## **Trends in Biochemical Sciences**

### **Review**

# A Tale of Two Moieties: Rapidly Evolving CRISPR/Cas-Based Genome Editing

Li Yang<sup>1,2,\*</sup> and Jia Chen<sup>2,3,\*</sup>

Two major moieties in genome editing are required for precise genetic changes: the locator moiety for target binding and the effector moiety for genetic engineering. By taking advantage of CRISPR/Cas, which consists of different modules for independent target binding and cleavage, a spectrum of precise and versatile genome editing technologies have been developed for broad applications in biomedical research, biotechnology, and therapeutics. Here, we briefly summarize the progress of genome editing systems from a view of both locator and effector moieties and highlight the advance of newly reported CRISPR-conjugated base editing and prime editing systems. We also underscore distinct mechanisms of off-target effects in CRISPR-conjugated systems and further discuss possible strategies to reduce off-target mutations in the future.

#### Genome Editing from a View of Two Moieties

The completion of human genome project in the beginning of this century [1,2] and the application of affordable high-throughput sequencing technologies in the past decade [3] have led life science researches to the post-genome era with genome-wide understanding of functional genomic elements related to human health and diseases. Importantly, the advent of practical genome editing technologies provides powerful methods to change genetic information, which benefits not only basic research aiming to decipher how different genotypes result in distinct phenotypes but also preclinical study to cure human diseases caused by genetic mutations. To target any genomic locus for desired DNA changes, two major moieties, a **locator** (see Glossary) and an **effector**, are usually required for competent genome editing. The locator moiety is designed to recognize and bind to a specific genomic locus, which guides the effector moiety for subsequent change of DNA sequence.

In last two decades or so, programmable genome editing systems have been mainly evolved from fusions of endonucleases to locators, such as zinc finger (ZF) motifs [4] and transcription activator-like effector (TALE) repeats [5] (Box 1), to the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein (Cas)-based technologies [6-8]. Unlike ZFs and TALEs, which are fused with a heterogeneous Fokl endonuclease for genome editing (Box 1), CRISPR/Cas proteins are featured by their dual functions. In addition to their DNA/RNA binding activity together with gRNA, CRISPR/Cas proteins can also process DNA/RNA cleavage activity with their endonuclease domains [9-12]. This makes CRISPR/Cas a convenient method for genome editing. Indeed, since it appeared in the early 2010s [11,13–15], CRISPR/Cas has been widely applied in genome editing of both single gene study and genome-wide screening, from bacteria to mammals [6-8]. However, although revolutionary, CRISPR/Cas systems were not always precise, but with unwanted side-products; there has been an aim to have improved precision in the application of genome editing to treat genetic diseases associated with single base mutations. Recently, by fusing CRISPR/Cas proteins (as the genome locator) with different types of effector moieties, such as nucleobase deaminases [16,17] or reverse transcriptases [18], more precise and versatile genome editing technologies have been developed to achieve single nucleotide editing

#### Highlights

CRISPR/Cas with different modules for independent target binding and cleavage has evolved to achieve convenient and precise genome editing.

CelPress

The endonuclease effector in conventional CRISPR/Cas genome editing systems can be replaced by nucleobase deaminases and the resulting base editors (BEs) enable single base changes.

By fusing CRISPR/Cas with reverse transcriptases, prime editors (PEs) represent a new way to accomplish genetic changes, including all types of base substitutions, small indels, and their combinations.

Both BEs and PEs are of potential in correcting disease-associated mutations.

Genome-wide and/or transcriptomewide off-target mutations are catalyzed by the nucleobase deaminase effector in BEs, which are independent of the fused gRNA/Cas moiety.

<sup>1</sup>CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup>School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

<sup>3</sup>CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China

\*Correspondence:

liyang@picb.ac.cn (L. Yang) and chenjia@shanghaitech.edu.cn (J. Chen).



#### Box 1. Genome Targeting Achieved by Protein Locators

To target any specific genomic site is one of the primary requirements for a programmable genome editing technology. Site-specific nucleases have long been applied in DNA recombination *in vitro* and therefore were first thought to be used for gene editing. For example, meganucleases, a type of endonucleases that recognize long DNA sequences (~12–40 bp), have been applied and engineered to generate DSBs at specific loci in genomic DNA [141,142]. However, due to their limited recognition sites and the difficulty to program their targeting specificities, meganucleases were not suitable in certain applications, such as in high-throughput screening assays.

The first applicable locator for genome editing was developed with ZF motifs, originally discovered in transcription factors in *Xenopus laevis* [143,144]. By fusing an array of ZF motifs as the locator with the cleavage domain of *Fokl* endonuclease as the effector, ZF nucleases (ZFNs) were developed to fulfill genome editing [145], theoretically at any given genomic locus. The specificity of ZFNs is rendered by the customized array of ZF motifs, each of which consists of about 30 amino acids to recognize a definite nucleotide triplet [146,147]. Within a designed ZFN, different ZF motifs can be combined to recognize ~9–18 bp at the targeted genomic locus for subsequent editing [148]. However, the application of ZFNs at most genomic target sites has remained challenging due to the crosstalk between adjacent ZF motifs that interferes with their binding to the corresponding DNA.

The ZFN-based technology was the only programmable method to engineer genomic DNA sequences for a while, prior to the appearance of TALE nucleases (TALENs) in 2011 [149]. The TALEN system uses TALE repeats, from a bacterial plant pathogen *Xanthomonas*, as the locator [5]. Each TALE repeat composes of 33–35 amino acids to distinguish a single base pair of DNA [150,151]; this leads to increased flexibility in designing customized TALENs to engineer most genetic loci by combining matched TALE repeats. By fusing an array of TALE DNA binding domains that recognize designated base pairs to the cleavage domain of *Fokl* endonuclease, the fusion protein can bind to a specific DNA sequence without the interference of each TALE domain in the array [149,152,153]. Nonetheless, the construction of TALEN vectors is complicated due to the homologous recombination of repetitive DNA sequences to express TALE repeats.

at target sites. In this review, we discuss the evolution of genome editing technologies in terms of two moieties, emphasize newly reported **base editing** and **prime editing** technologies based on CRISPR/Cas systems that have increased precision in gene editing, and further dissect underlined mechanisms that may account for their unwanted off-target (OT) effects for future improvement. (See Table 1).

# Programmable Locators Evolved to Ribonucleoproteins (RNPs) in CRISPR/Cas Platform

In nature, CRISPR/Cas functions as an adaptive immunity in bacteria and archaea against the invasion of foreign pathogens, such as phages [19–22]. Among many discovered CRISPR/Cas proteins, class 2 Cas systems use a single Cas protein [23], commonly type II Cas9 [13–15,24] and type V Cas12a (previously known as Cpf1) [11], for target DNA cleavage and have been well adopted for developing new genome editing technologies.

