1080/10495398.2020.1789869

147. Kim N, Kim HK, Lee S, et al. Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat Biotechnol*. Published online June 8, 2020:1-9. doi:10.1038/s41587-020-0537-9

148. Bell RT, Fu BXH, Fire AZ. Cas9 Variants Expand the Target Repertoire in Caenorhabditis elegans. *Genetics*. 2016;202(2):381-388. doi:10.1534/genetics.115.185041

149. Hu X, Meng X, Liu Q, Li J, Wang K. Increasing the efficiency of CRISPR-Cas9-VQR precise genome editing in rice. *Plant Biotechnol J*. 2018;16(1):292-297. doi:10.1111/pbi.12771

150. Yamamoto A, Ishida T, Yoshimura M, Kimura Y, Sawa S. Developing Heritable Mutations in Arabidopsis thaliana Using a Modified CRISPR/Cas9 Toolkit Comprising PAM-Altered Cas9 Variants and gRNAs. *Plant Cell Physiol*. 2019;60(10):2255-2262. doi:10.1093/pcp/pcz118

151. Zhong Z, Sretenovic S, Ren Q, et al. Improving Plant Genome Editing with High-Fidelity xCas9 and Non-canonical PAM-Targeting Cas9-NG. *Mol Plant*. 2019;12(7):1027-1036. doi:10.1016/j.molp.2019.03.011

152. Park D-S, Yoon M, Kweon J, Jang A-H, Kim Y, Choi S-C. Targeted Base Editing via RNA-Guided Cytidine Deaminases in Xenopus laevis Embryos. *Mol Cells*. 2017;40(11):823-827. doi:10.14348/molcells.2017.0262

153. Liu Z, Chen M, Chen S, et al. Highly efficient RNA-guided base editing in rabbit. *Nat Commun*. 2018;9. doi:10.1038/s41467-018-05232-2

154. Shi Z, Xin H, Tian D, et al. Modeling human point mutation diseases in Xenopus tropicalis with a modified CRISPR/Cas9 system. *FASEB J*. 2019;33(6):6962-6968. doi:10.1096/fj.201802661R

155. Zhou S, Cai B, He C, et al. Programmable Base Editing of the Sheep Genome Revealed No Genome-Wide Off-Target Mutations. *Front Genet*. 2019;10. doi:10.3389/fgene.2019.00215

156. Liu Y, Li Y, Liang Y, et al. Comparative analysis of genome editing systems, Cas9 and BE3, in silkworms. *Int J Biol Macromol*. 2020;158:486-492. doi:10.1016/j.ijbiomac.2020.04.170

157. Tran MTN, Khalid MKNM, Pébay A, et al. Screening of CRISPR/Cas base editors to target the AMD high-risk Y402H complement factor H variant. *Mol Vis*. 2019;25:174-182.

158. Lin Q, Zong Y, Xue C, et al. Prime genome editing in rice and wheat. *Nat Biotechnol*. 2020;38(5):582-585. doi:10.1038/s41587-020-0455-x

159. Tang X, Sretenovic S, Ren Q, et al. Plant Prime Editors Enable Precise Gene Editing in Rice Cells. *Mol Plant*. 2020;13(5):667-670. doi:10.1016/j.molp.2020.03.010

160. Liu Y, Li X, He S, et al. Efficient generation of mouse models with the prime editing system. *Cell Discov*. 2020;6. doi:10.1038/s41421-020-0165-z

161. Sürün D, Schneider A, Mircetic J, et al. Efficient Generation and Correction of Mutations in Human iPS Cells Utilizing mRNAs of CRISPR Base Editors and Prime Editors. *Genes*. 2020;11(5). doi:10.3390/genes11050511

162. Butt H, Rao GS, Sedeek K, Aman R, Kamel R, Mahfouz M. Engineering herbicide resistance via prime editing in rice. *Plant Biotechnol J*. n/a(n/a). doi:10.1111/pbi.13399

