

## Life Sciences and Medicine

Special Topic: Gene Editing towards Translation

# Nucleases in gene-editing technologies: past and prologue

Dan-Yuan Li<sup>1,2,#</sup>, Long-Qi Li<sup>1,2,#</sup> & Jun-Jie Gogo Liu<sup>1,2,\*</sup><sup>1</sup>*Beijing Advanced Innovation Center for Structural Biology, Frontier Research Center for Biological Structure, School of Life Sciences, Tsinghua University, Beijing 100084, China;*<sup>2</sup>*Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China*

#Contributed equally to this work.

\*Corresponding author (email: [junjiegogoliu@tsinghua.edu.cn](mailto:junjiegogoliu@tsinghua.edu.cn))

Received 25 October 2022; Revised 5 March 2023; Accepted 7 March 2023; Published online 18 July 2023

**Abstract:** In recent decades, gene-editing technologies, typically based on deoxyribonucleases to specifically modify genomic sequences, have dramatically remodeled various aspects of life sciences, including fundamental research, breeding, and medical therapeutics. So far, four types of endonucleases have been adopted and optimized as gene-editing tools: meganuclease, ZFN, TALEN, and Cas nuclease from the CRISPR-Cas system. Each tool comes with its own advantages and limitations. Over the last ten years, RNA-guided Cas nucleases have been extensively investigated and successfully implemented in almost all mammalian cells due to their remarkable editing efficacy, high specificity, and flexibility in targeting the specific locus. Diverse Cas nuclease, together with meganuclease, ZFN, and TALEN, represent the key strategies for nuclease-based gene editing. However, systematic introductions and comparisons among four types of nucleases are not yet available. Here, we overview the capabilities of four types of nucleases along the development history of gene editing and describe the molecular mechanisms of substrate recognition and cleavage. Particularly, we summarize the promising CRISPR-Cas systems as well as modified tools applied for gene editing in the eukaryotic genome. Moreover, how the re-modulated nucleases and other nucleases, either naturally occurring or AI-designed, might manipulate DNA sequences is discussed and proposed.

**Keywords:** gene editing, CRISPR, ZFN, TALEN, meganuclease, novel gene-editing system

## Introduction

It has long been a central question of genetics to elucidate how genotypes determine phenotypes and translate the sequence information into the knowledge of function. Classical genetics was based on the research of naturally occurring genetic polymorphisms. Mutagenesis, through radiation or chemical reagents, greatly enriched the pool of mutants. However, these mutant alleles are all derived from random mutations. Gene-editing technologies provide solutions for accurate manipulation of the DNA sequence, which allows for the generation of organisms with specific gene modifications and the construction of various disease models in order to further understand gene functions. In addition, according to Online Mendelian Inheritance in Man<sup>®</sup> ([www.omim.org](http://www.omim.org)), more than 4000 genes in the human genome are related to diseases. Precise sequence modification is necessary for the correction of certain genetic defects in therapeutic applications. The

foundation of gene editing was established in the 1970s and 1980s when researchers successively performed precise gene modifications in yeast and mammalian cell lines [1,2]. Such modifications were intermediated by cellular DNA homologous recombination pathway, which allowed the exchange of homologous sequences within two gene segments. Via offering exogenous DNA donor homologous to the target region, DNA sequence replacement at specific sites could be achieved without affecting the rest of the genome. Further, based on the mouse embryonic stem (ES) cell technology [3], researchers developed the “gene-targeting” technology to knock out or replace certain genes in mice specifically. In this process, genetically modified mouse ES cells were injected into normal blastocysts to generate chimeric mice, and the chimeric mice were then crossed with normal mice to produce genetically modified offspring. However, such precise editing is limited by extremely low efficiency and requires long and laborious large-scale positive and negative screening. In addition, gene-targeting practice is highly dependent on ES-cell culture, which makes it significantly challenging in the species with hardly isolated and cultured ES cells, such as rats.

The turnaround came since researchers had gradually realized that the efficacy of homologous recombination, one of the DNA repair mechanisms, was significantly enhanced by generating the double stranded DNA breaks (DSBs) in the target region. This phenomenon was observed in a variety of model organisms, such as yeasts and mammalian cells [4–6]. Thus, DNA nucleases that can effectively introduce DSBs at specific sites became the major tools for developing gene-editing technology. The nuclease-intermediated gene-editing in eukaryotes usually involves the following steps. First, a specific target site is recognized and cleaved by a natural or artificially designed DNA endonuclease to generate the DSB. Then, the DSB is repaired, and various gene alterations in the target region are introduced by endogenous DNA-repair pathways, including HR (homologous recombination) repair pathway or NHEJ (non-homologous end-joining) pathway. In the HR pathway, the homologous DNA donor template needs to be offered for precise replacement of the target region. While in NHEJ, DSB ends are directly linked, yielding different editing results, including single nucleotide substitutions, insertions, and deletions (INDEL) in the target region. Up to now, there have been four generations of such nuclease-based gene-editing tools, which will be discussed in the following text, including meganucleases, zinc-finger nucleases (ZFN), transcription activator-like effector endonucleases (TALEN) and clustered regularly interspaced short palindromic repeats (CRISPR) and their CRISPR-associated (Cas) proteins.

## Meganuclease

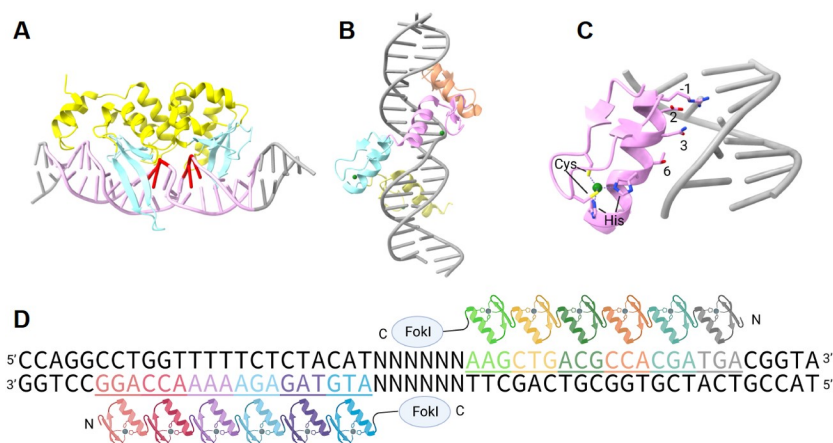
Meganuclease, a class of deoxyribonucleases that recognizes long DNA sequences (12 to 40 base pairs) [7], was the first type of nuclease repurposed as a gene-editing tool. In the case of meganuclease I-SceI from yeast, for example, it specifically recognizes and cleaves an 18 bp long dsDNA sequence. Theoretically, this 18-bp sequence would only occur once in a random sequence of 20 times the size of the human genome, which confers meganuclease with the highest specificity among natural restriction endonucleases [8,9]. Hundreds of meganucleases have been found in prokaryotes, eukaryotes, and archaea [10]. These proteins are encoded by mobile genetic elements and consist of two main families: intron-encoded endonucleases (containing the prefix “I-”), which are spliced and translated from precursor RNA, and intein endonucleases (containing the prefix “PI-”), which are spliced out of precursor proteins [11]. Meganuclease plays a role in

the replication and proliferation of mobile elements by introducing DSBs at specific genome sites, which stimulates endogenous DNA repair mechanisms leading to the amplification of introns or inteins at cleavage sites without disrupting other host genes or affecting normal functions [8,10].

The three-dimensional structure of the I-SceI and dsDNA complex solved by X-ray crystallography reveals a prototypical  $\beta$ -saddle conformation consisting of two pseudo-symmetric subunits (Figure 1A) [12]. The multiple  $\beta$  strand-hairpin structures embedded in the major groove of DNA empower I-SceI to recognize a long DNA sequence precisely, and the catalytic center of I-SceI is close to the scissile phosphodiester bond of dsDNA from the minor groove. Meganuclease provides an approach of precisely targeted cleavage on dsDNA *in vivo*. However, it is almost impossible to find identical recognition sequences on the gene of interest since recognition sites of meganucleases are fairly fixed, even though they can tolerate mutations at certain positions with reduced but not completely abolished cleavage activity [13]. Through assembling subdomain variants recognizing different sequences, researchers redesigned the meganuclease to target the human *XPC* (Xeroderma Pigmentosum Complementation Group C) gene and achieved efficient gene manipulation in mammalian cells [14,15]. Engineered meganucleases with altered recognition sequences have been implemented in site-specific gene editing in both mammals and plants [16,17]. Nevertheless, the rational design of the meganucleases requires extensive experience, and the procedure is time-consuming and arduous. Thereby, searching for modular nucleases with shorter DNA recognition sequences that allow flexible engineering and assembly became a new orientation in developing gene-editing tools.

## ZFN

Zinc-finger is a type of DNA recognition domain widely distributed in all categories of living organisms. The first zinc-finger domain was identified in the transcription factor IIIA (TFIIIA) from *Xenopus laevis*: nine tandem sequences exist in the N-terminus with a highly similar pattern:  $X_2$ -Cys- $X_{2-4}$ -Cys- $X_{12}$ -His- $X_{3,4}$ -His-



**Figure 1** (A) The structure of I-SceI bound to the DNA substrate (PDB ID: 1R7M). The meganuclease is yellow, while its multiple  $\beta$ -strands interacting with DNA bases are colored cyan. The 18-bp DNA sequence recognized by I-SceI is in pink, and the DNA cleavage sites are in red. The cartoon is depicted with ChimeraX 1.3. (B) Tandem zinc-finger repeats with the target DNA (PDB ID: 2I13). (C) An individual zinc finger repeat interacting with DNA. Key protein residues responsible for DNA base recognition and zinc ion coordination are shown as sticks. The zinc ion is presented as a green sphere. (D) Schematic diagram of ZFN. ZF modules are indicated in different colors, and DNA triplets are underlined and shown in the same color.

X<sub>2-6</sub>, where X represents variable amino acid residues [18]. Subsequent studies revealed the functions and structural information of zinc-finger domains. Each zinc-finger domain can recognize 3–4 bp on the dsDNA. Through the tandem linkage between multiple zinc-fingers units, a large zinc-finger protein can be formed that specifically recognizes a segment of DNA (Figure 1B). The contacts with bases by zinc finger proteins occur mostly on one strand of DNA, called the primary strand, from 3' to 5' direction. Zinc-finger domains adopt a highly conserved  $\beta$ - $\beta$ - $\alpha$  backbone structure, each of which binds a single zinc ion (Figure 1C). This zinc ion is coordinated by two cysteine residues near the turn of two  $\beta$ -sheets and two histidine residues in the C-terminus of  $\alpha$  helix forming a stable backbone structure. The amino acid residues on the surface of the  $\alpha$  helix and in the N-terminal loop region can dock into the DNA major groove to interact with bases laterally, generating the preference for different DNA bases [19,20]. Usually, residues 2, 3, 6, and residue –1 (according to their positions relative to the N-terminus of the  $\alpha$  helix) are the key residues that preferentially contact DNA bases: residue –1, 3, 6 recognize the 3', intermediate, and 5' bases in the triplet, respectively. Whereas residue 2 recognizes the DNA base on the other strand in the local or adjacent base triplet. However, many studies have reported non-canonical base recognition patterns in both natural and artificial zinc-finger proteins, including complex and cooperative interactions across residues [21,22]. In addition, designed zinc finger proteins obtained by linking deciphered zinc finger modules may recognize unintended sequences, which still need experimental proof, due to comprehensive recognition or site degeneracy (termed as “context-dependency”) [22].

The modularity and high programmability of zinc-finger confer great potential as an engineered protein for gene editing. By combining the tandem zinc-finger domains with a non-specific endonuclease domain from the C-terminus of bacterial *Fok I* restriction enzyme (Figure 1D), researchers devised the fusion protein ZFN and demonstrated that ZFN could specifically recognize and cleave target sites *in vitro* [23]. Subsequently, ZFN was injected into *Xenopus laevis* oocyte nuclei, where specific DNA cleavage and induced homologous recombination were observed [24]. Since dimerization is required by the *Fok I* domain for dsDNA cleavage activity [25], targeted cleavage of specific sites using ZFN usually involves a pair of tail-to-tail ZFNs designed to target the upstream and downstream sequences of the cleavage site, respectively. Through structure-based design, ZFN recognizing 9–18 bp long dsDNA was created by linking different ZF modules using conserved linker sequences [26]. The 18-bp recognition sequence allows for recognizing and targeting particular sites at the genomic scale (billions level) and empowers the practical value of ZFN technology. Engineered ZFNs were widely applied for gene editing in a variety of model plants and animals, as well as gene therapies [27–31]. It was demonstrated that ZFN could be applied to primary human cells for efficient repair of severe combined immune deficiency (SCID) mutation in the *IL2R $\gamma$*  gene through the HR pathway [32]. Targeting T-cell *CCR5* (HIV co-receptor) by ZFN to obtain HIV resistance has also been implemented in clinical trials [33]. In addition, ZFN technology allows researchers to perform efficient gene knock-outs or knock-ins in animals such as rats and cattle, where gene targeting was previously difficult using conventional methods due to the lack of ES cell culture technology [34,35].

With the progressive discovery of ZF modules that recognize different base triplets [36–39], various methods have been proposed in order to construct the corresponding ZFNs for specific sites. The “modular assembly” approach screens the identified zinc-finger library for modules that can recognize corresponding base triplets and assembles by rational design [40–42]. Though easy to manipulate, ZFNs designed by “modular assembly” often have low efficiency and high cytotoxicity [43,44]. This may be attributed to the

non-specific recognition and “context-dependency” nature of ZF arrays [20,21,45]. Another kind of construction method is based on pool selection, such as oligomerized pool engineering (OPEN), for suitable polydactyl ZFN proteins from the randomized recombination library of zinc fingers [46]. However, large-scale screening also poses difficulties and high costs for applications. ZFN also has some shortcomings, such as a high preference for guanines [47]. Nevertheless, ZFN technology has pioneered the paradigm of “modular recognition domain + non-specific cleavage domain” and greatly broadened the application of gene editing. Despite disadvantages, including difficult design and sequence preference, ZFN has taken its place in scientific research and gene therapy due to its high specificity and small size.