The ability of target binding in CRISPR/Cas-based systems is basically directed by a synthetic **guide RNA (gRNA)** and carried out by the gRNA/Cas RNP complex [24,25]. As exemplified by the CRISPR/Cas9 system in Figure 1A, the gRNA of gRNA/Cas9 RNP hybridizes to an intended DNA region containing the sequence (**protospacer**) complementary to gRNA and the Cas9 protein binds to the intended DNA region with a nearby **protospacer-adjacent motif (PAM)** [26–28]. Different Cas proteins have distinct PAM preferences. The PAM sequences for Cas9 proteins are generally G-rich and locate at the 3'-end of the protospacer (Figure 1A) [10,24], while Cas12a proteins recognize T-rich PAMs at the 5'-end of the protospacer (Figure 1B) [11]. Furthermore, engineering naturally existing Cas proteins can also diversify their targeting PAM sequences to extend editing scopes. For example, the wild type *Streptococcus pyogenes* Cas9 (SpCas9) recognizes a canonical NGG PAM [24], whereas engineered SpCas9 variants can recognize PAMs of NGA/NAG [29], NG [30,31], or even non-G PAMs [32,33]. Due to the strict requirement of PAMs for the binding of specific CRISPR/Cas to genomic sites, the availability of current CRISPR/Cas platforms with limited PAMs may impede genome editing pinpointed at any desired location. In this case, the discovery of new Cas proteins together

#### Glossary

Adenosine deaminase: a type of enzyme to deaminate adenosine by substituting the amino group for a keto group, resulting in adenosine-to-inosine editing in RNA. These enzymes are widely distributed in bacteria, plants, invertebrates, vertebrates, and mammals and considered to be key enzymes for purine metabolism. ADARs, adenosine deaminases acting on RNA, in human are crucial for embryonic/ neural development and also related with innate immune responses to viral RNAs.

**Base editing:** a gene editing technology that combines CRISPR/Cas system with nucleobase deaminase (e.g., cytidine deaminase or adenosine deaminase) to achieve precise base substitutions in target DNA or RNA.

Base excision repair (BER): an endogenous DNA repair pathway to remove damaged DNA bases, such as uracils from cytidine deamination. With DNA glycosylases, damaged bases are removed to form apurinic/apyrimidinic sites (AP sites, also known as abasic sites). The resulting AP sites are further cleaved by an AP endonuclease to generate DNA single-strand break, leading to either a single nucleotide replacement or multiple, commonly two to ten, nucleotide displacing synthesis. If a damaged DNA base appears in a single-stranded DNA region, BER can lead to a DNA double-strand break.

Cytidine deaminase (CDA): a type of enzyme to deaminate cytidine by substituting the amino group for a keto group, resulting in cytidine-to-uridine editing. A variety of APOBEC/AID family of cytidine deaminases have been found to catalyze cytidine-to-uridine editing in both RNA and DNA.

Effector: the moiety that can modify target site in a genome editing system. Guide RNA (gRNA): a synthesized RNA component that guides Cas proteins to bind at the target site in the CRISPR/Cas system.

Indel: random nucleotide insertion or deletion that is usually triggered by endjoining repair of a DNA double-strand break. Indels often lead to open reading frame shifts and ultimately disrupt the expression of protein products, which is commonly used for gene knockout. Locator: the moiety that can bind at target site in a genome editing system. Mismatch repair (MMR) pathway: an endogenous DNA repair pathway to



with their engineering can further expand the targeting range of CRISPR/Cas systems, hopefully to cover all regions across the whole genome.

In addition to targeting DNA, RNA editing technologies have recently gained attention due to their feature of no change of genomic DNA information in most species. With programmable CRISPR RNAs (crRNAs), class 2 type VI Cas13 proteins have been used to knockdown target RNA with crRNA-complementary sequence and a 3' protospacer flanking site [12,34,35], exemplified by Cas13a system in Figure 1C. Although still in its early stage, RNA editing technologies, including single base RNA editing [36,37], also hold potential in biomedical research and therapeutics, owing to a lower genomic OT concern and a reversible and temporary manner. The development and application of RNA editing technologies have been discussed elsewhere [38,39].

#### **Distinct Effectors for Various Genome Editing Outcomes**

In wild type CRISPR/Cas systems, Cas proteins themselves can function as both locators and effectors. After binding to corresponding gRNAs, the endonuclease activity of Cas proteins is activated to cut DNA double strands at a given target site that is complementary to gRNA [9,11], generating double-strand DNA breaks (DSBs). In general, these DSBs can be repaired by endogenous end-joining repair pathways (Box 2), which commonly introduce random **insertions or deletions (indels**) of nucleotides [40] for gene 'knockout' (KO). Although precise sequence replacement at CRISPR/Cas-triggered DSBs can be alternatively achieved by homology-directed repair (HDR) (Box 2), it not only requires the presence of an additional donor DNA with edit [11,41], but also is less efficient than imprecise end-joining [42].

donor DNA with edit [11,41], but also is less efficient than imprecise end-joining [42]. CRISPR/Cas endonuclease activity is carried out differently among different types of Cas proteins. For instance, Cas9 proteins have two individual endonuclease domains, HNH and RuvC. The HNH domain of Cas9 cleaves DNA at the target strand, which hybridizes with gRNA, while the RuvC domain cleaves the nontarget strand, which is cognate to the spacer region of gRNA [24]. Mutating one of these two domains results in two Cas9 nickases (nCas9s), D10A and H840A, for nicking only one strand of DNA helix (Figure 1A). Differently, Cas12a proteins have only a RuvC-like nuclease domain, which cleaves both nontarget and target strands (Figure 1B) [43]. In contrast, Cas13 proteins specifically cleave RNA with two HEPN domains (Figure 1C) [44]. Of note, nuclease activities of most Cas proteins are independent to their binding activities, as both *in vitro* and *in vivo* studies have shown that catalytically dead Cas9 (dCas9) [26,45], Cas12a (dCas12a) [43,46], and Cas13 (dCas13) [44,47] could still bind DNA/RNA substrates (Figure 1).

#### Adopting Naturally Existing Cytidine Deaminase Effector for C-to-T Base Editing

Distinct to convenient and efficient gene KO, the efficiency and product purity of precise editing by CRISPR/Cas has remained low [42], which hinders its application in therapeutics, such as correcting human genetic variants relevant to diseases. Considering that the majority of reported human pathogenetic variants are point mutations [48–50], new technologies are desired to achieve genome editing at single nucleotide resolution with high precision and efficiency. This dream came true in 2016, with the reports of efficient genome editing at single bases [16,51], originally referred to as base editors (BEs) and later as cytosine BE (CBE) more specifically. The original CBEs adapted gRNA/dCas9 as a locator and utilized apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC)/activation induced deaminase (AID) family of **cytidine deaminases** (**CDAs**) as an effector (Figure 2A, Key Figure). Naturally, APOBEC enzymes catalyze the deamination of cytidine (C) to uridine (U) in single-strand RNA or DNA (ssDNA) regions [52–54]. Since uracil in DNA is usually a signal for **base excision repair (BER)**, an endogenous DNA repair pathway to remove base lesions, such as uracil, in genome [55,56], a uracil DNA glycosylase inhibitor (UGI) [57]

repair erroneous short insertion, deletion, and mis-incorporation of bases. In D10A-mediated base editing where a C-to-U (or A-to-I) change happens, MMR resolves the U/G (or I/T) mismatch to a U/A (or I/C) pair, which can be then converted to a T/A (or G/C) base pair after DNA replication or repair. Prime editing: a genome editing technology that combines the CRISPR/ Cas system with reverse transcriptase (e.g., Moloney murine leukemia virus reverse transcriptase) to synthesize DNA according to the RNA template of a prime editing guide RNA (pegRNA) and finally achieve precise genome editing with great versatility.