997 # Tables

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

System	Target	Notes	Refs.
*Fn*Cas9	- Mouse (kidney cells*) - Human (PBMCs*, kidney cells*, liver carcinoma cells*)	- Comparable indel formation to *Sp*Cas9 with little to no off-target activity - Hirano et al. found it only worked when provided as an RNP complex	49,59,122
*Sp*Cas9	- Human (embryonic kidney cells*) - Mouse (embryonic stem cells) - Rat (one-cell stage embryos) - Zebrafish (one-cell stage embryos) - *Drosophila melanogaster* (embryonic cells) - *Arabidopsis thaliana* (one-cell stage embryos) - Liverwort (gametophytes) - *Caenorhabditis elegans* (germline syncytia) - Yeast (*Saccharomyces cerevisiae, Pichia pastoris, …)* - Gram-postitve/-negative bacteria *(Escherichia coli, Streptomyces lividans, Streptococcus pneumoniae, Bacillus subtilis, …)*	- First CRISPR-Cas9 system to be used *in vivo* - Has been used to edit genomes of a broad variety of organisms across most kingdoms of life	123–135
*Cj*Cas9	- Mouse (retinal cells, muscle cells, pancreatic cells)	- Comparable indel formation to *Sp*Cas9 and no off-target activity - No signs of toxicity 14 months after editing	52,136–138
*Sa*Cas9	- Mouse (hepatic cells, embryo fibroblasts) - Human (embryonic kidney cells*) - *Arabidopsis thaliana*	- No observable off-target activity at candidate sites (mouse and human) - No signs of toxicity 1-month post manipulation (mouse and human) - *Sa*Cas9 gave more DSB induction than *Sp*Cas9 in *Arabidopsis*	139–141
*Sp*Cas9 Nickase	- Human (HeLa cells*) - Brown Norway rat (midbrain neurons) - *Arabidopsis thaliana*	- Can be used in a 'paired nickase' approach for increased targeting specificity - Used in many more studies in more complex systems e.g. base editors	69,142,143
*Sp*Cas9 RFN	- Human (osteosarcoma cells*, embryonic kidney cells*)	- Less off-target cleavage than wild-type (WT) *Sp*Cas9 - Has limited target sites due to extra requirements - Greater specificity than the paired nickase approach	82,84,144

*Sa*Cas9 RFN	- Human (embryonic kidney-GFP cells*, embryonic stem cells*)	- More restrictive requirements than *Sp*Cas9 RFNs, but different PAM required so different target sites available - Can be paired with a *Sp*Cas9 RFN monomer for a heterodimer, higher efficiency than *Sa*Cas9 RFN dimer	83
*Sp*Cas9-HF1	- Human (osteosarcoma cells* embryonic kidney cells*) - Potato (protoplasts) - Chicken (embryo fibroblasts)	- 70% of WT *Sp*Cas9's target sites were targeted by *Sp*Cas9-HF1 - No activity at the off-target sites where WT *Sp*Cas9 was active	32,145–147
evoCas9	- *Saccharomyces cerevisiae* - Human (embryonic kidney cells*)	- Higher targeting efficiency than WT *Sp*Cas9 - Significantly more on-target cleavage than *Sp*Cas9-HF1 - Both *Sp*Cas9-HF1 and evoCas9 had almost no off-target cleavage, evoCas9 slightly less	60,147
VQR/VRER *Sp*Cas9	- Zebrafish (one-cell stage embryos) - Human (osteosarcoma cells*) - *E. coli* - *C. elegans* - Rice - *Arabidopsis thaliana*	- VQR targets 5'-NGAN-3' and 5'-NGCG-3' PAMs, VRER 5'-NGCG-3' - Both variants could target sites which WT *Sp*Cas9 cannot - VRER showed increased fidelity to WT *Sp*Cas9, possibly because of the 4th PAM base	51,148–150
xCas9(3.7)	- Human (embryonic kidney cells*) - Rice	- Targets 5'-NG-3', 5'-NNG-3', 5'-GAA-3', 5'-GAT -3' and 5'-CAA-3' PAMs - Targets 5'-NGG-3' PAMs with higher efficiency than WT *Sp*Cas9 - Much lower off-target activity than WT *Sp*Cas9 in human cells	21,151
*Sp*Cas9-BE1, BE2, BE3, BE4	- 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)	- 300-900 human genetic diseases are potential targets for correction via base editing - BE3 had the best editing yield of BE1, BE2 and BE3 - BE4 showed higher C to T editing efficiencies, lower indel formation and higher product formation than BE3	72,99,152–156
*Sa*Cas9-BE3	- Human (embryonic kidney cells*, osteosarcoma cells*)	- Can target sites not accessible to *Sp*Cas9-BE3	100,157
*Sa*KKH-BE3	- Human (embryonic kidney cells*, osteosarcoma cells*)	- Targets 5'-NNNRRT-3' PAMs - Higher efficiency of on-target editing than EQR-BE3 and VQR-BE3	100,157
EQR-BE3	- Human (embryonic kidney cells*, osteosarcoma cells*)	- Targets 5'-NGAG-3' PAMs - Less off-target activity than *Sa*BE3 and *Sa*KKH-BE3	100
VQR-BE3	- Human (embryonic kidney cells*, osteosarcoma cells*)	- Targets 5'-NGAN-3' PAMs - Less off-target activity than *Sa*BE3 and *Sa*KKH-BE3	100,157