## TALEN

The transcription-activator-like effector (TALE) was originally discovered in 1989 as an invasive element in the plant pathogen *Xanthomonas* spp. [48]. It was not until 20 years later that it stepped onto the center stage of gene editing after its function and structure were figured out. Within a few years, it quickly became the next generation of gene-editing tools. TALE is injected into plant cells through the bacterial Hrp (hypersensitive response and pathogenicity)-type III secretion system to target the promoters and induce the transcription of specific host genes in order to suppress the host immune system as well as modulate the transcriptome for pathogen proliferation. There is a type III translocation signal at the N-terminal end of the TALE protein and two or more nuclear localization signals (NLS) with a conserved activation domain (AD) at the C-terminal end. The most distinctive feature of the TALE proteins is the direct repeats in its central part. The first repeat in the N-terminal part is unique, called repeat 0. The rest of the repeats, each encoding 34 amino acids, are identical, except for two variable amino acids at positions 12 and 13, later termed as RVDs (repeat variable di-residues) [49]. The highly consistent direct repeats with different RVDs resemble the feature of the zinc-finger protein, which implies an unknown coding pattern between TALE repeats and DNA sequences.

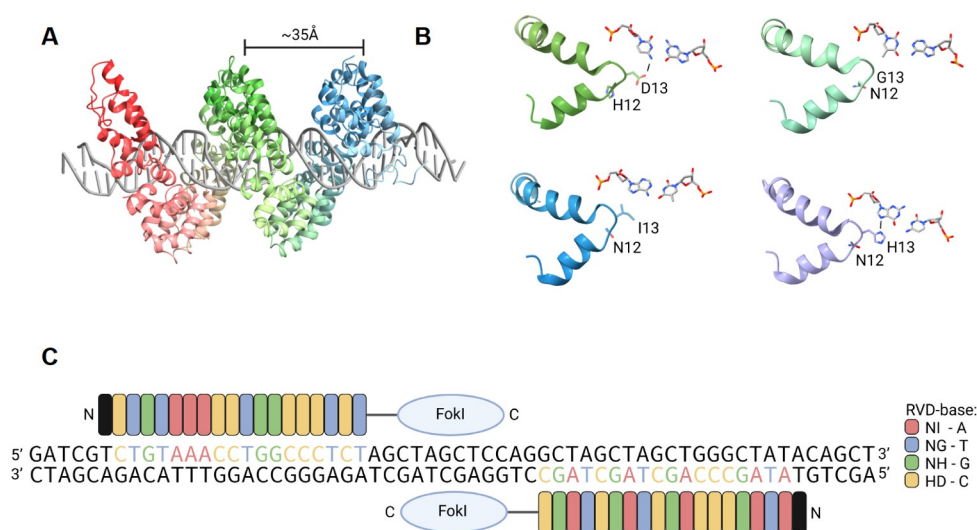
In 2009, two articles published in *Science* independently reported the correspondence between DNA sequences and the RVDs of TALE. Based on the known TALEs with target promoters, Moscou and Bogdanove [50] performed low-entropy alignments to obtain predicted binding sites assuming one-to-one correspondence between RVDs and bases. The different TALE-promoter combinations derived consistent RVD-base preferences. By further analysis of additional TALE and infection-activated promoter sequences from public microarray data, a correspondence between TAL effectors and DNA sequences was determined [50]. By analyzing sequence differences between TALE-induced and uninduced alleles, Bonas and her colleagues [51] predicted TALE binding sites and accordingly summarized the base frequency for different types of repeats, which was then experimentally validated by artificially designed TAL arrays. Both groups reached the same coding rule, which is relatively straightforward. The first position of the TALE recognition sequence is thymine, which is recognized by the unique repeat (Repeat 0) in the N-terminus of the repeat array. Then subsequent bases in the targeting sequence correspond strictly one-to-one to the repeats, with preference determined by the RVDs. The common RVDs mainly include HD, NG, NI, and NN, which recognize cytosine (C), thymine (T), adenine (A), and adenine/guanine (A/G), respectively. Subsequently, by biochemical assays on the artificial TALE array consisting of naturally absent RVDs, the decoding of TALE was



fully resolved, including all 400 (20×20) possible RVDs, which provided specific RVDs for each base, with recognition of G by NH [52,53].

Soon after the coding rules of TALE recognition had been clarified, the crystal structures of TALE bound to substrate dsDNA were resolved by two groups [54,55]. The TALE recognition region, consisting of multiple direct repeats, displays a right-handed, super-helical structure that recognizes the sense strand of dsDNA along the major groove (Figure 2A). The DNA in the TALE-DNA complex adopts undisturbed B-form with phosphate backbones localized along a positively charged helical ridge formed by invariable amino acids in the repeat backbone. TALE unbound to DNA exhibits a more extended structure, with a 60 Å pitch compared to 35 Å in the TALE-DNA complex. Each repeat consists of 33–35 amino acid residues and presents a highly conserved structure containing two helices joined by a loop containing the RVD (Figure 2B). The 12th residue of the RVD stabilizes the structure of repeat by forming a hydrogen bond with the carbonyl oxygen atom of Ala8 in the scaffold, while the 13th residue interacts with the base laterally [54]. Asp13 in HD RVDs recognizes cytosine through van der Waals interactions along the ring as well as a hydrogen bond to N4 of cytosine. Gly13 with no side chain in NG and HG RVDs leaves space for the 5'-methyl group of thymine, allowing for van der Waals interactions between the methyl group and the peptide chain backbone (average distance ~3.3 Å). NI RVD achieves recognition of adenine through van der Waals interactions between the aliphatic side chain of Ile and the base ring [55]. NH RVD recognizes G through the hydrogen bond between His13 and the purine N7 atom [56] (Figure 2B). In addition, Repeat 0, preceding the canonical repeats, forms a similar structure, with a tryptophan residue recognizing thymidine by forming van der Waals interactions with its 5'-methyl group [55]. The crystal structure sheds light on the molecular mechanism of TALE recognition with high plasticity and modularity, laying the foundation for engineering and applications.

Resembling the design of ZFN, TALEN was created by fusing the DNA recognition domain of the TALE protein with the *Fok* I endonuclease domain (Figure 2C). Once introduced, TALEN was rapidly applied to gene editing in various model organisms, including yeast [57], *C. elegans* [58], *Drosophila* [59], zebrafish



**Figure 2** (A) The structure of the TAL effector with DNA substrate (PDB ID: 3UGM). (B) The interactions between 4 kinds of TALE repeats with different RVDs and corresponding base pairs. The RVDs and base pairs are shown in sticks. Dashed lines indicate H bonds. The HD, NG, and NI RVDs are derived from PthXo1 (PDB ID: 3UGM). The NH RVD is derived from engineered TALE Hax3 (PDB ID: 4OSL). (C) Schematic diagram of TALEN. TALE modules are indicated in different colors, with the corresponding base shown in the same color.

[60], *Arabidopsis* [61], as well as medical disease treatment on Duchenne muscular dystrophy (DMD) [62] and Hepatitis B virus (HBV) [63], etc. Furthermore, using a pair of TALEN, two long-distance DSBs can be generated, potentially leading to large-scale deletions, inversions, and translocations. This allows researchers to manipulate elements not susceptible to frame-shift mutations, such as microRNA and long non-coding RNA (lncRNA), as well as to construct cancer models containing chromosome changes. Compared to zinc fingers, TALE repeats with better modularity can be linked tandemly without protein engineering or consideration of contextual correlation between repeats, which significantly decreases the threshold for TALEN technology application and shortens the time required for construction. Several programmed TALEN construction approaches have been proposed, including Golden Gate, iterative capped assembly (ICA), fast ligation-based automatable solid-phase high-throughput (FLASH) system, ligation-independent cloning (LIC), etc. [64–69]. Some of those have been reported to perform in-batch automated TALEN constructs. In addition, although still constrained to the first thymidine at the target site, TALEN has no sequence preference similar to that of ZFN, which allows it to target nearly any sites of interest theoretically, thus permitting the construction of genome-scale TALEN libraries in bulk [70]. Some cases demonstrated that TALEN exhibited lower off-target and cytotoxicity compared to ZFN, while more complete assessments are still required [71]. Despite many advantages, the large molecular weight of TALEN is a major drawback, which is not conducive to delivery, especially for vehicles with strict molecular weight constraints such as AAV (adeno-associated virus).

## CRISPR-Cas system

### *Discovery history of CRISPR*

The CRISPR-Cas system is widely distributed in bacteria and archaea as part of the adaptive immune system to defend against invading nucleic acids [72]. CRISPR was first discovered in 1987 [73], characterized by a cluster of 29 nt repeats with 32 nt non-repeat sequences (spacers) between such repeats. A series of specific CRISPR-associated protein (Cas) genes were identified in the vicinity of the CRISPR array, which presumed functional correlation [74]. The speculation of CRISPR as an adaptive immune system was raised in 2005 when 3 articles reported the origin of the spacers from phages or conjugative plasmid as well as the relevance between spacer and resistance to the corresponding invaders [75–77]. Such speculation was experimentally proven that after phage infection, the isolated resistant strains acquired novel spacers corresponding to the phage genome, and the resistance was correlated with the removal or addition of such spacers [78]. Also, researchers have noticed the consistency of several nucleotides adjacent to the sequences that spacers matched [77], which was later named PAM (protospacer adjacent motif).

In 2011, Siksnys and his colleagues [79] successfully conferred resistance to corresponding phages on *Escherichia coli* (*E. coli*) by reconstituting the Type II CRISPR system from *Streptococcus thermophilus* for the first time and proved that *Cas9* was the only gene necessary for interference, which was dependent on its RuvC and HNH domains. One year later, two papers published by Doudna and Charpentier groups [80] and Siksnys group [81] reported biochemical characteristics of *Cas9 in vitro* independently. *Cas9* could introduce double stranded break (DSB) to target DNA with the assistance of CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) *in vitro*. The target sequence was recognized by crRNA through base-pairing

and could be redesigned to target different sites. They also proved that the target strand (TS, the strand complementary to crRNA) was cleaved by the HNH domain, while the non-target strand (NTS, the strand complementary to TS) was cleaved by the RuvC domain. In addition, the fusion of crRNA with tracrRNA into a single-guide RNA (sgRNA) could also direct Cas9 to cleave target DNA. The model of sgRNA was widely used in subsequent applications.

The function of CRISPR-Cas as an adaptive immune system contains three main phases [82]. (1) Adaptation. The Cas1-Cas2 complex cleaves exogenous DNA and integrates it into the CRISPR locus. (2) crRNA Maturation. When the host is reinfected, the CRISPR array can be transcribed to form a long precursor crRNA (pre-crRNA). The pre-crRNA is then processed by Cas proteins or host RNases into mature crRNA. (3) Interference. Under the guidance of crRNA (with tracrRNA in some cases), Cas proteins can specifically recognize and cleave exogenous DNA complementary to the spacer, leading to target interference and immunity. Components of the effectors for interference are varied in different systems.

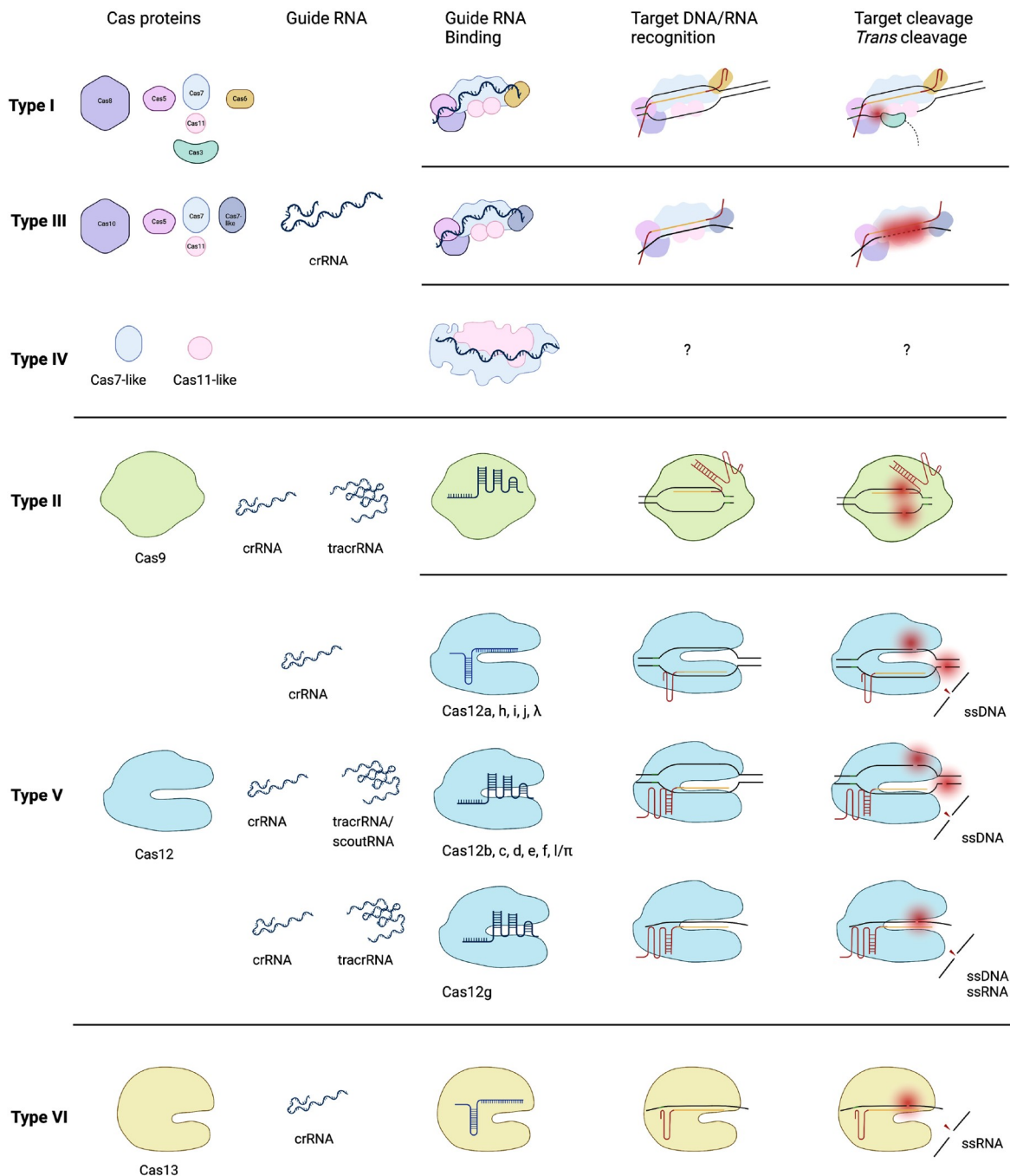
### ***Diversity of CRISPR nucleases***

Since the discovery of the interference capability of Cas9, numerous research groups have predicted and experimentally validated highly diverse CRISPR-Cas systems. And this diversity is continuously expanding. Based on the organization of the effector complex, sequence similarity, and the arrangement of Cas genes in the CRISPR locus, CRISPR systems can be classified into 2 classes, 6 types (I–VI), and over 30 subtypes [86] (Figure 3). The most significant difference between these 2 classes is that Class I effector nucleases consist of multi-subunits, while Class II effector complex is composed of a single multi-domain protein RNP complex.