**Protospacer:** a DNA region in invading viral or plasmid DNA that can be recognized by a CRISPR/Cas system.

Protospacer-adjacent motif (PAM): a short DNA sequence immediately following a protospacer that is targeted by a gRNA. A PAM can be at the 5' or 3' end of a protospacer.

# **Trends in Biochemical Sciences**



#### Table 1. Representative Genome Editors

Genome editor	Locator	Effector	PAM	Locator-dependent OT effects	Locator-independent OT effects	Refs
ZFN	ZF motif	Fokl nuclease	-	+++	-	[145,148]
TALEN	TALE repeat	Fokl nuclease	-	+++	-	[149,152,153]
Cas9	Cas9/gRNA	Cas9 nuclease	NGG	+++	-	[13,14,154]
Cas12a	Cas12a/gRNA	Cas12a nuclease	TTTV	++	-	[11]
Cas9-VQR	Cas9-VQR/gRNA	Cas9 nuclease	NGA	+++	-	[29]
xCas9	xCas9/gRNA	Cas9 nuclease	NG, GAA, GAT	++	-	[30]
Cas9-NG	Cas9-NG/gRNA	Cas9 nuclease	NG	+++	-	[31]
Cas9-NRRH	Cas9-NRRH/gRNA	Cas9 nuclease	NRRH	+++	-	[32]
Cas13a	Cas13a/gRNA	Cas13a nuclease	Н	+	-	[34]
nCas9	nCas9/gRNA pair	nCas9	NGG	++	-	[108,109]
dCas9-Fokl	dCas9/gRNA pair	Fokl nuclease	NGG	+	-	[110,155]
eSpCas9	eSpCas9/gRNA	Cas9 nuclease	NGG	-	-	[125]
SpCas9-HF	SpCas9-HF/gRNA	Cas9 nuclease	NGG	-	-	[126]
HypaCas9	HypaCas9/gRNA	Cas9 nuclease	NGG	-	-	[127]
Sniper-Cas9	Sniper-Cas9/gRNA	Cas9 nuclease	NGG	-	-	[128]
BE3	nCas9/gRNA	rA1	NGG	+++	DNA: +++, RNA: +++	[16]
YE1-BE3	nCas9/gRNA	rA1-YE1	NGG	+++	DNA: -, RNA: -	[68]
YEE-BE3	nCas9/gRNA	rA1-YEE	NGG	+++	DNA: -, RNA: -	[68]
BE4	nCas9/gRNA	rA1	NGG	+++	DNA: +++, RNA: +++	[72]
eBE	nCas9/gRNA	rA1	NGG	+++	DNA: +++, RNA: +++	[73]
hA3A-BE3	nCas9/gRNA	hA3A	NGG	+++	DNA: +++, RNA: +++	[63]
hA3A-BE3-Y130F	nCas9/gRNA	hA3A-Y130F	NGG	+++	DNA: +++, RNA: -	[63]
hA3A-BE3-Y132D	nCas9/gRNA	hA3A-Y132D	NGG	+++	DNA: +++, RNA: +	[63]
eA3A-BE3	nCas9/gRNA	A3A-N57Q	NGG	++	DNA: +++, RNA: +	[69]
SaKKH-BE3	nSaKKHCas9/gRNA	rA1	NNNRRT	+++	DNA: +++, RNA: +++	[68]
Target-AID	nCas9/gRNA	Sea lamprey CDA	NGG	+++	DNA: +++, RNA: -	[51]
dCas12a-BE	dLbCas12a/gRNA	rA1	TTTV	++	DNA: +++, RNA: +++	[59]
BEACON1	dLbCas12a/gRNA	Engineered hA3A	TTTV	++	DNA: +++, RNA: +	[75]
BEACON2	dLbCas12a/gRNA	Engineered hA3A	TTTV	++	DNA: +++, RNA: -	[75]
enAsBE	denAsCas12a/gRNA	rA1	VITV, TTTT, TTCN/TATV	+	DNA: +++, RNA: +++	[74]
PBE	nCas9/gRNA	rA1	NGG	+++	DNA: +++, RNA: +++	[61]
A3A-PBE	nCas9/gRNA	hA3A	NGG	+++	DNA: +++, RNA: +++	[64]
ABE7.10	nCas9/gRNA	TadA-TadA*	NGG	+++	DNA: -, RNA: +	[17]
ABE8e	nCas9/gRNA	TadA-TadA-8e	NGG	+++	DNA: +++, RNA: +++	[83]
ABE8e-V106W	nCas9/gRNA	TadA-TadA-8e-V106W	NGG	+++	DNA: +, RNA: ++	[83]
LbABE8e	dLbCas12a/gRNA	TadA-TadA-8e	TTTV	++	DNA: +++, RNA: +++	[83]
STEME-1	nCas9/gRNA	hA3A-TadA-TadA*	NGG	+++	DNA: +++, RNA: +++	[83]
ABE-P1	nCas9/gRNA	TadA-TadA*	NGG	+++	DNA: -, RNA: +	[92]
ABE-P2	nSaCas9/gRNA	TadA-TadA*	NNGRRT	+++	DNA: -, RNA: +	[92]
rBE14	nCas9/gRNA	TadA-TadA*	NGG	+++	DNA: -, RNA: +	[93]
PE1	dCas9/gRNA	M-MLV RTase	NGG	+	DNA: ?, RNA: ?	[18]

 $4 \qquad \text{Trends in Biochemical Sciences, Month 2020, Vol. xx, No. xx} \\$ 



#### Table 1. (continued)

Genome editor	Locator	Effector	PAM	Locator-dependent OT effects	Locator-independent OT effects	Refs
PE2	nCas9/gRNA	M-MLV RTase	NGG	+	DNA: ?, RNA: ?	[18]
PE3	nCas9/gRNA	M-MLV RTase	NGG	+	DNA: ?, RNA: ?	[18]
PPE	nCas9/gRNA	M-MLV RTase	NGG	+	DNA: ?, RNA: ?	[94]

was also fused in BE to inhibit BER and maintain the uracil, which can be recognized as thymine by cells to achieve C-to-T base editing. In order to enhance editing efficiency, dCas9 was further replaced by nCas9 D10A (with an inactive RuvC domain) in the most commonly used BE3 system. In BE3, the APOBEC/AID-generated C-to-U editing in nontarget strand together with the D10A-generated nick in target strand trigger the **mismatch repair (MMR) pathway** [58]. Then, MMR removes the unedited G-containing strand and resynthesizes it complementary to the U-containing sequence, resolving the U/G mismatch to a U/A pair, which can be then converted to a T/A base pair after DNA replication or repair processes. In most early versions of CBEs, the rat APOBEC (rA1) effector was used to catalyze the deamination of targeted cytosines to induce C-to-T editing [16,59–61]. For higher C-to-T editing efficiency, rA1 was replaced by other types of APOBEC deaminases, which also expands the editing scope [50,62]. For instance, conjugating



#### Trends in Blochemical Sciences

Figure 1. Schematic Drawing of Three Representative CRISPR/Cas Systems. (A) The class 2 type II Cas9 system. Together with a synthetic gRNA, Cas9 nuclease (top), Cas9 nickases (D10A and H840A, middle two), and catalytically-dead Cas9 (bottom) bind to target DNA. (B) The class 2 type V Cas12a system. Together with a crRNA, Cas12a nuclease (top) and catalytically-dead Cas12a (bottom) bind to target DNA. (C) The class 2 type VI Cas13 system. Together with a crRNA, Cas12a nuclease (top) and catalytically-dead Cas12a (bottom) bind to target DNA. (C) The class 2 type VI Cas13 system. Together with a crRNA, Cas13a nuclease (top) and catalytically-dead Cas13a (bottom) bind to target RNA. Targeted cleavage sites of Cas9 and Cas13a nucleases and two Cas9 nickases by corresponding endonuclease domains are highlighted with arrowhead. Abbreviations: crRNA, CRISPR RNA; gRNA, guide RNA; PAM, protospacer-adjacent motif; PFS, protospacer flanking site.