*Sp*Cas9 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

999 * Application of CRISPR-Cas9 system only shown *in vitro*.

Figures and captions

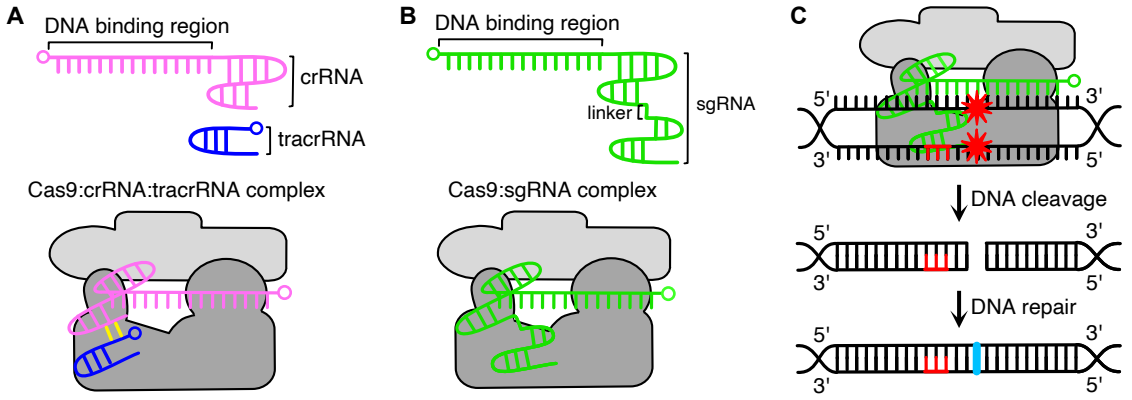

Figure 1: Core components of the CRISPR-Cas9 system. (**A**) In the native system a CRISPR RNA (crRNA; pink) and trans-activating crRNA (tracrRNA; blue), bind together (yellow interactions) to form a guide RNA (gRNA) that then complexes with the Cas9 protein (grey). The 5' end of each RNA is denoted 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.