Class I systems are distributed in both bacteria and archaea, accounting for about 90% of the reported CRISPR-Cas loci [87]. Class I systems are further divided into Type I, III, and IV based on the composition of signature effector nucleases (Figure 3). In Type I system, Cas5, Cas6, Cas7, Cas8, and Cas11 assemble along the crRNA and initiate the DNA unwinding and R-loop formation [88,89]. Then, an effector nuclease, as well as a helicase, Cas3, is recruited to the RNP complex to unwind further and degrade the DNA substrate [90,91]. In Type III system, targeting RNA is recognized by the effector complex through base-pairing with crRNA and cleaved by Cas7-like proteins. Upon binding crRNA, the Cas10 nuclease is activated and exhibits both nuclease activity for ssDNA cleavage and enzymatic activity for cyclic oligoadenylate [92] synthesis [93,94]. The effector nucleases in Type IV are the least understood among the six types. Four subtypes in Type IV-A, -B, -C, and -D are distinct in specific enzymes: dinG, cysH-like, Cas10-like, and a helicase of the RecD family, respectively [95,96].

Class II systems are mainly present in bacteria and rarely detected in archaea and can be divided into Types II, V, and VI based on the different effector nucleases [97] (Figure 3). The Type II systems have been extensively studied due to their widely used signature nuclease Cas9. Diverse Cas9 orthologs have been identified with versatile properties in the PAM spectrum, cleavage pattern, and cleavage activity at varying temperatures [92]. The systems in Type V mainly target dsDNA with Cas12 nucleases, which contain a conserved RuvC domain instead of two nuclease domains as Cas9 [98]. Despite the recently identified Cas12I (also denoted as Cas $\pi$ ) nucleases which prefer 5' C-rich PAM, most Cas12 homologs recognize 5' T-rich PAMs and cleave the NTS and TS sequentially at PAM distal end, forming a staggered end with the 5'





**Figure 3** The Cas proteins involved in target cleavage of six types. In Type I systems, multiple subunits assemble along the crRNA and initiate the R-loop formation, which then recruits Cas3 for dsDNA cleavage. The complex in Type III exhibits a similar architecture as that in Type I, and the difference is that this complex recognizes and degrades complementary target RNA. Complex in Type IV is less understood; here shows the ribonucleoprotein (RNP) complex of Type IV-B, whose structure has been determined [83]. Cas9 nuclease in Type II mediates dsDNA cleavage guided by crRNA and tracrRNA. The non-targeting DNA cleavage of Cas9 is reported to be RNA-independent in the presence of  $Mn^{2+}$  ions [84]. However, in Type V, Cas12 has both targeting and non-targeting DNA cleavage activity guided by crRNA alone or crRNA and tracrRNA (scoutRNA in Type V-C and Type V-D). Cas13 from Type VI is an RNA-guided RNA nuclease. And target-activated Cas13 complex can cleave surrounding RNA molecules non-specifically [85].

overhang [99,100]. The Cas12 family contains more diversity in phylogeny. Based on the bioinformatics analysis, several members of the Cas12 family, including Cas12a to Cas12l, have been reported gradually. Different Cas12 nucleases vary greatly in domain architecture, guide-RNA, PAM requirement, cutting velocity, targeting specificity, etc. To date, dsDNA-cleavage activity has been experimentally validated in 10 subtypes, including Cas12a (previously known as Cpf1), Cas12b (previously known as C2c1), Cas12d (previously known as CasY), Cas12e (previously known as CasX), Cas12f, Cas12h, Cas12i, Cas12j (previously known as CasΦ), Cas12l and Casλ [98,100–108]. Some naturally DNase-free Cas12 nucleases are demonstrated with various functions in biology. In bacteria, the Cas12c-scoutRNA complex binds to the targeted viral-derived DNA as a transcription repressor to provide antiviral immunity [109]. Naturally inactive Cas12k nuclease is co-opted for RNA-guided DNA transposition [110]. Distinguished from other Cas12 nucleases, Cas12g can cleave RNA guided by crRNA and tracrRNA without requiring a PAM [105]. Notably, after target cleavage mediated by guide RNA, all the active Cas12 nucleases can degrade non-specific ssDNA *in trans* [111]. Moreover, for Cas12a, the *trans*-cleavage activity also affects RNA and dsDNA in addition to ssDNA *in vitro* [112,113]. The magnitude of *trans* activity varies among nucleases in different types as well as orthologs from the same subtype.

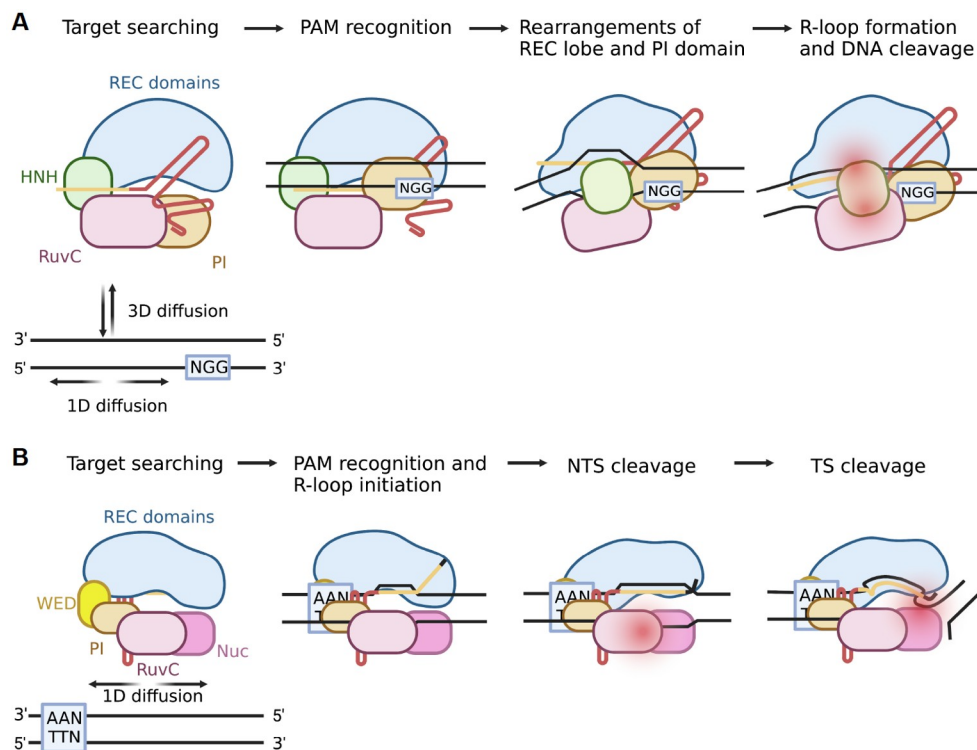
The effector nuclease for the Type VI system is Cas13, which contains two higher eukaryotes and prokaryotes nucleotide-binding (HEPN) RNase domains [102]. It recognizes, and cleaves targeted RNA guided by crRNA [85]. Since the review is to discuss gene editing, we will mainly focus on DNA nucleases rather than Cas13.

### ***Target recognition and cleavage of CRISPR nucleases***

The interactions between Cas-gRNA and target DNA are mechanistically diverse. PAM recognition and gRNA-DNA hybridization provide the specificity for gene editing. Various nucleic acid-binding domains are involved in inducing conformational changes for double-stranded DNA cleavage. Since CRISPR-Cas systems from Types II and V have been widely programmed to manipulate the genome of eukaryotic cells, we will mainly discuss the cleavage processes of these systems represented by CRISPR-Cas9 and CRISPR-Cas12a, respectively. We summarize and provide mechanistic models for DNA cleavage by *Streptococcus pyogenes* Cas9 (SpCas9) (Figure 4A) and *Acidaminococcus sp.* Cas12a (AsCas12a) (Figure 4B).

#### ***CRISPR-SpCas9***

The structure of SpCas9 has been best-elucidated [114,115]. SpCas9 consists of two lobes: a recognition lobe (REC lobe) and a nuclease lobe (NUC lobe) [114]. The NUC lobe contains a RuvC domain, an HNH domain, and a PAM-interacting (PI) domain. The PI domain is critical for PAM recognition and specificity [115]. The PAM duplex is nestled in the positively charged cleft on the PI domain [114]. HNH and RuvC nuclease domains are responsible for the cleavage of TS and NTS, respectively. RuvC domain has an RNase H fold and requires Mg<sup>2+</sup> for cleavage activity. In similar positions to the RuvC nuclease in *E. coli*, Asp10, Glu762, His983, and Asp986 of the SpCas9 are catalytic residues for nuclease activity [115]. The HNH domain contains a ββ-metal fold and cleaves the nucleic acid substrate catalyzed by Asp839, His840, and Asn863 [115]. Mutation of the catalytic residues in these two domains results in Cas9 nickase or dead Cas9 without



**Figure 4** Schematic models of staggered cleavage of DNA by SpCas9 (A) and Cas12a (B). (A) The SpCas9-sgRNA complex searches the target via both 1D and 3D diffusion. The PAM recognition by the PI domain initiates the R-loop formation. As the R-loop elongates, the REC domains and PI domain undergo significant conformational changes and distort the RNA-DNA heteroduplex, enabling the activation of the HNH domain. Finally, the NTS and TS are cleaved by RuvC and HNH domains, respectively. (B) The AsCas12a-crRNA complex searches the target through 1D diffusion. Unlike SpCas9, AsCas12a recognizes bases T and A from both strands in the PAM region. As the formation of the R-loop, the REC lobe and Nuc lobe become more open to accommodate the RNA-DNA heteroduplex. The NTS is displaced and further cleaved by the RuvC domain. The TS is then loaded to the active site of the RuvC domain and cleaved, which is speculated to be mediated by the Nuc domain.

harming the target binding activity, which is an advantage for precise gene editing. REC lobe interacts with repeat: anti-repeat duplex of guide RNA to facilitate RNA-guided DNA cleavage [115]. As described later, mutants in the REC domain or target-binding region would affect on-target specificity. And mutations of arginines in the PI domain enable Cas9 to recognize alternative PAMs.

Under the physiological salt condition, the SpCas9-gRNA complex combines 3- and 1-dimensional asymmetric diffusions to search 5'-NGG-3' PAM and the flanking region [116,117]. Once the complex finds the target, two conserved arginine residues (Arg 1333 and Arg 1335) form major groove interactions with the GG dinucleotide in the NTS [114]. Interactions with the minor groove of the PAM duplex and the phosphodiester group at the +1 position in the target DNA strand by the Lys 1107-Ser 1109 loop promote the local melting immediately upstream of the PAM [114]. Then the spacer sequence hybrids with the target to form a catalytically competent R-loop. Along with this process, the domains in the REC lobe go through rearrangements and relocate the HNH domain [118]. Once 17 or more base pairs form, the HNH domain is conformationally activated, driven by the displacement of REC domains and distortion of the gRNA-DNA duplex [118]. Meanwhile, both NTS and TS are located to and cleaved by the catalytic sites of the RuvC domain and HNH domain, respectively. The scissile phosphate of the target strand is hydrolyzed via the one-metal-ion mechanism, whereas the cleavage of NTS is mediated by the two-metal-ion mechanism. In the

HNH domain, an  $Mg^{2+}$  is coordinated into the catalytic center with six surrounding oxygen atoms from a water molecule, three polar residues (Asn863, Asp839, and Asn854), and the scissile phosphate between nucleotide +3 and +4 of the TS [115,119,120]. Then, the side chain of H840 activates a water molecule for nucleophilic attack on the former scissile phosphate, generating two products with 5' phosphate and 3' hydroxyl [115,119]. For the cleavage of NTS, two  $Mg^{2+}$  are positioned into the catalytic center surrounded by water molecules, some acidic residues, and the target scissile phosphate [115,119]. Subsequently, the cleavage is mediated by His986 in a way similar to that of TS cleavage. It is conceived that the coordination of two ions induces the conformational change toward the active state and bridges the gap between the NTS and RuvC domains in varying degrees [120].

The HNH domain-mediated cleavage occurs at the position of 3 nucleotides upstream of PAM on TS [81,121]. Unlike TS, the exact cleavage site on NTS was controversial at first. It was generally believed that SpCas9 created a blunt end at 3 base pairs (-3 position) upstream of the PAM [80,81,114,115]. Subsequently, the NTS was trimmed from 3' to 5', mediated by the exonuclease activity of Cas9 [80]. This conclusion is drawn from the patterns of cleaved products on the denaturing polyacrylamide gel [80]. The staggered cleavage pattern was not deeply analyzed until 2017 when Wu and his team [122] demonstrated that Cas9 could cleave the NTS endonucleolytically and generate a staggered end with 1 to 3-nucleotide (mainly 1 nucleotide) overhang at the 5' end. Actually, the double  $Mg^{2+}$  binding at the phosphate between -4 and -5 is energetically more favorable than binding at other positions [120]. More importantly, Wu and his team's [122,123] novel work suggested that Cas9-mediated nucleotide insertions can be predictable. Similar findings were demonstrated by another two groups through a machine-learning model [124,125]. These works have inspired investigators to persistently improve the editing outcomes by manipulating DNA repair pathways for research and clinical applications [123,126].