#### Box 2. Pathways for DNA Double-Strand Break Repair

When the endonuclease effector of genome editors cleaves both strands of DNA, a double-strand break (DSB) is generated. DSBs cannot be fixed by endonuclease effector, but by endogenous DNA repair enzymes [42]. Mechanically, DSBs trigger two endogenous repair pathways, nonhomologous end-joining (NHEJ) and homology-directed repair (HDR).

In general, NHEJ is the major repair pathway for DSBs introduced by genome editors. NHEJ (or microhomologymediated end-joining) can induce random insertions or deletions (indels) in the genomic DNA regions around a DSB, which can result in open reading frame shifts and, finally, gene inactivation. However, endonuclease-generated DSBs can also trigger HDR to achieve sequence replacement with high precision when a donor DNA is present [41,42]. Compared with gene knockout by NHEJ, accurate sequence changes by HDR is more desirable for therapies, such as correcting human pathogenic-related mutations. However, DSB-triggered HDR is not efficient enough for most gene correction purposes. Even with a foreign DNA donor aiming for HDR, high levels of random indels rather than efficient and precise replacement were observed [42]. The attempt to develop other efficient genome editing methods for precise gene correction has been long standing.

human APOBEC3A (hA3A) in CBEs can efficiently edit cytosines in highly methylated regions and in the GpC dinucleotide content [63–67]. Furthermore, fusing engineered and/or *in vitro* evolved APOBEC effectors (e.g., rA1-YE1, rA1-YEE, hA3A-Y130F, hA3A-Y132D, eA3A and evoA1) in CBEs can narrow the base editing window (a context region within gRNA target site in which all cytosines can be potentially converted to thymines) to reduce unintended bystander mutations (unintended C-to-T changes within editing windows) and to diversify editing scopes for C-to-T changes [63,68–70].

Although CBEs do not induce DSBs directly, indels were still observed in treatments with CBEs [16,51], resulting from the cleavage of fused nCas9 in most CBEs and the further breakage of the abasic site after the excision of U by uracil DNA glycosylase [71]. To reduce indel formation, more UGIs were fused into or coexpressed with nCas9-CBEs, which enhanced editing efficiency as well [72,73]. Differently, nCas9 could be replaced by dCpf1/dCas12a in some recently developed CBEs [59,74,75], which were shown to induce efficient C-to-T editing with only a basal level of DNA damage response [75], due to the fusion of catalytic dead dCpf1/dCas12a in these CBEs.

As all cytosines in the editing window of CBEs can be potentially converted to thymines, wide editing windows are not suitable for precise single base changes, but are useful to induce diversified mutagenesis for high-throughput screening of functional variants [76,77]. In contrast, narrow editing windows, despite limiting editing scopes, are precise to pinpoint desired single base changes [63,68].

#### Developing In Vitro Evolved Adenosine Deaminase Effector for A-to-G Base Editing

Other than pathogenic T-to-C (or A-to-G) mutations that can be potentially corrected by CBEs, the majority of reported human pathogenic variants are G-to-A (or C-to-T) [48–50]. In this case, another type of genome editing technology was desired to reverse pathogenic G-to-A (or C-to-T) variants for treatment. It is known that the deamination of adenosine leads to adenosine-to-inosine editing (A-to-I) naturally only at RNA, but not at DNA [78–80]. Thus, native **adenosine deaminases** cannot be directly used in developing adenine BEs (ABEs). To solve this problem, *Escherichia coli* tRNA-specific adenosine deaminase (TadA) was selected for seven rounds of directed evolution *in vitro* to gain TadA\* that exhibits adenosine effector, which contains a wild type TadA linked with the *in vitro* evolved TadA\*, to nCas9 (D10A) for A-to-I DNA editing (Figure 2B) [17]. Similar to CBEs, the subsequent MMR or DNA replication resolves the resulted I/T mismatch to an I/C pair and eventually installs a G/C pair at the target site for A-to-G base editing. As inosines rarely exist in DNA, no DNA glycosylase is yet known to efficiently remove inosines from deoxyribose. As a result, no DNA glycosylase inhibitor is required to be fused into



### **Key Figure**

Conjugating Nucleobase Deaminases or Reverse Transcriptases with CRISPR/Cas Proteins to Achieve Precise Genome Editing at Single Nucleotide Resolution



Figure 2. (A,B) Schematic drawing of cytosine base editor (CBE) (A) and adenine base editor (ABE) (B). Single base change of cytosine or adenine has been achieved by fusing cytidine (A) or adenosine (B) deaminase with Cas9 nickase (D10A). Of note, uracil DNA glycosylase inhibitor (UGI) is included in CBEs, but not ABEs, to reduce the formation of unwanted indels by CBEs. (C) Schematic drawing of dual function BE for simultaneous cytosine and adenine deamination. (D) Schematic drawing of prime editor (PE). The conjugation of reverse transcriptase (RTase) with Cas9 nickase (H840A) leads to a versatile PE system for all type of base substitutions, small indels and their combinations. Abbreviations: AID, Activation induced deaminase; APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; crRNA, CRISPR RNA; dCas12a, catalytically dead Cas12a; gRNA, guide RNA; nCas9, Cas9 nickase; PAM, protospacer-adjacent motif; pegRNA, prime editing guide RNA.



ABEs and no significant indel formation was induced by ABEs [81,82]. Recently, Cas12a-derived ABEs have also been reported by combining further evolved adenosine deaminases with dCas12a for A-to-G base editing [83].

As adenosine deamination naturally happens at the RNA level, A-to-I base editing in RNA was also obtained by conjugating the deaminase domain of adenosine deaminase acting on RNA (ADAR, mainly with that of ADAR2) as the effector with RNA-targeting dCas13 protein as the locator [36]. Interestingly, the adenine deaminase domain of ADAR2 has been evolved *in vitro* to deaminate cytidine and further used to perform targeted C-to-U RNA base editing by fusing with dCas13 protein [37]. However, it is untested whether the unusual cytidine deamination by evolved ADAR2 could also occur in DNA. In addition, whether other types of base changes, such as cytosine-to-guanine observed in somatic hypermutation of immunoglobulin genes [84], can be adapted for corresponding base editing remains unreported.

# Combining Cytosine and Adenosine Deaminase Effectors for Simultaneous C-to-T and A-to-G Base Editing

Despite being valuable, the utility of CBEs and ABEs in correcting pathogenic variants is limited, as CBEs are solely for T-to-C mutations and ABEs are for G-to-A ones. To further expand editing competency, dual-functional BEs were developed by fusing both cytidine and adenosine deaminases with nCas9 in both plants and mammals (Figure 2C) [85,86]. These dual functional base editing systems were reported to induce simultaneous C-to-T and A-to-G changes efficiently in tested editing windows. As hundreds of known pathogenic T-to-C and G-to-A point mutations coexist close enough to fit in same editing windows, these dual-functional base editing systems are promising in therapeutics [86].