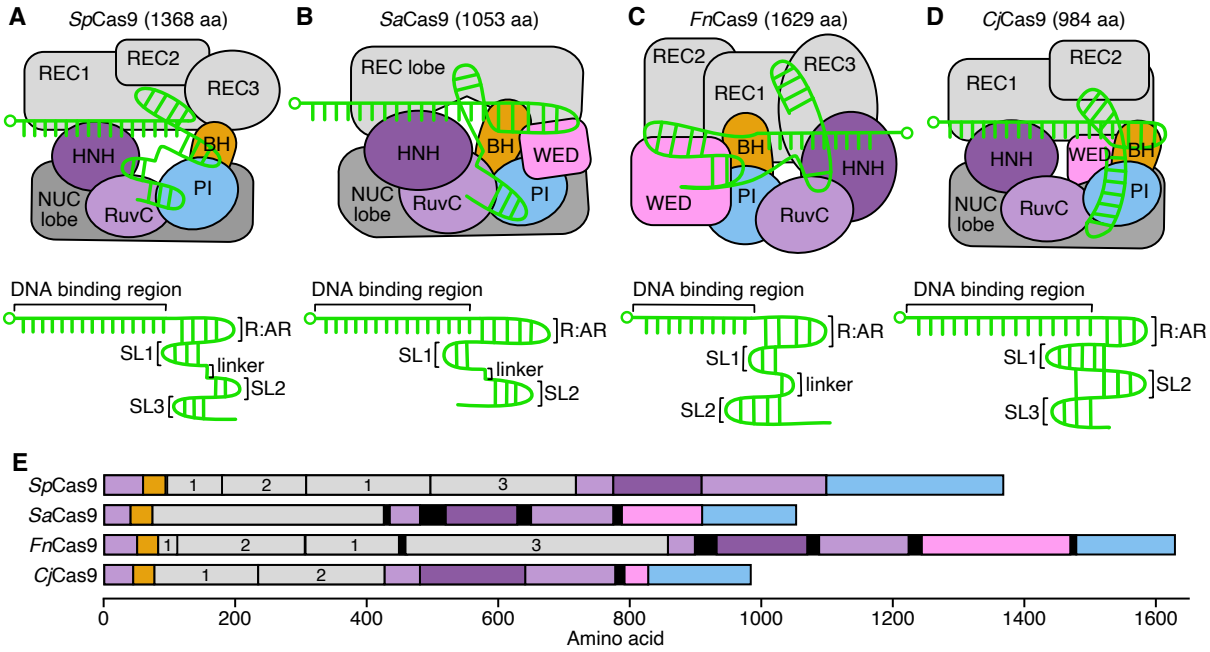

Figure 2: Naturally occurring variants of Cas9 and their respective gRNA structures. Top diagrams show the Cas9:gRNA complex and interactions of the gRNA with core Cas9 domains (labelled). Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting, CTD = C-terminal 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 (*Sp*Cas9). (**B**) *Staphylococcus aureus* Cas9 (*Sa*Cas9). (**C**) *Francisella novicida* Cas9 (*Fn*Cas9). (**D**) *Campylobacter jejuni* Cas9 (*Cj*Cas9). (**E**) Domain structure of the Cas9 variants. Linkers are shown by black regions and REC domains are numbered.

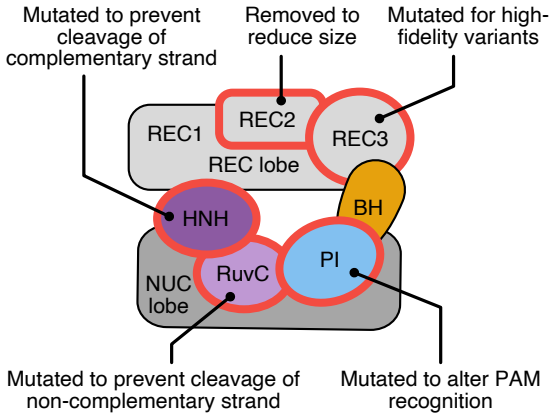

Figure 3: Key domains of Cas9 and the effect of modifications of each on phenotype. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. HNH and RuvC are nuclease domains. Thick red outlines indicate domains which have been modified.