### CRISPR-AsCas12a

The structures of AsCas12a and *Lachnospiraceae bacterium* ND2006 Cas12a (LbCas12a) were first resolved after the identification of type V effector proteins in 2015 [98,127,128]. Similar to SpCas9, AsCas12a adopts a bilobed architecture consisting of the REC lobe and NUC lobe. The REC lobe is composed of helices-rich REC1 and REC2 domains. They form a positively charged channel with the RuvC domain to stabilize the crRNA-DNA duplex. The NUC lobe contains the RuvC domain and three unique domains referred to as Wedge [93], PI, and Nuc domains [127]. WED and PI domains play functional roles similar to those of domains in Cas9 [114,115,129,130]. The 5' handle of crRNA accommodates within the groove between the RuvC domain and WED domain. Specifically, the WED domain recognizes the U(-1)•U(-16) bases in the 5' handle of crRNA in a base-specific manner [127,131]. The Nuc domain is considered responsible for the TS cleavage, for the mutation R1226A on the Nuc domain shows slightly reduced NTS cleavage activity, whereas the TS cleavage is almost blocked [127]. In addition to these common structural features, more unique domains have been identified recently in other Cas12 homologs, such as the non-target strand binding (NTSB) domain in CasX (Cas12e), zinc-finger in CasΦ (Cas12j), lock-catch (LC) and proline-rich string (PRS) domains in Casπ (Cas12l), provided new insights into the mechanism and evolution of Type V CRISPR family [100,103,132,133].

AsCas12a-crRNA complex searches for the target site by one-dimensional diffusion along the DNA [134].

The target DNA recognition initiates with PAM-containing base pairs binding by residues from PI, WED, and REC1 domains via hydrogen bond and salt bridge [131]. Among these residues, Lys607 in the PI domain specifically recognizes both N3 from the second dA and O2 from the first dT upstream of the 20-nt target, contributing to the 5'-TTN-3' PAM specificity of AsCas12a [131]. PAM recognition promotes DNA melting. Then the crRNA hybrids to the target DNA and the dramatic rearrangements of the REC lobe form a channel to accommodate the crRNA-DNA heteroduplex [135]. After the formation of the complete R-loop, the NTS is displaced into the catalytic site of the RuvC domain and cleaved [131]. Interestingly, the formation of the crRNA-target DNA duplex is limited to 20 bp, for the remaining base pairings are blocked by stacking interaction from Trp382 in the REC2 domain [127]. In the studies of *Francisella novicida* Cas12a (FnCas12a), the PAM distal target DNA is perceived to further separate, facilitating the cleavage of TS [136]. This sequential cleavage manner of Cas12a has been validated by single-molecule fluorescence assays [134,137–139]. It is speculated that the Nuc domain in FnCas12a induces a kink in the TS during the interaction, so the TS is well-loaded into the catalytic site of the RuvC domain [136]. Bulk cleavage assays by Cofsky et al. [140] suggest that FnCas12a cleaves the target strand within a tract of DNA destabilized by adjacent R-Loop. In summary, NTS cleavage occurs in the region opened by R-loop formation, whereas TS cleavage is likely dependent on the melting of the double stranded DNA beyond the crRNA-TS hybrid region at the 5' end of the TS, leaving the cleavage sites on TS exposed to the catalytic pocket in a single-strand form. In this way, Cas12a cuts NTS and TS inside and outside the R-loop, respectively, generating staggered ends. In the case of Cas9, the cleavage site on TS by HNH is more unambiguously identified as occurring at the phosphate between +3 and +4. However, the non-specific cleavage of the single strand by RuvC is impacted by the degrees of the coordination of divalent ions and the TS-loading process by Nuc or zinc-ribbon domains [141]. In addition, cleavages of NTS and TS occur within the R-Loop, so that both blunt and staggered ends are produced [80,122]. These conclusions well-explain the fundamental cleavage mechanisms by a single-RuvC domain of Type V CRISPR-Cas systems.

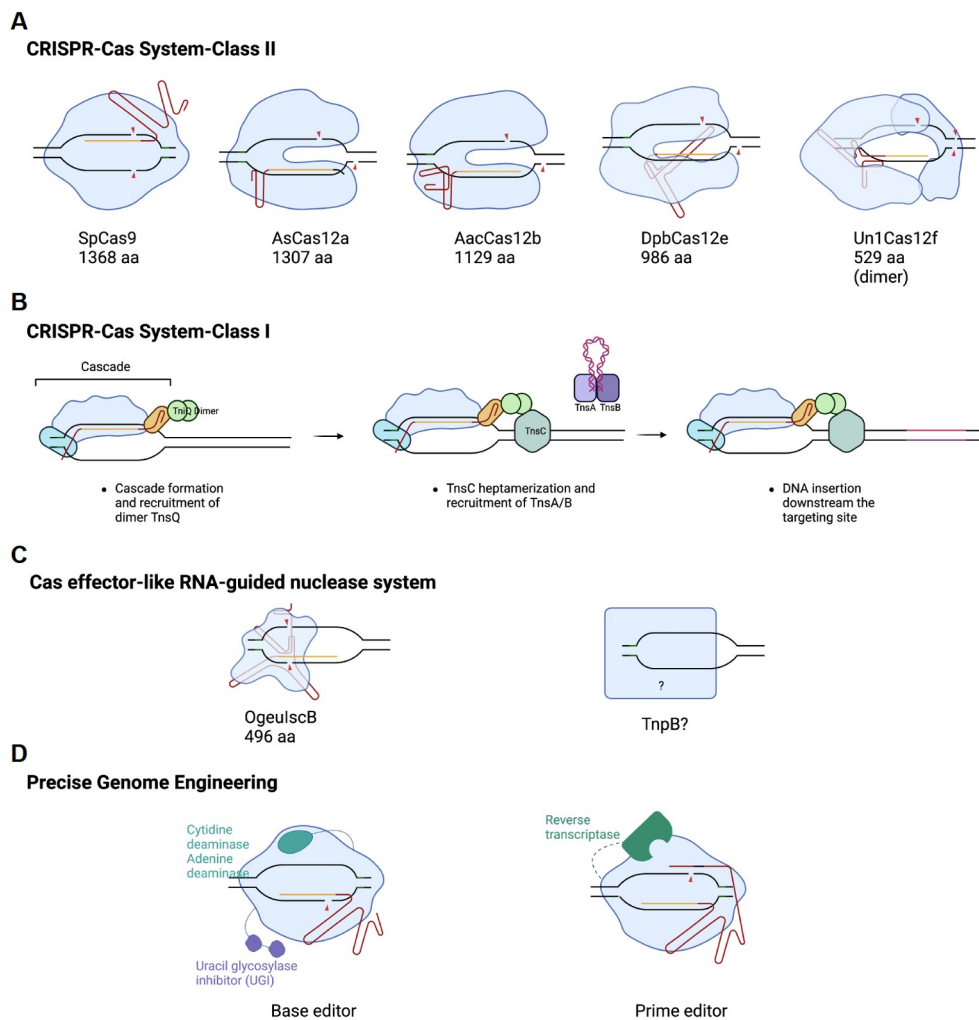
## Gene editor derived from Cas nucleases

Due to the programmability of targeting sites by simply changing spacer sequence, CRISPR-Cas systems that cut the dsDNA with single-protein nucleases, especially Cas9 and Cas12 (Figure 5A), have been widely utilized for gene editing. Additionally, the naturally evolved diversity of Cas nucleases provides various options in PAM, type of substrate, efficacy, and specificity for desired genetic manipulation. In the following sections, we will overview the gene-editing capacity of these nucleases, including some naturally occurring Cas nucleases (Cas9, Cas12a, Cas12b, Cas12e, Cas12f, engineered Cas variants, Cas proteins from Class I); Cas-like nucleases reported with promising gene-editing efficacy, like IscB and TnpB. We will also briefly summarize the precise gene-editing tools developed from CRISPR-Cas systems.

### Cas9

CRISPR-Cas9 is one of the best-characterized, dual-RNA guided DNA endonuclease systems. Cas9 is the first Cas nuclease reported to mediate programmable genome editing in mammalian cells [142–145], which





**Figure 5** (A) CRISPR-Cas systems for genome editing in Class II. (B) The positioning mechanism of transposons associated with Type-I-F system. (C) The ancestors of some effector proteins from the CRISPR-Cas systems have been identified and, subsequently, shown to have the capacity for human genome editing. More investigations on the diversity and structural information are necessary. (D) Strategies for precise genome editing.

greatly inspires interest in developing more gene-editing tools using CRISPR-Cas9 systems. Although SpCas9 is most widely used in gene-editing experiments, it still has many limitations. It is not sensitive to mismatches in the gRNA-DNA duplex. It can also recognize alternative PAMs, like 5'-NAG-3' and 5'-NGA-3' [146]. Additionally, it is difficult to deliver into eukaryotic cells efficiently for its relatively large size. Therefore, researchers have developed various CRISPR-Cas9 systems over the past few years. Some of these naturally-occurring variants show improved properties. Besides, structural studies have provided instructive information for engineering Cas9 nuclease by rational designs or evolution strategies.

#### Naturally-occurring Cas9 variants

Two orthologs of SpCas9, *Streptococcus thermophilus* Cas9 (St1Cas9) and *Staphylococcus aureus* Cas9 (SaCas9), function efficiently in human cells [146]. The small protein size and robust genome editing

efficiency comparable to SpCas9 make SaCas9 a good candidate for AAV delivery in therapeutic applications [147]. Mediated by AAV, SaCas9 has successfully edited specific neuronal subpopulations in rats and mutated genes in difficult-to-isolate cell types in mice [148,149]. Apart from the neuron system, the AAV-derived CRISPR-SaCas9, promoted by a liver-specific promoter, is reported to inhibit Hepatitis B virus (HBV) replication both *in vitro* and *in vivo* in mice [150]. Newly discovered compact Cas9 orthologs from *Staphylococcus auricularis* (SauriCas9) and chimeric SlugCas9 (*Staphylococcus lugdunensis* Cas9)-SaCas9 also maintain high activity for genome editing [151,152]. Recently, a batch of Cas9 orthologs has been biochemically identified with extraordinary variations [92]. Advanced studies on these systems may provide more efficient and safer tools for clinical applications in the future.

### *Engineered Cas9 alternatives and variants*

Off-target effects could be detrimental and concerning for gene editing using CRISPR-Cas systems in human cells. Previous studies on the off-target cleavage of guide RNA: Cas9 complex suggested that using a shorter less-active guide RNA or decreasing the RNP concentration would limit the enzyme activity and lead to higher target specificity [153]. Other engineered Cas9 nucleases show diverse improved properties, rapidly expanding the genome editing toolbox. The enhanced SpCas9 variant, eSpCas9 (1.1), maintained a solid on-target cleavage ability but a reduced off-target effect due to the neutralization of some positively charged residues from the non-target groove, which resulted in weaker binding affinity between RNP and DNA substrates [154]. Similarly, four substitutions in SpCas9-HF1 (N497A/R661A/Q695A/Q926A) decreased the energetics of interaction between RNP and the target DNA that the non-specific interaction could be diminished [155]. The hyper-accurate Cas9 (HypaCas9) introduced mutations in the REC3 domain, a part of the REC lobe that can sense mismatch, to avoid cleavage of mismatched DNA substrates due to the improved proofreading before cleavage [156]. This variant has been successfully implemented to edit mouse zygotes to generate allele-specific genetic mutations [157]. Expanding the range of PAM sequences is another way to increase the scope of the application of CRISPR-Cas systems. Many engineered Cas9 nucleases with mutations, including key arginines in the PI domain, could recognize alternative PAM sequences [146]. An engineered Cas9 variant by rational design, SpCas9-NG, was proven to recognize a relaxed NG PAM [158]. Phage-assisted continuous evolution evolved a SpCas9 variant (xCas9) recognizing a broad range of PAMs, including NG, GAA, and GAT [159]. Moreover, this variant has higher target specificity than SpCas9 [159]. Remarkably, xCas9 displayed the lowest tolerance for mismatched target sequences compared to SpCas9 and SpCas9-NG [160]. A near-PAMless SpCas9 variant named SpRY (NRN and, to a lesser extent, NYN PAMs) (Y is C or T; R is A or G) was developed and exhibited robust activities on previously inaccessible targets [161]. By introducing mutants in the hinge region between the HNH and RuvC domains, seven Cas9 variants (G915F, F916P, ΔF916, K918A, R919P, Q920P, R780A) are demonstrated to enable the altered scissile patterns with precision and predictability of nucleotide insertions or specific cleavage events [122].

Similar to SpCas9, SaCas9 was also modified for better applications. A rationally engineered SaCas9 variant (SaCas9-HF) was reported with highly genome-wide specificity in human cells without compromising on-target efficacy [162]. Another variant, KKHSaCas9, could effectively target sites with NNARRT, NNCRRT, and to a lesser extent, NNTRRT, and PAM sequences, thereby broadening the targeting scope by 2–4 times relative to the wild type SaCas9, which recognizes an NNGRRT PAM [163].

### Research and clinical applications of Cas9

Compared to ZFN and TALEN, the RNA-guided CRISPR-Cas9 system offers significant advantages in high programmability and low cost. In 2013, Cong *et al.* [142] and Mali *et al.* [143] independently reported the first successful applications of CRISPR in human cells. Soon after, CRISPR-Cas9 was reported to be successfully applied in the gene-editing of different species of model organisms, including bacterium [164], yeast [165], rice [166], fruit fly [167], zebrafish [168], and mouse [169]. Besides point mutations, insertions, or deletions, applications of CRISPR-Cas9 have also been extended to scenarios including chromosome translocations [170], genome-wide screening [171], gene regulation [172,173], and genomic locus imaging [174]. In medical applications, CRISPR-Cas9 has also shown great potential. As the first clinically proven CRISPR therapy, exagamglogene autotemcel (exa-cel, previously named CTX001), which treats sickle cell disease and  $\beta$ -thalassemia through editing the patient's own stem cells by CRISPR-Cas9 to produce high level of fetal hemoglobin [175], has been submitted to the U.S. FDA for rolling review and is expected to be the first CRISPR-based gene editing therapy. A clinical trial (NCT03164135) in which transplantation of CRISPR-edited CCR5-ablated hematopoietic stem and progenitor cells (HSPCs) was used to treat HIV infection is also reported [176].