#### Exploiting Reverse Transcriptase Effector for Versatile Genome Editing

In addition to C-to-T and/or A-to-G editing, new strategies for any targeted base change have long been desired. Recently, a versatile gene editing tool, prime editor (PE), has been developed to induce all types of base substitutions, small indels and their combinations with high efficiency and product purity (Figure 2D) [18]. In the PE system, a multifunctional prime editing guide RNA (pegRNA) that binds with nCas9 H840A (with an inactive HNH domain) is used as the locator and a conjugated reverse transcriptase (RTase) is used as the effector. The featured pegRNA contains three functional parts of sequences: a typical sgRNA with a spacer region for PE targeting, a primer binding site (PBS) for reverse transcription (RT) primer binding and RT initiation, and an RT template with edit(s) for intended DNA changes (Figure 2D). Mechanically, with the spacer sequence in pegRNA, the H840A locator binds to the target genomic DNA site and nicks the nontarget strand to generate a single-strand break (SSB) as RT primer, which binds to PBS in pegRNA to initiate RT by the conjugated RTase effector and then to convert the pegRNA template sequence with intended edit information to cDNA. The synthesized cDNA is finally incorporated into the target region by taking advantage of the endogenous MMR pathway [18,87]. Several steps of improvements have been fulfilled to ensure high levels of genome editing outcomes by PEs in mammalian cells [18]. For example, the editing efficiency was much improved by engineering Moloney murine leukemia virus (M-MLV) RTase as the effector, owing to the enhanced binding ability at the RT initiation site, thermostability, and enzyme processivity. In addition, a canonical gRNA (nicking gRNA) was introduced to make a flanking nick in the target strand, which triggers the MMR pathway to remove the unedited strand and to maintain the edited strand for even higher prime editing efficiency. Although PE can induce precise editing with great versatility, the use of the PE system requires comprehensive design and, therefore, multiple parameters need to be considered with delicacy, such as the length of the PBS, the



sequence of RT template, the location of the edit, and the selection of nicking gRNA. In addition, OT effects of PEs have not been tested in a genome-wide manner. In the future, the new PE system will definitely be improved as a promising technology in gene therapy to correct most disease-associated genetic variants, including 12 types of base changes, small indels, and their combinations [87].

#### Applications of BEs and PEs

Ever since their recent advent, CBEs have been widely used in biological and biomedical researches, such as correcting or modeling human pathogenic variants. Kim et al. applied BE3 into mouse embryos to mimic Duchenne muscular dystrophy and albinism [60]. Later on, Li et al. compared the editing of BE3 and hA3A-BE3-Y130F at multiple genomic loci in mice and found that hA3A-BE3-Y130F induced higher editing efficiency in G/C-rich regions [65]. Chadwick et al. packaged BE3 into an adenoviral vector to disrupt proprotein convertase subtilisin/kexin type 9 (PCSK9) and found that both plasma PCSK9 and cholesterol levels were significantly reduced [88]. By using BE3 in utero gene editing, Rossidis et al. also disrupted Pcsk9 and thus reduced the serum cholesterol level [156]. SaKKH-BE3, a BE with SaCas9-KKH locator, was used to treat phenylketonuria in adult mice through the delivery of adeno-associated virus [89]. Recently, A3A (N57Q)-BE3 was used to edit the enhancer region of B cell lymphoma/leukemia 11A (BCL11A) gene and expression of fetal hemoglobin (HbF) was induced successfully, which showed therapeutic benefits for sickle cell disease and  $\beta$ -thalassemia [90]. In addition to animals, Zong et al. successfully applied codon-optimized BE3 [plant base editor (PBE)] in plants [61] and later, the same lab also optimized hA3A-BE3 to develop the plant version of A3A-PBE to achieve higher editing efficiencies in plants [64].

As for ABEs, Ryu *et al.* used ABE7.10 to edit *Tyrosinase* (*Tyr*) and *Duchenne muscular dystrophy* (*DMD*) in mouse embryos, which modeled Himalayan mouse type and rescued Duchenne muscular dystrophy, respectively [82]. Liu *et al.* also used ABE7.10 to introduce mutations in *Androgen Receptor* (*AR*) and *Homeobox protein-D13* (*HOXD13*) in mice embryos and the relevant phenotypes of sex reversal and fused digits were observed [91]. Meanwhile, plant versions of ABEs have been also developed and applied. In rice, Hua *et al.* developed the ABE-P1 and ABE-P2 to induce mutations in six genes [92] and Yan *et al.* constructed rBE14 to introduce mutations in four genes [93].

Shortly after its first report, PE has been already applied in plants. Lin *et al.* developed plant versions of PEs (PPEs) to induce precise editing in rice and wheat [94]. Meanwhile, Li *et al.* and Xu *et al.* also used PEs to introduce mutations in rice with high precision [95,96]. We envision that other precise editing applications by PEs, such as in animal embryos and somatic cells, will be booming.

#### Understanding OT Mechanisms to Achieve Better Genome Editing

#### OT Binding by Mismatched Pairing of gRNA with Nonspecific Sites

With the broad applications of CRISPR/Cas genome editing in biomedical and translational research, unintended OT effects were widely reported at nontargeted sites in the genome [97–99], hindering their potential in cases when precise genome change is required. Most of these OT effects were caused by the nonspecific binding of gRNA to potential OT sites with mismatch(es) compared with the on-target (ON) site (Figure 3A) [100,101]. This type of OT sites can be cataloged or predicted by searching sites with high sequence similarity to the ON site [102–104]. Thus, a common and practical strategy to reduce OT effects is to find a unique ON site that has maximal sequence difference from other sites in the genome.





**Trends in Biochemical Sciences** 

Figure 3. Distinct Mechanisms of Off-Target (OT) Effects in CRISPR Base Editing Technologies. (A) Nonspecific binding of gRNA to OT sites with mismatch (es). Different OT effects can be resulted due to the OT binding of gRNA, such as OT double-strand breaks (DSBs) by Cas9 nuclease, OT nicks by Cas9 nickase (D10A), and OT base editing by Cas9-BE. (B) Formation of unwanted indels by nCas9-base technologies, including nCas9-BE. APOBECs and a series of DNA repair enzymes participate in the formation of unwanted indels near the nicking site by nCas9 and nCas9-BE. (C) Unintended C-to-T mutation can be catalyzed by the cytidine deaminase moiety of CBEs at OT genomic sites independent of gRNA/Cas9 locator. (D) Unintended C-to-U editing can be catalyzed by the cytidine deaminase moiety of CBEs in RNA independent of gRNA/Cas9 locator. (E) Unintended A-to-I editing can be catalyzed by the adenosine deaminase moiety of ABEs in RNA independent of gRNA/Cas9 locator. Abbreviations: ABE, Adenine base editor; APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; BE, base editor; cytosine base editor; gRNA, guide RNA; PAM, protospacer-adjacent motif; UGI, uracil DNA glycosylase inhibitor.