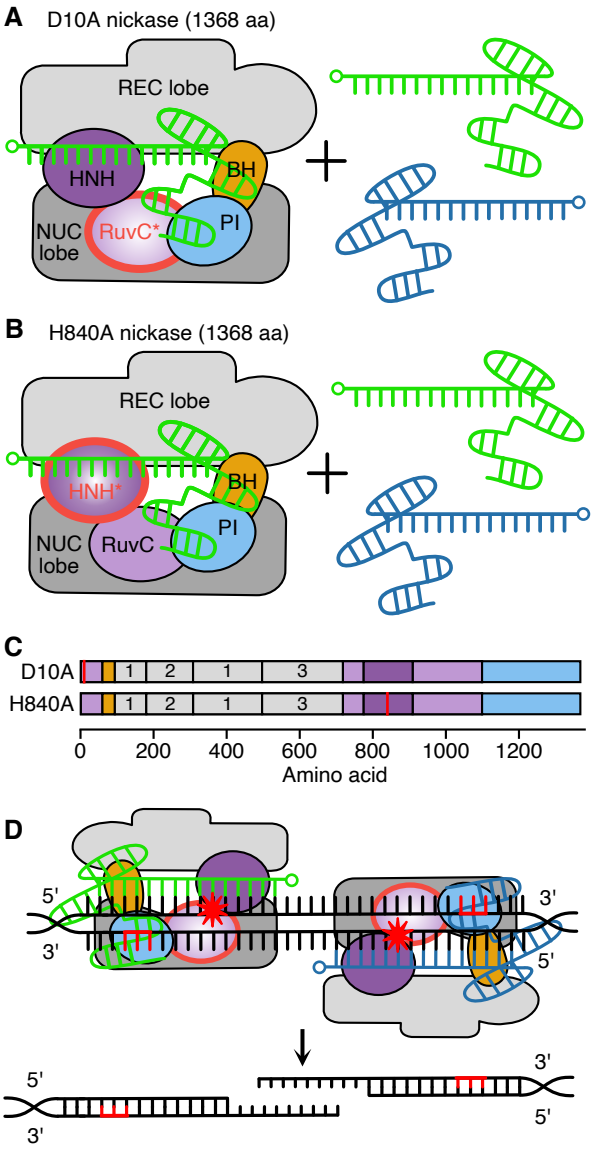

Figure 4: Cas9D10A and Cas9H840A nickase systems. (**A**) The Cas9D10A nickase system which nicks the complementary strand. This Cas9D10A is used in conjunction with a pair of guides to target each strand independently. The 5' end of each gRNA is denoted by a small circle and inactive domains are outlined 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.

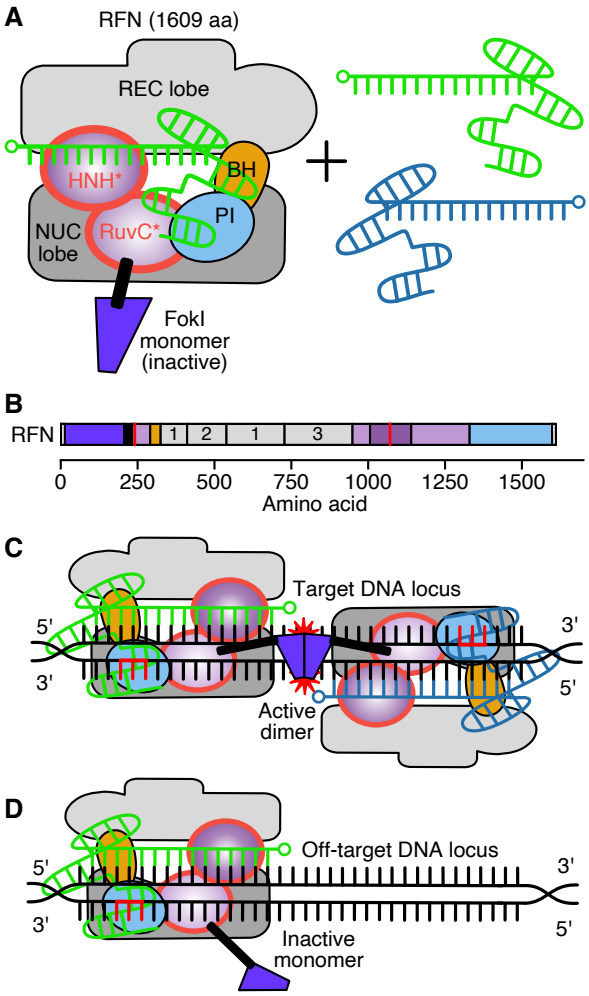

Figure 5: An RNA-guided FokI Nuclease (RFN) system. (**A**) An RFN system consists of a dCas9-FokI fusion and two gRNA (green and blue) with targets ~15 bp apart. Two FokI monomers (blue) are required for the active dimer (purple) to cleave DNA, so off-target binding of a single RFN does not (usually) 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.