### Cas12a

Cas12a is the first described Cas12 with the genome-editing ability [98]. It has been widely used in eukaryotic genome editing [177–179]. The Cas12a orthologs cleave dsDNA with a T-rich PAM: 5'-TTTV-3' (V is A, G, or C) for LbCas12a and AsCas12a; and 5'-TTV-3' for FnCas12a [98]. To expand the targeting scope of Cas12a, Zhang's group [180] performed a structure-guided mutagenesis screen and successfully engineered two AsCas12a variants: S542R/K607R and S542R/K548V/N552R, which recognized TYCV and TATV PAMs. Inspired by these designs, Welker's group [181] engineered an improved LbCas12a (im-pLbCas12a) with altered PAM specificities by combining the RR and RVR mutations in Zhang's work [180]. Joung's group [182] engineered an enhanced AsCas12a (enCas12a) with a substantially expanded PAM scope, including canonical TTTV and non-canonical TTYN, VTTN, and TRTV. Besides, they also developed a high-fidelity version of enAsCas12a (enAsCas12a-HF1) with a lower off-target effect [182].

Compared to SpCas9, Cas12 is generally more specific, for it has lower mismatch tolerance for gene therapy [183,184]. The staggered ends in DNA post-cleavage make Cas12a useful for genetic manipulation via the HDR pathway [140]. Another advantage of using Cas12a is the simplicity of multiplex targeting by introducing a single crRNA array with multiple targeting sites, albeit only feasible in bacteria so far [185].

### Cas12b

CRISPR-Cas12b is a dual-guide DNA endonuclease system with minimal off-target effects [186]. However, 37°C in mammals is unsuitable for efficient dsDNA cleavage by AacCas12b from *Alicyclobacillus acidoterrestris* ATCC 4902 in the early study [99]. Thus, researchers have been devoted to discovering and engineering more efficient Cas12b orthologs for mammalian genome editing. Teng *et al.* [187] reported a thermo-acidophilic Cas12b nuclease from *Alicyclobacillus acidophilus* (AaCas12b) that maintained maximal

nuclease activity between 31°C and 59°C. Moreover, AaCas12b enabled robust genome editing in mammalian cell lines and mice [187]. Strecker *et al.* [188] identified a mutant BhCas12b from *Bacillus hisashii* that exhibited robust genome editing in primary human T cells with greater specificity compared to SpCas9. Ming *et al.* [189] demonstrated that AaCas12b enabled multiplexed genome editing in rice with high sequence specificity. A vital advantage of these systems is that their high mismatch sensitivity reduces off-target cleavage, which makes them suitable for target recognition in base editors. Besides, engineering either Cas protein or RNA scaffold would help to evolve the system with higher editing efficacy in mammalian cells.

### **Cas12e**

Given that the well-established AAV delivery *in vivo* is limited by DNA packaging size, CRISPR-Cas systems with compact effector proteins have been intensively described. CRISPR-Cas12e (also denoted CasX) is a more compact dual-RNA-guided tool for gene editing [107]. Both chemical and structural features of 2 known Cas12e systems, DpbCas12e and PlmCas12e, have been elucidated in detail [103,132]. Among these studies, structure-based domain switches between Cas proteins and rational gRNA modification significantly improved the genome editing efficiency in mammalian cells [132]. However, the editing efficiency at different targeting sites varies a lot, and almost half of the sites cannot be edited robustly [132]. One possible reason is that some spacer sequences interfere with the proper folding of sgRNA and reduce RNP complexes in cells. In other words, the guide RNA scaffold of the CRISPR-Cas12e system is less stable before RNP formation or R-Loop formation. Therefore, efforts in guide RNA engineering to overcome this problem are necessary.

### **Cas12f**

The Type V-F effectors, Cas12f (originally denoted Cas14), are hypercompact Cas proteins ranging between 400–700 amino acid residues. So far, only 2 natural Cas12f1 nucleases, SpCas12f1 (497 aa) and AsCas12f1 (422 aa), have been characterized with genome editing ability in eukaryotic cells [190,191]. And AsCas12f1 is superior in human genome editing [190]. While the firstly characterized Un1Cas12f was unable to edit bacteria or mammalian genome, guide RNA and protein engineering turned it into an efficient gene editing tool in mammalian cells [192,193]. Interestingly, Wang *et al.* [194] transformed SpCas12f1, which was less efficient for mammalian genome editing earlier, into a robust nuclease comparable to FnCas12a by guide RNA engineering. These transformations of editing capacity do expand our understanding of the Cas-RNA complex. Although Cas12f homologs prefer low salt concentration and higher temperatures (45°C–55°C) for target dsDNA binding and cleavage [104], the formation of stable RNP might be a key factor for successful dsDNA targeting. Given that CRISPR-Cas12f systems have remarkably long tracrRNAs and function as a Cas-dimer upon one single RNA, the properly folded guide RNA is vital for the whole process. The interaction between modified RNA and Cas protein may overcome some thermo-kinetic requirements for conformational changes in Cas protein. Additionally, RNA engineering involves improving RNA transcription levels to promote the RNP constitution in cells. As a tool for genome editing, Cas12f systems contain advantages both in editing efficiency as high as Cas12a or Cas9 at sites indicated and in size-limited

AAV-delivery therapy.

### ***Class I: Transposon-associated CRISPR-Cas system***

Several Cas proteins form a multi-subunit effector complex in Class I systems for crRNA binding and target interference. These systems utilize a number of nuclease activities for the cleavage of dsDNA, ssDNA, and RNA. Multiple subunits in Type I assemble across a single crRNA to form a seahorse-like architecture. Once PAM is recognized, the local bending of dsDNA followed by base-pairing along the crRNA will unwind the dsDNA to form an R-loop structure “Cascade” (CRISPR-associated complex for antiviral defense) complex, which further recruits Cas3, a helicase-nuclease effector protein for target dsDNA degradation. Notably, some minimal Type I-F and Type I-B systems that lack the adaptation module and Cas3 are associated with Tn7-like transposons composed of TnsA/B/C and TniQ [195]. Thus, it was hypothesized that these systems recognize but do not cleave the dsDNA to allow the transfer of transposons. It was demonstrated by studies on Type I-F CRISPR system from *Vibrio cholerae* that cargo DNA can be integrated into the site around 50 bp downstream of the DNA target recognized by the Cascade complex [196]. This insertion of transposable elements by guide RNA-assisted targeting (INTEGRATE) is further engineered for multiplexed, kilobase-scale genome integration in bacteria [197]. Recently, Jinek’s group [198] resolved the structure of Cascade-TniQ and found that the TnsC heptamerized upon binding ATPs and wrapped around the DNA, which explained the specificity of the insertion site of the exogenous DNA (Figure 5B). Similarly, two Tn7-like transposons encoding subtype V-K CRISPR-Cas systems termed CASTs, or CRISPR-associated transposase, were also demonstrated to be directed to target sites by CRISPR RNA and insert DNA into the genome of *E. coli* [188]. In this context, a naturally inactive Cas12k with a RuvC-like domain corresponds to the lack of Cas3 effector protein in Type I. Although these systems successfully insert cargo fragments, none were found to function in mammalian cells. Besides, the large size of Cascade in INTEGRATE from Type I and relatively low specificity in CRISPR-associated transposase from Type V-K are major challenges for further application in mammalian cells. Given that a single effector protein is involved in target recognition in Class II CRISPR systems, deeper exploration for various transposon-associated systems in Class II may provide utility for gene insertion in mammalian cells. In addition, natural eukaryotic transposases might be co-opted for site-specific gene insertion guided by the CRISPR complex in mammalian cells.

### ***More compact RNA-guided protein: IscB and TnpB***

Recently, a brief system composed of IscB, likely an ancestor of Cas9, and associated noncoding RNA,  $\omega$ RNA, was identified [199]. Co-expression of IscB and  $\omega$ RNA was capable of cleaving DNA complementary to the putative RNA guide in a target-adjacent motif [94] dependent way [199]. Among the six IscB candidates, OgeuIscB (496 aa) exhibited varying editing efficiency of up to 4.5% at 28 out of 46 sites in the human genome [199]. Another transposon-encoded nuclease: ISDra2 TnpB (408 aa) from *Deinococcus radiodurans*, guided by ~230 nt RNA transcript derived from right-end element RNA (reRNA), was also demonstrated to induce genomic DNA cleavage both in bacteria and human cells [200]. The structure of IscB- $\omega$ RNA-dsDNA has been determined [201], whereas the structure of TnpB is not available (Figure 5C). IscB and TnpB are thought to be the ancestors of Cas9 and Cas12, respectively, since they have highly



conserved RuvC-like domain and/or HNH domain [199]. Newly found RNA-guided nuclease activities of IscB and TnpB strongly supported this theory. Structure-based analysis can reveal their evolutionary relationships more directly and help us optimize the editing systems. These small transposase-derived systems show huge potential as genome editing tools that can be packaged into AAV for therapeutic applications. Based on our RNA engineering trials in Cas12b, Cas12e, and especially Cas12f, modifications in RNA are likely to show higher editing efficacy *in vivo*. This highlights the importance of studying RNA structure in these systems for further engineering.

### ***Precise gene editing: base editing (BE) and prime editing (PE)***

Genome editing based on the Cas-gRNA complex relies on introducing DSBs on the locus to generate random indels, while base editing technologies enable us to rewrite DNA more precisely. Liu's group [202] engineered the fusion of dSpCas9 and a cytidine deaminase enzyme that could mediate the conversion from cytidine to uridine within a small window. This mis-pair was then turned into desired T-A pair via cellular repair processes [202]. Among the designs in the study, BE3 construction (APOBEC-XTEN-dCas9 (A840H)-UGI) could induce efficient single-base editing in human cells [202]. A-T pair can also be converted to G-C pair by adenine base editors (ABEs) that fuse dCas9 with various evolved adenine deaminases [203] (Figure 5D, left). Gao's group [204] has successfully developed or optimized new base editors for precise editing in plants, especially crops, since 2017. Fusion of Cas9 nickase and human APOBEC3A (A3A-PBE) can efficiently convert C to T in wheat, rice, and potato. Another plant adenine base editor is based on the fusion of nCas9 and an evolved tRNA adenosine deaminase, which shows A•T to G•C conversion at frequencies of up to 59% in rice and wheat [205].

Recently, a more versatile genome-editing tool, prime editing, has expanded the scope for single-base transversions and transitions [206]. This strategy fuses an impaired Cas9 with an engineered reverse transcriptase (RT), which generates desired DNA sequence taking prime guide RNA (pegRNA) as the template [206] (Figure 5D, right). Using an engineered Moloney murine leukemia virus (M-MLV) RT variant fused with nCas9 (H840A) and primer binding sites in proper length exhibits the most efficient editing results [206]. Engineered pegRNAs (epegRNAs), which incorporate structured RNA patterns at the 3' ends of pegRNAs, can improve plasmid editing efficiency by 3–4 fold in human cells without increasing off-target editing activity, as this modification enhances gRNA stability and prevents degradation of the 3' extension [207]. Other engineered modifications on PE have proven effective, including untethering Cas9 from the reverse transcriptase [208], splitting the pegRNA into a single guide RNA and a circular RNA RT template [208], and compacting the PE by deleting the RNase H domain of RT with or without a split site on Cas9 [209].

Instead of using the CRISPR-Cas system, Liu's group [210] developed a DddA-derived cytosine base editors (DdCBE) system that uses TALE proteins to specifically target sites for base editing on human mitochondrial DNA. The successful editing by DdCBEs has been reported in human embryos, mice, zebrafish, and plants [211–216]. Soon after, Kim's group [217] developed DdABE combining DdCBE and adenine base editor ABE8e to enable adenine base editing of mitochondrial DNA. Recently, through phage-assisted directed evolution, they further evolved DddA variants with enhanced editing efficiency in mitochondrial and nuclear DNA, overcoming the constraints imposed by target cytosine within the spacing

region between the two target protospacers [218]. Although DdCBE is a promising approach for the treatment of mitochondrial diseases, extensive off-target editing has been observed in the nuclear genome [219], which has promoted further efforts to mitigate such off-targets.

Since DSBs harmful to cells are avoided during the editing process theoretically, base editing and prime editing have a higher safety profile for medical applications. In addition, the cell division-independent nature of prime editing allows it to be used for the genetic modification of non-dividing cells. Although both the base editor and prime editor face the packaging constraints as other less compact CRISPR-Cas systems, these technologies can be appropriately applied for genome editing in plants and mice [220–222]. Base editors have been successfully implemented in rice, wheat, and tomato [223–225]. More technological developments and biological applications in plants are reviewed in this article [222]. This review [226] summarizes the research progress of base editing and the application of this technology in medical therapeutics before 2021. Last year, Liu's team [227] developed an upgraded version of pilot editing technology, twin prime editing (twinPE), which consists of a prime editor protein and two pegRNAs to cleave target DNA at different sites creating two single-stranded nicks. Combined with a site-specific recombinase, this editor successfully inverted the sequence of about 40,000 base pairs and also inserted a plasmid of more than 5000 bp in length, up to the gene size, into the specific site in the human genome [227]. Gao's group [225] is the first to successfully establish and optimize the prime editor in plants (PPE). They found that the editing efficacy could be increased by optimizing the expression vector of pegRNA, the length of PBS and template for RT, and editing conditions. Further, they substantially enhanced the efficiency by combining designing PBS with a melting temperature of 30°C and using two pegRNAs in rice [228]. Another strategy is to optimize the reverse transcriptase. Gao's group proved that deleting the RNase H domain of M-MLV RT or fusing the viral nucleocapsid protein at the N terminal of M-MLV RT could significantly improve the PE efficiency [229]. All these pieces of work have laid a good foundation for applying PE in plants [230].