#### OT Effects at gRNA/Cas-Dependent Sites

After the binding of gRNA/Cas at OT sites with mismatches, it is the moiety of catalytic Cas protein or other conjugated effector enzymes, such as deaminases in BEs, editing DNA to result in unintended OT effects [105–107]. For instance, a gRNA was originally designed to guide a Cas9 nuclease to generate indels at ON sites. However, when bound at OT sites, Cas9 nuclease can also cut DNA double-strand to trigger unintended indels (Figure 3A) [100,101]. To inhibit these gRNA/Cas-dependent OT indels, nCas9 is applied with a pair of offset gRNAs targeting the upstream and downstream regions of ON sites to improve specificity [108,109]. In this case, nCas9 generates two opposite DNA SSBs at the ON site, but likely only an SSB at a specific OT site, which avoids triggering unintended indels at OT sites by DSBs. However, in a previous



study for reducing OT effects of nCas9, Tsai *et al.* found that nCas9 monomer could induce unintended C-to-T base substitutions in the R-loop region at ON sites and most of the mutated cytosines were in TpC dinucleotides, manifesting a typical APOBEC mutational signature [110]. Meanwhile, Chen *et al.* also found that endogenous human APOBEC family members can induce C-to-T base substitutions during the repair of a DNA nick in an episomal shutter vector [111]. These studies thus implied the possible mechanism of unintended point mutations in the nCas9-processed genome editing through the crosstalk between APOBEC and CRISPR/Cas9. Later on, nCas9-generated SSBs, including those by nCas9-CBEs, were also found to induce indels at some OT sites because these SSBs could be converted to DSBs through the steps involving endogenous APOBEC CDAs and DNA repair proteins (Figure 3B) [71]. Thus, repression of endogenously expressed APOBECs can inhibit these unwanted indels at nCas9-generated SSB sites [71].

#### OT Effects at Nonspecific Binding Sites by Deaminase Effectors in BEs

In addition to those aforementioned OT effects in a gRNA/Cas-dependent manner, gRNA/Casindependent OT effects were also identified in recently developed BE systems. In mice and plants treated with several versions of CBEs that contain different APOBEC CDAs, unintended C-to-T mutations were identified at OT sites that have no sequence similarity to ON sites [112–114], indicating that these unintended C-to-T mutations occur independent of the gRNA/Cas moiety (Figure 3C). Despite being used to perform DNA C-to-T base editing, some CBEs were found to induce massive C-to-U editing in transcriptome RNAs (Figure 3D) [115,116]. These findings are unexpected but not totally surprising, because APOBEC CDAs intrinsically bind both RNA and ssDNA substrates for cytidine deamination [52–54]. Specifically, APOBEC1 was originally discovered to induce C-to-U editing in apolipoprotein B mRNA [117]. Later on, AID, APOBEC3, and their homologs were found to commonly trigger C-to-U deamination in ssDNA regions generated during various cellular processes (e.g., transcription, DNA replication, or repair) [111,118–121]. Indeed, a significant amount of mutations in tumor genomes were identified to be related with APOBEC activity [122,123]. In this case, a strategy to reduce OT effects of CBEs is to engineer their deaminase effectors [115,116].

Although evolved to perform A-to-G DNA editing, the TadA-TadA\* heterodimer deaminase in ABEs did not likely induce global OT effects at genomic DNAs. However, its original function of RNA adenosine deamination might contribute to the observed massive A-to-I OT editing in transcriptome RNAs (Figure 3E) [115,116]. Correspondingly, by mutating the residues of TadA-TadA\* involved in RNA binding, the RNA OT editing by ABEs was greatly reduced with little effect on the DNA ON editing [116,124].

In the most recently developed PE systems, an RTase from murine retrovirus was used to achieve versatile genome editing. While the conjugated RTase effector in PEs seemed harmless to cell viability and transcriptomic gene expression [18], whether it induces genome- or transcriptome-wide OT effects or not remains unexploited.

#### Strategies to Reduce OT Effects of BEs

Different strategies can be applied to reduce OT editing in BEs by tethering their locator and/or effector moieties. It has been reported that, by changing residues involved in the interaction between Cas9 protein and deoxyribose backbone, engineered Cas9 proteins, (e.g., eSpCas9 [125], SpCas9-HF [126], HypaCas9 [127], and Sniper-Cas9 [128]) could reduce their binding at OT sites, but their binding and editing ability at ON sites largely remain. Meanwhile, the modification of gRNA has been also reported to eliminate OT effects, such as by altering the length of spacer sequence in gRNA [100, 129] or by adding an RNA secondary structure at the 5' end of



gRNA [130]. In principle, these engineered Cas9 proteins and modified gRNAs can be adapted to develop new BEs with high specificity. Since delivery methods also affect the specificity of CRISPR/Cas9-mediated gene editing [106,131–133], the delivery of RNP complex or RNA of BEs can offer higher editing specificity than that of plasmid DNA.

In addition, different engineering strategies have been applied to modify the APOBEC effector in BEs to reduce unwanted genome- and transcriptome-wide editing [115,116,124,134]. For example, mutating the APOBEC residues involved in RNA binding could greatly reduce the RNA OT editing while maintaining DNA editing capability [115,116,124]. However, structure analysis showed that only one active CDA domain in APOBEC was responsible for both DNA binding and deamination [53,54,135]. This finding suggests that mutating the active CDA domain could lead to controversial consequences, possibly with suppressed editing at both ON and OT sites in the genome.

#### **Concluding Remarks and Future Perspectives**

In view of two moieties of genome editing, a locator and an effector are mainly required to fulfill different genome editing purposes. In the last decade, the locator moiety has evolved from ZF and TALE proteins to CRISPR/Cas nucleoproteins. With great convenience, efficiency, and precision, CRISPR/Cas systems (e.g., CRISPR/Cas9 and CRISPR/Cas12a) have been dominantly chosen for single gene KO and genome-wide screening. Moreover, CRISPR/Cas proteins have been widely used to develop a variety of genome engineering technologies, such as fusing or recruiting transcription activator/repressor, fluorescent protein or transposase to perform transcription activation/repression [136], nucleic acid imaging [137,138], or targeted gene integration [139,140]. More strikingly, by tethering gRNA/Cas locators to catalytically active effectors with DNA processing activities (e.g., nucleotide deaminase and reverse transcriptase), BEs or PEs were recently shown to enable precise editing with high efficiency and versatility, lifting genome editing to a new height. Although questions regarding developing reliable genome editing tools for in vivo application and especially for clinic trials still remain (see Outstanding Questions), great efforts have been made to better understand mechanisms of the specificity, efficiency, and OT effects of these newly emerging technologies. We envision that better CRISPR/Cas-evolved genome editing systems will be invented to not only facilitate the research in biomedical fields, but also shed new light on treatments of human genetic diseases.

#### Acknowledgments

We are grateful to Ling-Ling Chen for critical reading. We apologize to colleagues whose studies could not be discussed here owing to space/content limitations. Our work is supported by grants 2019YFA0802804 (L.Y.), 2018YFA0801401 (J.C.), and 2018YFC1004602 (J.C.) from National Key R&D Program of China and 31925011 (L.Y.), 91940306 (L.Y.), 31822016 (J.C.), and 81872305 (J.C.) from National Natural Science Foundation of China.