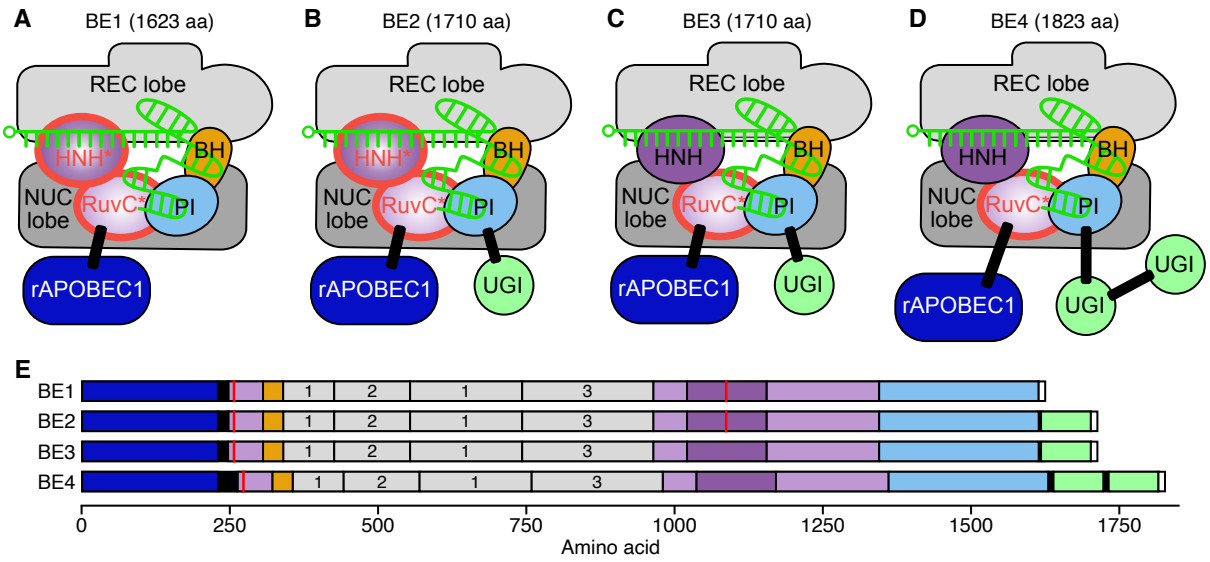

Figure 6: Base editing systems. (**A**) Base editor 1 (BE1) consists of a *Sp*dCas9 with a cytidine deaminase (rAPOBEC1) fused to its N-terminus. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. HNH and RuvC are nuclease domains. (**B**) Base editor 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.

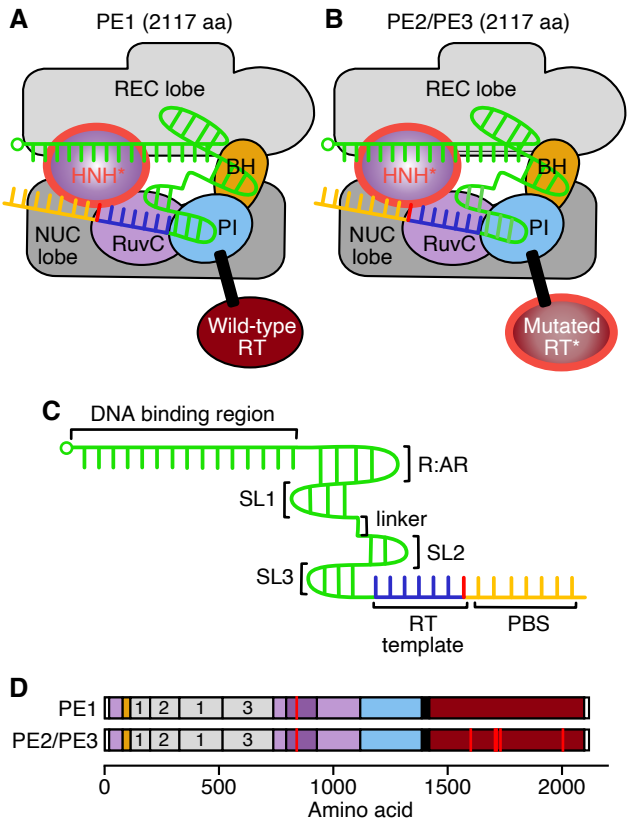

Figure 7: Prime editing systems and pegRNA. (**A**) Prime editor 1 (PE1) consists of a H840A nickase with a flexible linker fusing an M-MLV wild-type (WT) reverse transcriptase (RT; red) to the C-terminus. Domains abbreviated as: REC = recognition, NUC = nuclease, BH = bridge helix, PI = PAM-interacting. 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.