## **Concluding remarks and future perspectives**

In this review, we have discussed the history and application scenarios of four gene editing technologies. The structural mechanisms of targeted DNA recognition by different nucleases are briefly described. Updates in gene editing technologies have been accompanied by an easier implementation of targeting specific sequences, from cooperative recognition patterns between amino acid residues to modular base recognition motifs, then to gRNA recognition through base complementary pairing. ZFN and TALEN target specific sites by tandem DNA recognition modules. Although relatively complicated to be constructed, they have acquired an essential position in medical applications due to high specificity. Furthermore, such entirely protein-based editing systems have unique advantages in the mitochondrial editing scenario where delivery of nucleic acids is quite difficult [210,231,232].

CRISPR-Cas systems have been extensively adopted and successfully implemented in numerous fundamental studies in eukaryotic cells and mammals. Some clinical trials using Cas9 seem promising so far. Obviously, the CRISPR-Cas system has been a dominant strategy for gene editing. In future, the feasibility of using precise editing tools for clinical applications should be emphasized and developed. This highlights the importance of studying those compact systems and the delivery vector used for gene therapy. Although

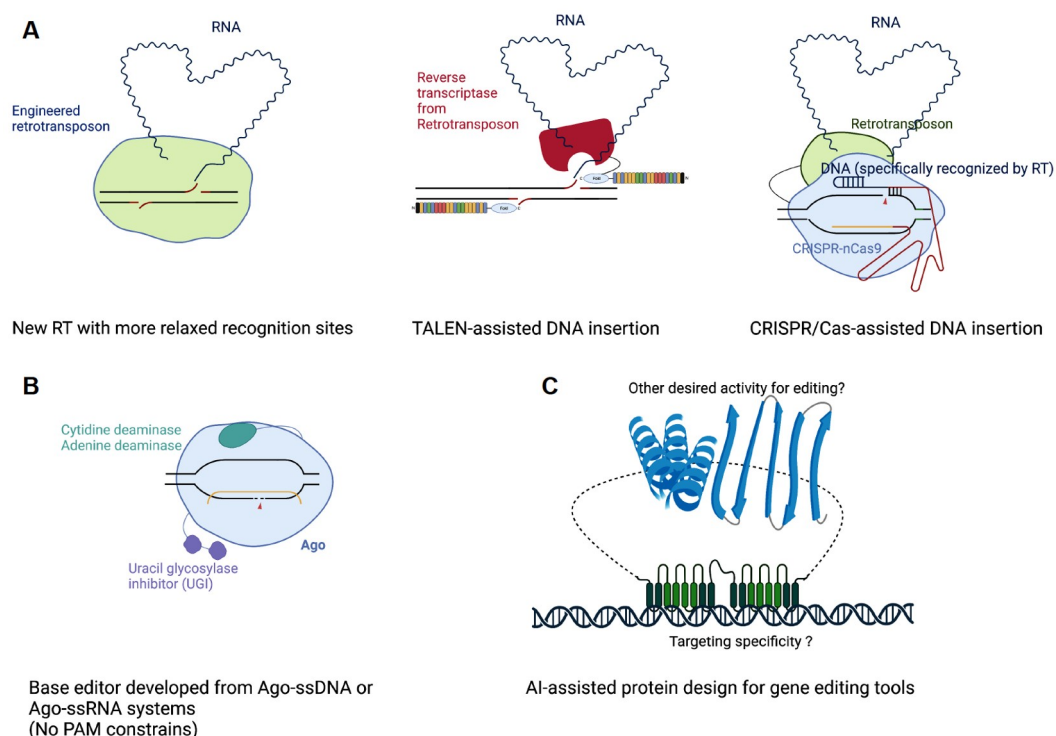
various gene-editing tools have been developed, their efficiency, off-target effects, and feasibility of delivery remain major issues for clinical treatments. Precise editing has become a major direction for the future of gene editing. Still, current precise editing tools require more effort to simplify components, reduce the size, and improve editing efficiency while decreasing off-target effects. In fundamental research, how to select a proper editing tool should depend on our major aims. Each strategy needs to be carefully evaluated and optimized, especially when using base editors or prime editors.

Given the never-ending optimization of the CRISPR-Cas system, new gene editing systems, naturally occurring or engineered, may substantially improve gene editing. The numerous varieties of biology resources provide the endless possibility to develop more advanced tools. Can we find or evolve a single chimeric protein with all the activities required for base editing or prime editing? Or could we do this simply with a single RNA, DNA, or by using an intrinsic mechanism and modifying it slightly? In addition, other types of nucleic acid-directed or binding systems could be exploited for gene editing.

Gene editing by transposons like INTEGRATE and CAST has emerged as a promising tool to achieve site-specific insertion of long sequences. However, it is still challenging to enable DNA insertion in human cells. In addition to insertion via the “cut and paste” mechanism, we can achieve this in human cells by reverse transcription, given that retrotransposons are abundant in eukaryotes. However, the loci they recognize are highly specific and conserved. Perhaps, we can explore new retrotransposon-derived reverse transcriptase with substantial locus tolerance to achieve targeted insertion in combination with TALEN (Figure 6A, middle). Alternatively, using the CRISPR-Cas system that inserts specific motifs before retrotransposons recognize them may also help (Figure 6A, right). On the other hand, we can engineer the domain responsible for target-site recognition and develop suitable reverse transcriptase with high engineerability (Figure 6A, left).

A possible direction for gene editing technology is to exploit DNA-guided systems. Some prokaryotic Argonaute (pAgo) protein members, which exhibit DNA endonuclease activity under the guidance of short ssDNA at complementary pairing sites, are potential candidates [233]. Gene editing based on pAgo is more flexible because there is no such requirement for PAM in this system as in CRISPR. The convenient synthesis of short guide DNA simplifies the cloning and delivery procedure. A retracted study claiming to achieve gene editing in mammalian cells based on *Natronobacterium gregoryi* Argonaute (NgAgo) [234], however, was not reproducible by other groups [235]. The exploration of new pAgos or engineering current members into other functions like base editing (Figure 6B) remains a potential direction for the development of novel gene editing technologies. Another type of DNA-guided gene editing tool, structure-guided nuclease (SGN), consists of a hairpin DNA probe and a flap structure-specific endonuclease 1 (FEN1)-*Fok I* effector [236,237]. FEN1 recognizes the 3' flap structure formed by hairpin DNA binding to the target site and guides *Fok I* cleavage, resulting in decent editing efficacy in both bacterial and human cells [237]. More comprehensive assessments, including a broader range of targeting sequences across different model organisms and cytotoxicity, are still required.

The specificity of target-site recognition can be achieved by artificial intelligence (AI)- assisted modular design of proteins capable of binding nucleic acids, similar to ZFN and TALEN technology. In recent years, many achievements have been made using AI for protein-structure predictions and structure-based protein designs. Recent studies have shown that new machine-learning algorithms could design proteins faster and more accurately than before [238–240]. Designing new DNA or RNA binding proteins may rely on extensive



**Figure 6** The potential engineered tools for genome editing. (A) Combining the reverse transcriptases (RTs) from retrotransposon with TALEN or CRISPR system may overcome the limitation of specific recognizing motifs by RTs. Or we can evolve new RTs that have less strict recognizing motifs. (B) The Ago system can provide specific targeting for precise editing, relying on the complementation of ssDNA or RNA with the locus regardless of PAM. Compared with CRISPR, its relatively weak cleavage ability will reduce the hazards of off-target cleavage events. (C) AI-assisted protein design may be able to generate proteins with higher targeting specificity, efficiency, and low redundancy.

machine learning and screening. With AI-assisted design, we may develop more compact, diverse, and specific nucleic acid-binding proteins (Figure 6C). Although it sounds feasible, the difficulty of designing such proteins with high specificity and efficiency has not been evaluated.

## Funding

This work was supported by the Ministry of Agriculture and Rural Affairs of China, the National Natural Science Foundation of China (32150018), and start-up funds from Tsinghua University, Beijing (J.J.G.L.).

## Author contributions

J.J.G.L. conceived the structure and logic of the manuscript. D.Y.L, L.Q.L, and J.J.G.L. wrote the manuscript. D.Y.L. and L.Q.L. prepared the figures.

## Conflict of interest

The authors declare no conflict of interest.

## References

- 1 Smithies O, Gregg RG, Boggs SS, *et al*. Insertion of DNA sequences into the human chromosomal  $\beta$ -globin locus by

- homologous recombination. *Nature* 1985; **317**: 230–234.
- 2 Scherer S, Davis RW. Replacement of chromosome segments with altered DNA sequences constructed *in vitro*. *Proc Natl Acad Sci USA* 1979; **76**: 4951–4955.
- 3 Capecchi MR. The new mouse genetics: Altering the genome by gene targeting. *Trends Genet* 1989; **5**: 70–76.
- 4 Orr-Weaver TL, Szostak JW. Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc Natl Acad Sci USA* 1983; **80**: 4417–4421.
- 5 Rouet P, Smih F, Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol*, 1994; **14**: 8096–8106.
- 6 Rudin N, Sugarman E, Haber JE. Genetic and physical analysis of double-strand break repair and recombination in *Saccharomyces cerevisiae*. *Genetics* 1989; **122**: 519–534.
- 7 Belfort M. Homing endonucleases: keeping the house in order. *Nucleic Acids Res* 1997; **25**: 3379–3388.
- 8 Jacquier A, Dujon B. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* 1985; **41**: 383–394.
- 9 Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* 1996; **12**: 224–228.
- 10 Stoddard BL. Homing endonuclease structure and function. *Q Rev Biophys*, 2005; **38**: 49–95.
- 11 Roberts RJ. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 2003; **31**: 1805–1812.
- 12 Moure CM, Gimble FS, Quijcho FA. The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J Mol Biol* 2003; **334**: 685–695.
- 13 Chevalier BS. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* 2001; **29**: 3757–3774.
- 14 Smith J, Grizot S, Arnould S, *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res* 2006; **34**: e149.
- 15 Arnould S, Perez C, Cabaniols JP, *et al.* Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol* 2007; **371**: 49–65.
- 16 Gao H, Smith J, Yang M, *et al.* Heritable targeted mutagenesis in maize using a designed endonuclease. *Plant J* 2010; **61**: 176–187.
- 17 Ménoret S, Fontanière S, Jantz D, *et al.* Generation of *Rag1*-knockout immunodeficient rats and mice using engineered meganucleases. *FASEB J* 2013; **27**: 703–711.
- 18 Berg JM. Zinc finger domains: hypotheses and current knowledge. *Annu Rev Biophys Chem* 1990; **19**: 405–421.
- 19 Pabo CO, Peisach E, Grant RA. Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem* 2001; **70**: 313–340.
- 20 Elrod-Erickson M, Rould MA, Nekudova L, *et al.* Zif268 protein-DNA complex refined at 1.6Å: a model system for understanding zinc finger-DNA interactions. *Structure* 1996; **4**: 1171–1180.
- 21 Wolfe SA, Grant RA, Elrod-Erickson M, *et al.* Beyond the “Recognition Code”. *Structure* 2001; **9**: 717–723.
- 22 Segal DJ, Crotty JW, Bhakta MS, *et al.* Structure of aart, a designed six-finger zinc finger peptide, bound to DNA. *J Mol Biol* 2006; **363**: 405–421.
- 23 Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to *Fok* I cleavage domain. *Proc Natl Acad Sci USA* 1996; **93**: 1156–1160.
- 24 Bibikova M, Carroll D, Segal DJ, *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* 2001; **21**: 289–297.
- 25 Bitinaite J, Wah DA, Aggarwal AK, *et al.* *Fok* I dimerization is required for DNA cleavage. *Proc Natl Acad Sci USA* 1998; **95**: 10570–10575.
- 26 Liu Q, Segal DJ, Ghiara JB, *et al.* Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci USA* 1997; **94**: 5525–5530.



- 27 Bibikova M, Golic M, Golic KG, *et al.* Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics* 2002; **161**: 1169–1175.
- 28 Doyon Y, McCammon JM, Miller JC, *et al.* Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* 2008; **26**: 702–708.
- 29 Meng X, Noyes MB, Zhu LJ, *et al.* Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* 2008; **26**: 695–701.
- 30 Perez-Pinera P, Ousterout DG, Brown MT, *et al.* Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res* 2012; **40**: 3741–3752.
- 31 Zhang F, Maeder ML, Unger-Wallace E, *et al.* High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc Natl Acad Sci USA* 2010; **107**: 12028–12033.
- 32 Urnov FD, Miller JC, Lee YL, *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 2005; **435**: 646–651.
- 33 Tebas P, Stein D, Tang WW, *et al.* Gene editing of *CCR5* in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* 2014; **370**: 901–910.
- 34 Geurts AM, Cost GJ, Freyvert Y, *et al.* Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* 2009; **325**: 433.
- 35 Liu X, Wang Y, Tian Y, *et al.* Generation of mastitis resistance in cows by targeting human lysozyme gene to  $\beta$ -casein locus using zinc-finger nucleases. *Proc Biol Sci*, 2014; **281**: 20133368.
- 36 Segal DJ, Dreier B, Beerli RR, *et al.* Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci USA* 1999; **96**: 2758–2763.
- 37 Dreier B, Segal DJ, Barbas Iii CF. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* 2000; **303**: 489–502.
- 38 Dreier B, Fuller RP, Segal DJ, *et al.* Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA Sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 2005; **280**: 35588–35597.
- 39 Dreier B, Beerli RR, Segal DJ, *et al.* Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 2001; **276**: 29466–29478.
- 40 Beerli RR, Barbas Iii CF. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* 2002; **20**: 135–141.
- 41 Bae KH, Do Kwon Y, Shin HC, *et al.* Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol* 2003; **21**: 275–280.
- 42 Mandell JG, Barbas CF. Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* 2006; **34**: W516–W523.
- 43 Cornu TI, Thibodeau-Beganny S, Guhl E, *et al.* DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol Ther* 2008; **16**: 352–358.
- 44 Pruett-Miller SM, Connelly JP, Maeder ML, *et al.* Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther* 2008; **16**: 707–717.
- 45 Isalan M, Choo Y, Klug A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci USA* 1997; **94**: 5617–5621.
- 46 Maeder ML, Thibodeau-Beganny S, Osiak A, *et al.* Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* 2008; **31**: 294–301.
- 47 Isalan M. Zinc-finger nucleases: How to play two good hands. *Nat Methods*, 2011; **9**: 32–34.
- 48 Bonas U, Stall RE, Staskawicz B. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol Gen Genet* 1989; **218**: 127–136.
- 49 Boch J, Bonas U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* 2010; **48**: 419–436.
- 50 Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* 2009; **326**: 1501.