#### References

- Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
   8.
- 2. Venter, J.C. *et al.* (2001) The sequence of the human genome. Science 291, 1304–1351
- 3. Metzker, M.L. (2010) Sequencing technologies the next generation. *Nat. Rev. Genet.* 11, 31–46
- Beerli, R.R. and Barbas III, C.F. (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol.* 20, 135–141
- Boch, J. and Bonas, U. (2010) Xanthomonas AvrBs3 familytype III effectors: discovery and function. Annu. Rev. Phytopathol. 48, 419–436
- Hsu, P.D. et al. (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278
- Wright, A.V. et al. (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell 164, 29–44

- . Komor, A.C. *et al.* (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* 168, 20–36
- Jinek, M. et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science 343, 1247997
- Ran, F.A. et al. (2015) In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186–191
- 11. Zetsche, B. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771
- Abudayyeh, O.O. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573
- 13. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823
- 14. Jinek, M. et al. (2013) RNA-programmed genome editing in human cells. Elife 2, e00471

#### **Outstanding Questions**

In addition to cysteine/adenine deaminases and reverse transcriptases, can other types of effectors be tethered for developing novel CRISPR/Cas-based genome editing tools?

Is it possible to fuse deaminase activators to specifically enhance editing efficiency at target sites, or repressor to dampen editing efficiency at off-target sites?

Can off-target effects by BEs be feasibly examined by simple methods, rather than genome and transcriptome sequencing?

Can newly developed PEs induce global off-target effects?

How can large editing tools (e.g., BEs and PEs) be efficiently delivered *in vivo* to achieve desired genetic changes?

**CellPress** REVIEWS

- 15. Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823–826
- Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424
- Gaudelli, N.M. et al. (2017) Programmable base editing of A\*T to G\*C in genomic DNA without DNA cleavage. Nature 551, 464–471
- Anzalone, A.V. et al. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149–157
- Bolotin, A. *et al.* (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561
- Mojica, F.J. et al. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–182
- Pourcel, C. et al. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653–663
- Barrangou, R. et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712
- Makarova, K.S. et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67–83
- Jinek, M. et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821
- Gasiunas, G. et al. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. 109, E2579–E2586
- Anders, C. *et al.* (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573
- 27. Nishimasu, H. *et al.* (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949
- Sternberg, S.H. *et al.* (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67
- Kleinstiver, B.P. et al. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485
- Hu, J.H. et al. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556, 57–63
- Nishimasu, H. et al. (2018) Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259–1262
- Miller, S.M. et al. (2020) Continuous evolution of SpCas9 variants compatible with non-G PAMs. Nat. Biotechnol. 38, 471–481
- Walton, R.T. et al. (2020) Unconstrained genome targeting with near-PAMIess engineered CRISPR-Cas9 variants. Science 368, 290–296
- 34. Abudayyeh, O.O. *et al.* (2017) RNA targeting with CRISPR-Cas13. *Nature* 550, 280–284
- Konermann, S. et al. (2018) Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell 173, 665–676
- 36. Cox, D.B.T. *et al.* (2017) RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027
- Abudayyeh, O.O. et al. (2019) A cytosine deaminase for programmable single-base RNA editing. Science 365, 382–386
- Wang, F. et al. (2019) Advances in CRISPR-Cas systems for RNA targeting, tracking and editing. *Biotechnol. Adv.* 37, 708–729
- Smargon, A.A. *et al.* (2020) RNA-targeting CRISPR systems from metagenomic discovery to transcriptomic engineering. *Nat. Cell Biol.* 22, 143–150
- Deriano, L. and Roth, D.B. (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu. Rev. Genet. 47, 433–455
- Jasin, M. and Rothstein, R. (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* 5, a012740
- Ceccaldi, R. et al. (2016) Repair pathway choices and consequences at the double-strand break. Trends Cell Biol. 26, 52–64
- 43. Yamano, T. *et al.* (2016) Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* 165, 949–962

- 44. Liu, L. *et al.* (2017) The molecular architecture for RNA-guided RNA cleavage by Cas13a. *Cell* 170, 714–726
- Qi, L.S. *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173–1183
- Tak, Y.E. et al. (2017) Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods 14, 1163–1166
- Yang, L.Z. et al. (2019) Dynamic imaging of RNA in living cells by CRISPR-Cas13 systems. Mol. Cell 76, 981–997
- Landrum, M.J. *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44, D862–D868
- Wang, Y. et al. (2019) Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. *Genome Biol.* 20, 218
- Rees, H.A. and Liu, D.R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788
- Nishida, K. et al. (2016) Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729
- Harris, R.S. and Liddament, M.T. (2004) Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* 4, 868–877
- Salter, J.D. *et al.* (2016) The APOBEC protein family: united by structure, divergent in function. *Trends Biochem. Sci.* 41, 578–594
- 54. Yang, B. et al. (2017) APOBEC: from mutator to editor. J. Genet. Genomics 44, 423–437
- Beard, W.A. et al. (2019) Eukaryotic base excision repair: new approaches shine light on mechanism. Annu. Rev. Biochem. 88, 137–162
- Mullins, E.A. et al. (2019) Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* 44, 765–781
- Mol, C.D. et al. (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. Cell 82, 701–708
- Kunkel, T.A. and Erie, D.A. (2015) Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.* 49, 291–313
- Li, X. et al. (2018) Base editing with a Cpf1-cytidine deaminase fusion. Nat. Biotechnol. 36, 324–327
- Kim, K. et al. (2017) Highly efficient RNA-guided base editing in mouse embryos. Nat. Biotechnol. 35, 435–437
- Zong, Y. et al. (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat. Biotechnol. 35, 438–440
- 62. Yang, B. et al. (2019) Development and application of base editors. CRISPR J. 2, 91–104
- Wang, X. et al. (2018) Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. Nat. Biotechnol. 36, 946–949
- Zong, Y. *et al.* (2018) Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* 36, 950–953
- Li, J. et al. (2019) Efficient base editing in G/C-rich regions to model androgen insensitivity syndrome. Cell Res. 29, 174–176
- Liu, Z. et al. (2019) Efficient and precise base editing in rabbits using human APOBEC3A-nCas9 fusions. Cell Discov. 5, 31
- Xie, J. et al. (2019) Efficient base editing for multiple genes and loci in pigs using base editors. Nat. Commun. 10, 2852
- Kim, Y.B. et al. (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. 35, 371–376
- Gehrke, J.M. et al. (2018) An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat. Biotechnol. 36, 977–982
- Thuronyi, B.W. et al. (2019) Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* 37, 1070–1079
- Lei, L. et al. (2018) APOBEC3 induces mutations during repair of CRISPR-Cas9-generated DNA breaks. Nat. Struct. Mol. Biol. 25, 45–52
- Komor, A.C. et al. (2017) Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A