- 51 Boch J, Scholze H, Schornack S, *et al.* Breaking the code of DNA binding specificity of TAL-Type III effectors. *Science* 2009; **326**: 1509–1512.
- 52 Streubel J, Blücher C, Landgraf A, *et al.* TAL effector RVD specificities and efficiencies. *Nat Biotechnol* 2012; **30**: 593–595.
- 53 Yang J, Zhang Y, Yuan P, *et al.* Complete decoding of TAL effectors for DNA recognition. *Cell Res* 2014; **24**: 628–631.
- 54 Deng D, Yan C, Pan X, *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 2012; **335**: 720–723.
- 55 Mak ANS, Bradley P, Cernadas RA, *et al.* The crystal structure of TAL Effector PthXo1 bound to its DNA target. *Science* 2012; **335**: 716–719.
- 56 Deng D, Yan C, Wu J, *et al.* Revisiting the TALE repeat. *Protein Cell* 2014; **5**: 297–306.
- 57 Christian M, Cermak T, Doyle EL, *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 2010; **186**: 757–761.
- 58 Wood AJ, Lo TW, Zeitler B, *et al.* Targeted genome editing across species using ZFNs and TALENs. *Science* 2011; **333**: 307.
- 59 Liu J, Li C, Yu Z, *et al.* Efficient and specific modifications of the drosophila genome by means of an easy TALEN strategy. *J Genet Genomics* 2012; **39**: 209–215.
- 60 Sander JD, Cade L, Khayter C, *et al.* Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* 2011; **29**: 697–698.
- 61 Christian M, Qi Y, Zhang Y, *et al.* Targeted mutagenesis of *Arabidopsis thaliana* using engineered TAL effector nucleases. *G3 Genes|Genomes|Genet* 2013; **3**: 1697–1705.
- 62 Ousterout DG, Perez-Pinera P, Thakore PI, *et al.* Reading frame correction by targeted genome editing restores dystrophin expression in cells from duchenne muscular dystrophy patients. *Mol Ther* 2013; **21**: 1718–1726.
- 63 Bloom K, Ely A, Mussolino C, *et al.* Inactivation of hepatitis B virus replication in cultured cells and *in vivo* with engineered transcription activator-like effector nucleases. *Mol Ther* 2013; **21**: 1889–1897.
- 64 Cermak T, Doyle EL, Christian M, *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 2011; **39**: e82.
- 65 Briggs AW, Rios X, Chari R, *et al.* Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res* 2012; **40**: e117.
- 66 Reyon D, Tsai SQ, Khayter C, *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 2012; **30**: 460–465.
- 67 Schmid-Burgk JL, Schmidt T, Kaiser V, *et al.* A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* 2013; **31**: 76–81.
- 68 Yang J, Yuan P, Wen D, *et al.* Ultimate system for rapid assembly of customized tal effectors. *PLoS One*, 2013; **8**: e75649.
- 69 Liang J, Chao R, Abil Z, *et al.* FairyTALE: A high-throughput TAL effector synthesis platform. *ACS Synth Biol* 2014; **3**: 67–73.
- 70 Kim Y, Kweon J, Kim A, *et al.* A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* 2013; **31**: 251–258.
- 71 Mussolino C, Morbitzer R, Lütge F, *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* 2011; **39**: 9283–9293.
- 72 Marraffini LA. CRISPR-Cas immunity in prokaryotes. *Nature* 2015; **526**: 55–61.
- 73 Ishino Y, Shinagawa H, Makino K, *et al.* Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 1987; **169**: 5429–5433.
- 74 Jansen R, Embden JDA, Gaastra W, *et al.* Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002; **43**: 1565–1575.

- 75 Mojica FJM, Díez-Villaseñor C, García-Martínez J, *et al.* Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 2005; **60**: 174–182.
- 76 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**: 653–663.
- 77 Bolotin A, Quinquis B, Sorokin A, *et al.* Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 2005; **151**: 2551–2561.
- 78 Barrangou R, Fremaux C, Deveau H, *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007; **315**: 1709–1712.
- 79 Sapranas R, Gasiunas G, Fremaux C, *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 2011; **39**: 9275–9282.
- 80 Jinek M, Chylinski K, Fonfara I, *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; **337**: 816–821.
- 81 Gasiunas G, Barrangou R, Horvath P, *et al.* Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 2012; **109**: E2579–2586.
- 82 Hille F, Richter H, Wong SP, *et al.* The biology of CRISPR-Cas: backward and forward. *Cell* 2018; **172**: 1239–1259.
- 83 Zhou Y, Bravo JPK, Taylor HN, *et al.* Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *iScience* 2021; **24**: 102201.
- 84 Sundaresan R, Parameshwaran HP, Yogesha SD, *et al.* RNA-independent DNA cleavage activities of Cas9 and Cas12a. *Cell Rep* 2017; **21**: 3728–3739.
- 85 Abudayyeh OO, Gootenberg JS, Konermann S, *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 2016; **353**: aaf5573.
- 86 Koonin EV, Makarova KS. Evolutionary plasticity and functional versatility of CRISPR systems. *PLoS Biol* 2022; **20**: e3001481.
- 87 Makarova KS, Zhang F, Koonin EV. SnapShot: Class 1 CRISPR-Cas systems. *Cell* 2017; **168**: 946–946.e1.
- 88 Xiao Y, Luo M, Hayes RP, *et al.* Structure basis for directional R-loop formation and substrate handover mechanisms in type I CRISPR-Cas system. *Cell* 2017; **170**: 48–60.e11.
- 89 Rutkauskas M, Sinkunas T, Songailiene I, *et al.* Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep* 2015; **10**: 1534–1543.
- 90 Sinkunas T, Gasiunas G, Fremaux C, *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J* 2011; **30**: 1335–1342.
- 91 Hochstrasser ML, Taylor DW, Bhat P, *et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci USA* 2014; **111**: 6618–6623.
- 92 Gasiunas G, Young JK, Karvelis T, *et al.* A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nat Commun* 2020; **11**: 5512.
- 93 Niewoehner O, Garcia-Doval C, Rostøl JT, *et al.* Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* 2017; **548**: 543–548.
- 94 Kazlauskienė M, Kostiuk G, Venclovas Č, *et al.* A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* 2017; **357**: 605–609.
- 95 Taylor HN, Laderman E, Armbrust M, *et al.* Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. *Front Microbiol* 2021; **12**: 671522.
- 96 Pinilla-Redondo R, Mayo-Muñoz D, Russel J, *et al.* Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res* 2020; **48**: 2000–2012.
- 97 Makarova KS, Zhang F, Koonin EV. SnapShot: Class 2 CRISPR-Cas systems. *Cell* 2017; **168**: 328–328.e1.
- 98 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**: 759–771.
- 99 Makarova KS, Wolf YI, Iranzo J, *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and

- derived variants. *Nat Rev Microbiol* 2020; **18**: 67–83.
- 100 Sun A, Li CP, Chen Z, *et al.* The compact Cas $\pi$  (Cas12I) ‘bracelet’ provides a unique structural platform for DNA manipulation. *Cell Res* 2023; **33**: 229–244.
- 101 Urbaitis T, Gasiunas G, Young JK, *et al.* A new family of crispr-type v nucleases with c-rich pam recognition. *EMBO Rep*, 2022; **23**: e55481.
- 102 Shmakov S, Abudayyeh OO, Makarova KS, *et al.* Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol Cell* 2015; **60**: 385–397.
- 103 Liu JJ, Orlova N, Oakes BL, *et al.* CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* 2019; **566**: 218–223.
- 104 Karvelis T, Bigelyte G, Young JK, *et al.* PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res* 2020; **48**: 5016–5023.
- 105 Yan WX, Hunnewell P, Alfonse LE, *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* 2019; **363**: 88–91.
- 106 Pausch P, Al-Shayeb B, Bisom-Rapp E, *et al.* CRISPR-Cas $\Phi$  from huge phages is a hypercompact genome editor. *Science* 2020; **369**: 333–337.
- 107 Burstein D, Harrington LB, Strutt SC, *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* 2017; **542**: 237–241.
- 108 Al-Shayeb B, Skopintsev P, Soczek KM, *et al.* Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. *Cell* 2022; **185**: 4574–4586.e16.
- 109 Huang CJ, Adler BA, Doudna JA. A naturally DNase-free CRISPR-Cas12c enzyme silences gene expression. *Mol Cell* 2022; **82**: 2148–2160.e4.
- 110 Strecker J, Ladha A, Gardner Z, *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 2019; **365**: 48–53.
- 111 Thompson MK, Sobol RW, Prakash A. Exploiting DNA endonucleases to advance mechanisms of DNA repair. *Biology* 2021; **10**: 530.
- 112 Marino ND, Pinilla-Redondo R, Bondy-Denomy J. CRISPR-Cas12a targeting of ssDNA plays no detectable role in immunity. *Nucleic Acids Res* 2022; **50**: 6414–6422.
- 113 Fuchs RT, Curcuru J, Mabuchi M, *et al.* Cas12a trans-cleavage can be modulated *in vitro* and is active on ssDNA, dsDNA, and RNA. *bioRxiv*, 2019, doi: 10.1101/600890.
- 114 Anders C, Niewoehner O, Duerst A, *et al.* Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 2014; **513**: 569–573.
- 115 Nishimasu H, Ran FA, Hsu PD, *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014; **156**: 935–949.
- 116 Globy V, Lee SH, Bae T, *et al.* CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J* 2019; **38**: e99466.
- 117 Yang M, Sun R, Deng P, *et al.* Nonspecific interactions between SpCas9 and dsDNA sites located downstream of the PAM mediate facilitated diffusion to accelerate target search. *Chem Sci* 2021; **12**: 12776–12784.
- 118 Pacesa M, Loeff L, Querques I, *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature* 2022; **609**: 191–196.
- 119 Chen JS, Doudna JA. The chemistry of Cas9 and its CRISPR colleagues. *Nat Rev Chem* 2017; **1**: 0078.
- 120 Zuo Z, Liu J. Cas9-catalyzed DNA cleavage generates staggered ends: evidence from molecular dynamics simulations. *Sci Rep* 2016; **6**: 37584.
- 121 Jinek M, Jiang F, Taylor DW, *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 2014; **343**: 1247997.
- 122 Shou J, Li J, Liu Y, *et al.* Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol Cell* 2018; **71**: 498–509.e4.

- 123 Wu Q, Shou J. Toward precise CRISPR DNA fragment editing and predictable 3D genome engineering. *J Mol Cell Biol* 2021; **12**: 828–856.
- 124 Shen MW, Arbab M, Hsu JY, *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 2018; **563**: 646–651.
- 125 Allen F, Crepaldi L, Alsinet C, *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol* 2019; **37**: 64–72.
- 126 Yeh CD, Richardson CD, Corn JE. Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol* 2019; **21**: 1468–1478.
- 127 Yamano T, Nishimasu H, Zetsche B, *et al.* Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 2016; **165**: 949–962.
- 128 Dong D, Ren K, Qiu X, *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 2016; **532**: 522–526.
- 129 Nishimasu H, Cong L, Yan WX, *et al.* Crystal structure of *Staphylococcus aureus* Cas9. *Cell* 2015; **162**: 1113–1126.
- 130 Hirano H, Gootenberg JS, Horii T, *et al.* Structure and Engineering of *Francisella novicida* Cas9. *Cell* 2016; **164**: 950–961.
- 131 Gao P, Yang H, Rajashankar KR, *et al.* Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res* 2016; **26**: 901–913.
- 132 Tsuchida CA, Zhang S, Doost MS, *et al.* Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity. *Mol Cell* 2022; **82**: 1199–1209.e6.
- 133 Pausch P, Soczek KM, Herbst DA, *et al.* DNA interference states of the hypercompact crispr-cas $\phi$  effector. *Nat Struct Molecular Biology*, 2021; **28**: 652–661.
- 134 Jeon Y, Choi YH, Jang Y, *et al.* Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nat Commun* 2018; **9**: 2777.
- 135 Swarts DC, van der Oost J, Jinek M. Structural basis for guide RNA processing and seed-dependent DNA Targeting by CRISPR-Cas12a. *Mol Cell* 2017; **66**: 221–233.e4.
- 136 Swarts DC, Jinek M. Mechanistic Insights into the *cis*- and *trans*-Acting DNase Activities of Cas12a. *Mol Cell* 2019; **73**: 589–600.e4.
- 137 Zhang L, Sun R, Yang M, *et al.* Conformational dynamics and cleavage sites of Cas12a are modulated by complementarity between crRNA and DNA. *iScience* 2019; **19**: 492–503.
- 138 Stella S, Mesa P, Thomsen J, *et al.* Conformational activation promotes CRISPR-Cas12a catalysis and resetting of the endonuclease activity. *Cell* 2018; **175**: 1856–1871.e21.
- 139 Son H, Park J, Choi YH, *et al.* Exploring the dynamic nature of divalent metal ions involved in DNA cleavage by CRISPR-Cas12a. *Chem Commun* 2022; **58**: 1978–1981.
- 140 Cofsky JC, Karandur D, Huang CJ, *et al.* CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *eLife* 2020; **9**: e55143.
- 141 Wang JY, Pausch P, Doudna JA. Structural biology of CRISPR-Cas immunity and genome editing enzymes. *Nat Rev Microbiol* 2022; **20**: 641–656.
- 142 Cong L, Ran FA, Cox D, *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013; **339**: 819–823.
- 143 Mali P, Yang L, Esvelt KM, *et al.* RNA-guided human genome engineering via Cas9. *Science* 2013; **339**: 823–826.
- 144 Cho SW, Kim S, Kim JM, *et al.* Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 2013; **31**: 230–232.
- 145 Jinek M, East A, Cheng A, *et al.* RNA-programmed genome editing in human cells. *eLife* 2013; **2**: e00471.
- 146 Kleinstiver BP, Prew MS, Tsai SQ, *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 2015; **523**: 481–485.
- 147 Ran FA, Cong L, Yan WX, *et al.* *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 2015; **520**: 186–