base editors with higher efficiency and product purity. *Sci. Adv.* 3, eaao4774

- Wang, L. et al. (2017) Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. Cell Res. 27, 1289–1292
- Kleinstiver, B.P. et al. (2019) Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* 37, 276–282
- Wang, X. et al. (2020) Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* 31, 107723
- Hess, G.T. et al. (2016) Directed evolution using dCas9targeted somatic hypermutation in mammalian cells. Nat. Methods 13, 1036–1042
- Ma, Y. et al. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* 13, 1029–1035
- Ramaswami, G. *et al.* (2013) Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* 10, 128–132
- Zhu, S. *et al.* (2013) Prediction of constitutive A-to-I editing sites from human transcriptomes in the absence of genomic sequences. *BMC Genomics* 14, 206
- Eisenberg, E. and Levanon, E.Y. (2018) A-to-I RNA editing immune protector and transcriptome diversifier. *Nat. Rev. Genet.* 19, 473–490
- Li, C. *et al.* (2018) Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 19, 59
- Ryu, S.M. et al. (2018) Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* 36, 536–539
- Richter, M.F. et al. (2020) Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat. Biotechnol. Published online March 15, 2020. https://doi.org/10.1038/s41587-020-0453-z
- Jansen, J.G. *et al.* (2006) Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J. Exp. Med.* 203, 319–323
- Li, C. et al. (2020) Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* Published online January 13, 2020. https://doi.org/10.1038/ s41587-019-0393-7
- Zhang, X. et al. (2020) Dual base editor catalyzes both cytosine and adenine base conversions in human cells. Nat. Biotechnol. https://doi.org/10.1038/s41587-020-0527-y. Published online June 1, 2020
- Yang, L. *et al.* (2019) One prime for all editing. *Cell* 179, 1448–1450
- Chadwick, A.C. et al. (2017) In vivo base editing of PCSK9 (proprotein convertase subtilisin/kexin type 9) as a therapeutic alternative to genome editing. Arterioscler. Thromb. Vasc. Biol. 37, 1741–1747
- Villiger, L. *et al.* (2018) Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. Nat. Med. 24, 1519–1525
- Zeng, J. *et al.* (2020) Therapeutic base editing of human hematopoietic stem cells. *Nat. Med.* 26, 535–541
- Liu, Z. *et al.* (2018) Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat. Commun.* 9, 2338
- Hua, K. et al. (2018) Precise A.T to G.C base editing in the rice genome. Mol. Plant 11, 627–630
- Yan, F. *et al.* (2018) Highly efficient A.T to G.C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Mol. Plant* 11, 631–634
- 94. Lin, Q. et al. (2020) Prime genome editing in rice and wheat. Nat. Biotechnol. 38, 582–585
- Li, H. et al. (2020) Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* 13, 671–674
- Xu, W. et al. (2020) Versatile nucleotides substitution in plant using an improved prime editing system. Mol. Plant 13, 675–678
- Fu, Y. et al. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826

- Hsu, P.D. et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832
- Pattanayak, V. *et al.* (2013) High-throughput profiling of offtarget DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 31, 839–843
- 100. Kim, D. et al. (2015) Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat. Methods 12, 237–243
- Tsai, S.Q. *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197
- 102. Bae, S. et al. (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30, 1473–1475
- Kim, H.K. et al. (2018) Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. Nat. Biotechnol. 36, 239–241
- Listgarten, J. et al. (2018) Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. Nat. Biomed. Eng. 2, 38–47
- Kim, D. et al. (2017) Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat. Biotechnol. 35, 475–480
- Rees, H.A. et al. (2017) Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat. Commun. 8, 15790
- 107. Kim, D. et al. (2019) Genome-wide target specificity of CRISPR RNA-guided adenine base editors. Nat. Biotechnol. 37, 430–435
- Ran, F.A. et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389
- Shen, B. et al. (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. Nat. Methods 11, 399–402
- Tsai, S.Q. *et al.* (2014) Dimeric CRISPR RNA-guided Fokl nucleases for highly specific genome editing. *Nat. Biotechnol.* 32, 569–576
- 111. Chen, J. *et al.* (2014) Repair of naturally occurring mismatches can induce mutations in flanking DNA. *Elife* 3, e02001
- 112. Chen, J. et al. (2019) To BE or not to BE, that is the question. Nat. Biotechnol. 37, 520–522
- Jin, S. *et al.* (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 364, 292–295
- Zuo, E. *et al.* (2019) Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 364, 289–292
- Grunewald, J. et al. (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* 569, 433–437
- Zhou, C. et al. (2019) Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* 571, 275–278
- 117. Lau, P.P. et al. (1990) A 40 kilodalton rat liver nuclear protein binds specifically to apolipoprotein B mRNA around the RNA editing site. *Nucleic Acids Res.* 18, 5817–5821
- Kazanov, M.D. et al. (2015) APOBEC-induced cancer mutations are uniquely enriched in early-replicating, gene-dense, and active chromatin regions. *Cell Rep.* 13, 1103–1109
- Seplyarskiy, V.B. et al. (2016) APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Bes.* 26, 174–182
- Roberts, S.A. et al. (2012) Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* 46, 424–435
- 121. Chen, J. and Furano, A.V. (2015) Breaking bad: the mutagenic effect of DNA repair. *DNA Repair (Amst)* 32, 43–51
- Burns, M.B. et al. (2013) APOBEC3B is an enzymatic source of mutation in breast cancer. Nature 494, 366–370
- Burns, M.B. et al. (2013) Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. 45, 977–983
- Grunewald, J. *et al.* (2019) CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* 37, 1041–1048



- 125. Slaymaker, I.M. *et al.* (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88
- Kleinstiver, B.P. et al. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495
- 127. Chen, J.S. et al. (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550, 407–410
- Lee, J.K. et al. (2018) Directed evolution of CRISPR-Cas9 to increase its specificity. Nat. Commun. 9, 3048
- 129. Fu, Y. et al. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 32, 279–284
- Kocak, D.D. et al. (2019) Increasing the specificity of CRISPR systems with engineered RNA secondary structures. Nat. Biotechnol. 37, 657–666
- Kim, S. et al. (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24, 1012–1019
- Ramakrishna, S. et al. (2014) Gene disruption by cellpenetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res. 24, 1020–1027
- 133. Vin, H. et al. (2017) Structure-guided chemical modification of guide RNA enables potent non-viral *in vivo* genome editing. *Nat. Biotechnol.* 35, 1179–1187
- Doman, J.L. et al. (2020) Evaluation and minimization of Cas9independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* 38, 620–628
- Shi, K. et al. (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* 24, 131–139
- Gilbert, L.A. et al. (2014) Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661
- 137. Chen, B. et al. (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell 155, 1479–1491
- Ma, H. et al. (2015) Multicolor CRISPR labeling of chromosomal loci in human cells. Proc. Natl. Acad. Sci. U. S. A. 112, 3002–3007
- Klompe, S.E. et al. (2019) Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature 571, 219–225
- Strecker, J. *et al.* (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 365, 48–53

- Rouet, P. et al. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rarecutting endonuclease. *Mol. Cell. Biol.* 14, 8096–8106
- Epinat, J.C. *et al.* (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* 31, 2952–2962
- Miller, J. et al. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 4, 1609–1614
- Klug, A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.* 79, 213–231
- Kim, Y.G. et al. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U. S. A. 93, 1156–1160
- Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252, 809–817
- Rebar, E.J. and Pabo, C.O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 263, 671–673
- Bibikova, M. et al. (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161, 1169–1175
- Miller, J.C. et al. (2011) A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29, 143–148
- 150. Boch, J. et al. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509–1512
- Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501
- Cermak, T. et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 39, e82
- Christian, M. et al. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186, 757–761
- 154. Mali, P. et al. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838
- Guilinger, J.P. et al. (2014) Fusion of catalytically inactive Cas9 to Fokl nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 32, 577–582
- Rossidis, A.C. et al. (2018) In utero CRISPR-mediated therapeutic editing of metabolic genes. Nat. Med. 24, 1513–1518