- 191.
- 148 Sun H, Fu S, Cui S, *et al.* Development of a CRISPR-SaCas9 system for projection- and function-specific gene editing in the rat brain. *Sci Adv* 2020; **6**: eaay6687.
- 149 Hunker AC, Soden ME, Krayushkina D, *et al.* Conditional single vector CRISPR/SaCas9 viruses for efficient mutagenesis in the adult mouse nervous system. *Cell Rep* 2020; **30**: 4303–4316.e6.
- 150 Yan K, Feng J, Liu X, *et al.* Inhibition of hepatitis B virus by AAV8-Derived CRISPR/SaCas9 expressed from liver-specific promoters. *Front Microbiol* 2021; **12**: 665184.
- 151 Hu Z, Wang S, Zhang C, *et al.* A compact cas9 ortholog from *Staphylococcus auricularis* (saucas9) expands the DNA targeting scope. *PLoS Biol*, 2020; **18**: e3000686.
- 152 Hu Z, Zhang C, Wang S, *et al.* Discovery and engineering of small SlugCas9 with broad targeting range and high specificity and activity. *Nucleic Acids Res* 2021; **49**: 4008–4019.
- 153 Pattanayak V, Lin S, Guilinger JP, *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 2013; **31**: 839–843.
- 154 Slaymaker IM, Gao L, Zetsche B, *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* 2016; **351**: 84–88.
- 155 Kleinstiver BP, Pattanayak V, Prew MS, *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 2016; **529**: 490–495.
- 156 Chen JS, Dagdas YS, Kleinstiver BP, *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 2017; **550**: 407–410.
- 157 Ikeda A, Fujii W, Sugiura K, *et al.* High-fidelity endonuclease variant HypaCas9 facilitates accurate allele-specific gene modification in mouse zygotes. *Commun Biol* 2019; **2**: 371.
- 158 Nishimasu H, Shi X, Ishiguro S, *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 2018; **361**: 1259–1262.
- 159 Hu JH, Miller SM, Geurts MH, *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 2018; **556**: 57–63.
- 160 Kim HK, Lee S, Kim Y, *et al.* High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. *Nat Biomed Eng* 2020; **4**: 111–124.
- 161 Walton RT, Christie KA, Whittaker MN, *et al.* Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 2020; **368**: 290–296.
- 162 Tan Y, Chu AHY, Bao S, *et al.* Rationally engineered *Staphylococcus aureus* Cas9 nucleases with high genome-wide specificity. *Proc Natl Acad Sci USA* 2019; **116**: 20969–20976.
- 163 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**: 1293–1298.
- 164 Jiang W, Bikard D, Cox D, *et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 2013; **31**: 233–239.
- 165 DiCarlo JE, Norville JE, Mali P, *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 2013; **41**: 4336–4343.
- 166 Shan Q, Wang Y, Li J, *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 2013; **31**: 686–688.
- 167 Gratz SJ, Cummings AM, Nguyen JN, *et al.* Genome engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 nuclease. *Genetics* 2013; **194**: 1029–1035.
- 168 Hwang WY, Fu Y, Reyon D, *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 2013; **31**: 227–229.
- 169 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**: 910–918.
- 170 Maddalo D, Manchado E, Concepcion CP, *et al.* *In vivo* engineering of oncogenic chromosomal rearrangements with

- the CRISPR/Cas9 system. *Nature* 2014; **516**: 423–427.
- 171 Koike-Yusa H, Li Y, Tan EP, *et al.* Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 2014; **32**: 267–273.
- 172 Gilbert LA, Larson MH, Morsut L, *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 2013; **154**: 442–451.
- 173 Guo Y, Xu Q, Canzio D, *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 2015; **162**: 900–910.
- 174 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**: 1479–1491.
- 175 Frangoul H, Altshuler D, Cappellini MD, *et al.* CRISPR-Cas9 gene editing for sickle cell disease and  $\beta$ -thalassemia. *N Engl J Med* 2021; **384**: 252–260.
- 176 Xu L, Wang J, Liu Y, *et al.* CRISPR-edited stem cells in a patient with HIV and acute lymphocytic leukemia. *N Engl J Med* 2019; **381**: 1240–1247.
- 177 Zetsche B, Heidenreich M, Mohanraju P, *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* 2017; **35**: 31–34.
- 178 Kim D, Kim J, Hur JK, *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* 2016; **34**: 863–868.
- 179 Kim HK, Song M, Lee J, *et al.* *In vivo* high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods* 2017; **14**: 153–159.
- 180 Gao L, Cox DBT, Yan WX, *et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol* 2017; **35**: 789–792.
- 181 Tóth E, Varga É, Kulcsár PI, *et al.* Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases. *Nucleic Acids Res*, 2020; **48**: 3722–3733.
- 182 Kleinstiver BP, Sousa AA, Walton RT, *et al.* Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* 2019; **37**: 276–282.
- 183 Strohkendl I, Saifuddin FA, Rybarski JR, *et al.* Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Mol Cell* 2018; **71**: 816–824.e3.
- 184 Zhang L, Rube HT, Vakulskas CA, *et al.* Systematic *in vitro* profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. *Nucleic Acids Res* 2020; **48**: 5037–5053.
- 185 Zhang X, Wang J, Cheng Q, *et al.* Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discov* 2017; **3**: 17018.
- 186 Liu L, Chen P, Wang M, *et al.* C2c1-sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. *Mol Cell* 2017; **65**: 310–322.
- 187 Teng F, Cui T, Feng G, *et al.* Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discov* 2018; **4**: 63.
- 188 Strecker J, Jones S, Koopal B, *et al.* Engineering of CRISPR-Cas12b for human genome editing. *Nat Commun* 2019; **10**: 212.
- 189 Ming M, Ren Q, Pan C, *et al.* CRISPR-Cas12b enables efficient plant genome engineering. *Nat Plants* 2020; **6**: 202–208.
- 190 Bigelyte G, Young JK, Karvelis T, *et al.* Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nat Commun* 2021; **12**: 6191.
- 191 Wu Z, Zhang Y, Yu H, *et al.* Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat Chem Biol* 2021; **17**: 1132–1138.
- 192 Xu X, Chemparathy A, Zeng L, *et al.* Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol Cell* 2021; **81**: 4333–4345.e4.
- 193 Kim DY, Lee JM, Moon SB, *et al.* Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat Biotechnol* 2022; **40**: 94–102.

- 194 Wang Y, Wang Y, Pan D, *et al.* Guide RNA engineering enables efficient CRISPR editing with a miniature *Syntrophomonas palmitica* Cas12f1 nuclease. *Cell Rep* 2022; **40**: 111418.
- 195 Peters JE, Makarova KS, Shmakov S, *et al.* Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc Natl Acad Sci USA*, 2017; **114**: E7358–E7366.
- 196 Klompe SE, Vo PLH, Halpin-Healy TS, *et al.* Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* 2019; **571**: 219–225.
- 197 Vo PLH, Ronda C, Klompe SE, *et al.* CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat Biotechnol* 2021; **39**: 480–489.
- 198 Querques I, Schmitz M, Oberli S, *et al.* Target site selection and remodelling by type V CRISPR-transposon systems. *Nature* 2021; **599**: 497–502.
- 199 Altae-Tran H, Kannan S, Demircioglu FE, *et al.* The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* 2021; **374**: 57–65.
- 200 Karvelis T, Druteika G, Bigelyte G, *et al.* Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 2021; **599**: 692–696.
- 201 Schuler G, Hu C, Ke A. Structural basis for RNA-guided DNA cleavage by IscB-ωRNA and mechanistic comparison with Cas9. *Science* 2022; **376**: 1476–1481.
- 202 Komor AC, Kim YB, Packer MS, *et al.* Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 2016; **533**: 420–424.
- 203 Gaudelli NM, Komor AC, Rees HA, *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 2017; **551**: 464–471.
- 204 Zong Y, Song Q, Li C, *et al.* Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat Biotechnol* 2018; **36**: 950–953.
- 205 Li C, Zong Y, Wang Y, *et al.* Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol* 2018; **19**: 59.
- 206 Anzalone AV, Randolph PB, Davis JR, *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 2019; **576**: 149–157.
- 207 Nelson JW, Randolph PB, Shen SP, *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol* 2022; **40**: 402–410.
- 208 Liu B, Dong X, Cheng H, *et al.* A split prime editor with untethered reverse transcriptase and circular RNA template. *Nat Biotechnol* 2022; **40**: 1388–1393.
- 209 Zheng C, Liang S-Q, Liu B, *et al.* Development of a flexible split prime editor using truncated reverse transcriptase. *bioRxiv*, 2021, doi: 10.1101/2021.08.26.457801.
- 210 Mok BY, de Moraes MH, Zeng J, *et al.* A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 2020; **583**: 631–637.
- 211 Guo J, Zhang X, Chen X, *et al.* Precision modeling of mitochondrial diseases in zebrafish via ddcbe-mediated mtdna base editing. *Cell Discov*, 2021; **7**: 78.
- 212 Kang BC, Bae SJ, Lee S, *et al.* Chloroplast and mitochondrial DNA editing in plants. *Nat Plants* 2021; **7**: 899–905.
- 213 Lee H, Lee S, Baek G, *et al.* Mitochondrial DNA editing in mice with DddA-TALE fusion deaminases. *Nat Commun* 2021; **12**: 1190.
- 214 Sabharwal A, Kar B, Restrepo-Castillo S, *et al.* The fusx tale base editor (fusxtbe) for rapid mitochondrial DNA programming of human cells *in vitro* and zebrafish disease models *in vivo*. *Crispr j*, 2021, 4: 799–821.
- 215 Guo J, Chen X, Liu Z, *et al.* DdCBE mediates efficient and inheritable modifications in mouse mitochondrial genome. *Mol Ther Nucleic Acids* 2022; **27**: 73–80.
- 216 Wei Y, Xu C, Feng H, *et al.* Human cleaving embryos enable efficient mitochondrial base-editing with DdCBE. *Cell Discov* 2022; **8**: 7.
- 217 Cho SI, Lee S, Mok YG, *et al.* Targeted A-to-G base editing in human mitochondrial DNA with programmable

- deaminases. *Cell* 2022; **185**: 1764–1776.e12.
- 218 Mok BY, Kotrys AV, Raguram A, *et al.* CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat Biotechnol* 2022; **40**: 1378–1387.
- 219 Lei Z, Meng H, Liu L, *et al.* Mitochondrial base editor induces substantial nuclear off-target mutations. *Nature* 2022; **606**: 804–811.
- 220 Levy JM, Yeh WH, Pendse N, *et al.* Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* 2020; **4**: 97–110.
- 221 Xu L, Zhang C, Li H, *et al.* Efficient precise *in vivo* base editing in adult dystrophic mice. *Nat Commun* 2021; **12**: 3719.
- 222 Molla KA, Sretenovic S, Bansal KC, *et al.* Precise plant genome editing using base editors and prime editors. *Nat Plants* 2021; **7**: 1166–1187.
- 223 Shimatani Z, Kashojiya S, Takayama M, *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol* 2017; **35**: 441–443.
- 224 Zong Y, Wang Y, Li C, *et al.* Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol* 2017; **35**: 438–440.
- 225 Lin Q, Zong Y, Xue C, *et al.* Prime genome editing in rice and wheat. *Nat Biotechnol* 2020; **38**: 582–585.
- 226 Porto EM, Komor AC, Slaymaker IM, *et al.* Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov* 2020; **19**: 839–859.
- 227 Anzalone AV, Gao XD, Podracky CJ, *et al.* Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat Biotechnol* 2022; **40**: 731–740.
- 228 Lin Q, Jin S, Zong Y, *et al.* High-efficiency prime editing with optimized, paired pegRNAs in plants. *Nat Biotechnol* 2021; **39**: 923–927.
- 229 Zong Y, Liu Y, Xue C, *et al.* An engineered prime editor with enhanced editing efficiency in plants. *Nat Biotechnol* 2022; **40**: 1394–1402.
- 230 Jin S, Lin Q, Gao Q, *et al.* Optimized prime editing in monocot plants using PlantPegDesigner and engineered plant prime editors (ePPEs). *Nat Protoc* 2023; **18**: 831–853.
- 231 Gammage PA, Moraes CT, Minczuk M. Mitochondrial genome engineering: the revolution may not be CRISPR-ized. *Trends Genet* 2018; **34**: 101–110.
- 232 Gammage PA, Viscomi C, Simard ML, *et al.* Genome editing in mitochondria corrects a pathogenic mtDNA mutation *in vivo*. *Nat Med* 2018; **24**: 1691–1695.
- 233 Hegge JW, Swarts DC, van der Oost J. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nat Rev Microbiol*, 2018; **16**: 5–11.
- 234 Gao F, Shen XZ, Jiang F, *et al.* DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. *Nat Biotechnol* 2016; **34**: 768–773.
- 235 Burgess S, Cheng L, Gu F, *et al.* Questions about NgAgo. *Protein Cell* 2016; **7**: 913–915.
- 236 Xu S, Cao S, Zou B, *et al.* An alternative novel tool for DNA editing without target sequence limitation: the structure-guided nuclease. *Genome Biol* 2016; **17**: 186.
- 237 Tian K, Guo Y, Zou B, *et al.* DNA and RNA editing without sequence limitation using the flap endonuclease 1 guided by hairpin DNA probes. *Nucleic Acids Res* 2020; **48**: e117.
- 238 Wang J, Lisanza S, Juergens D, *et al.* Scaffolding protein functional sites using deep learning. *Science* 2022; **377**: 387–394.
- 239 Dauparas J, Anishchenko I, Bennett N, *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 2022; **378**: 49–56.
- 240 Wicky BIM, Milles LF, Courbet A, *et al.* Hallucinating symmetric protein assemblies. *Science* 2022; **378**: 56–61.