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

Li Yang\(^1,2,*\) and Jia Chen\(^2,3,*\)

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 FokI 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.

1 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, Shanghai 200031, China
2 School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
3 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 FokI 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 TALE system uses TALE repeats, from a bacterial plant pathogen Xanthomonas, as the locator [9]. 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 FokI 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 proteins 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...
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].

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 locater 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.
Table 1. Representative Genome Editors

<table>
<thead>
<tr>
<th>Genome editor</th>
<th>Locator</th>
<th>Effector</th>
<th>PAM</th>
<th>Locator-dependent OT effects</th>
<th>Locator-independent OT effects</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZFN</td>
<td>ZF motif</td>
<td>FokI nuclease</td>
<td>–</td>
<td>+++</td>
<td>–</td>
<td>[145,148]</td>
</tr>
<tr>
<td>TALEN</td>
<td>TALE repeat</td>
<td>FokI nuclease</td>
<td>–</td>
<td>+++</td>
<td>–</td>
<td>[149,152,153]</td>
</tr>
<tr>
<td>Cas9</td>
<td>Cas9/gRNA</td>
<td>Cas9 nuclease</td>
<td>NGG</td>
<td>+++</td>
<td>–</td>
<td>[13,14,154]</td>
</tr>
<tr>
<td>Cas12a</td>
<td>Cas12a/gRNA</td>
<td>Cas12a nuclease</td>
<td>TTTV</td>
<td>++</td>
<td>–</td>
<td>[11]</td>
</tr>
<tr>
<td>Cas9-VQR</td>
<td>Cas9-VQR/gRNA</td>
<td>Cas9 nuclease</td>
<td>NGA</td>
<td>+++</td>
<td>–</td>
<td>[29]</td>
</tr>
<tr>
<td>xCas9</td>
<td>xCas9/gRNA</td>
<td>Cas9 nuclease</td>
<td>NG, GAA, GAT</td>
<td>++</td>
<td>–</td>
<td>[30]</td>
</tr>
<tr>
<td>Cas9-NG</td>
<td>Cas9-NG/gRNA</td>
<td>Cas9 nuclease</td>
<td>NG</td>
<td>+++</td>
<td>–</td>
<td>[31]</td>
</tr>
<tr>
<td>Cas9-NRRH</td>
<td>Cas9-NRRH/gRNA</td>
<td>Cas9 nuclease</td>
<td>NRRH</td>
<td>+++</td>
<td>–</td>
<td>[32]</td>
</tr>
<tr>
<td>Cas13a</td>
<td>Cas13a/gRNA</td>
<td>Cas13a nuclease</td>
<td>H</td>
<td>+</td>
<td>–</td>
<td>[34]</td>
</tr>
<tr>
<td>nCas9</td>
<td>nCas9/gRNA</td>
<td>nCas9</td>
<td>NGG</td>
<td>++</td>
<td>–</td>
<td>[108,109]</td>
</tr>
<tr>
<td>dCas9-FokI</td>
<td>dCas9/gRNA</td>
<td>FokI nuclease</td>
<td>NGG</td>
<td>+</td>
<td>–</td>
<td>[110,155]</td>
</tr>
<tr>
<td>eSpCas9</td>
<td>eSpCas9/gRNA</td>
<td>Cas9 nuclease</td>
<td>NGG</td>
<td>–</td>
<td>–</td>
<td>[125]</td>
</tr>
<tr>
<td>SpCas9-HF</td>
<td>SpCas9-HF/gRNA</td>
<td>Cas9 nuclease</td>
<td>NGG</td>
<td>–</td>
<td>–</td>
<td>[126]</td>
</tr>
<tr>
<td>HyaCas9</td>
<td>HyaCas9/gRNA</td>
<td>Cas9 nuclease</td>
<td>NGG</td>
<td>–</td>
<td>–</td>
<td>[127]</td>
</tr>
<tr>
<td>Sniper-Cas9</td>
<td>Sniper-Cas9/gRNA</td>
<td>Cas9 nuclease</td>
<td>NGG</td>
<td>–</td>
<td>–</td>
<td>[128]</td>
</tr>
<tr>
<td>BE3</td>
<td>nCas9/gRNA</td>
<td>rA1</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[16]</td>
</tr>
<tr>
<td>YE1-BE3</td>
<td>nCas9/gRNA</td>
<td>rA1-YE1</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: –, RNA: –</td>
<td>[69]</td>
</tr>
<tr>
<td>BE4</td>
<td>nCas9/gRNA</td>
<td>rA1</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[72]</td>
</tr>
<tr>
<td>eBE</td>
<td>nCas9/gRNA</td>
<td>rA1</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[73]</td>
</tr>
<tr>
<td>hA3A-BE3</td>
<td>nCas9/gRNA</td>
<td>hA3A</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[63]</td>
</tr>
<tr>
<td>hA3A-BE3-Y130F</td>
<td>nCas9/gRNA</td>
<td>hA3A-Y130F</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: –</td>
<td>[63]</td>
</tr>
<tr>
<td>hA3A-BE3-Y132D</td>
<td>nCas9/gRNA</td>
<td>hA3A-Y132D</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[63]</td>
</tr>
<tr>
<td>eA3A-BE3</td>
<td>nCas9/gRNA</td>
<td>A3A-N57Q</td>
<td>NGG</td>
<td>+</td>
<td>DNA: +++, RNA: +</td>
<td>[69]</td>
</tr>
<tr>
<td>SaKKh-BE3</td>
<td>nSaKKhCas9/gRNA</td>
<td>rA1</td>
<td>NNNRRT</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[68]</td>
</tr>
<tr>
<td>Target-AID</td>
<td>nCas9/gRNA</td>
<td>Sea lamprey CDA</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: –</td>
<td>[61]</td>
</tr>
<tr>
<td>dCas12a-BE</td>
<td>dLbCas12a/gRNA</td>
<td>rA1</td>
<td>TTTV</td>
<td>++</td>
<td>DNA: +++, RNA: +</td>
<td>[59]</td>
</tr>
<tr>
<td>BEACON1</td>
<td>dLbCas12a/gRNA</td>
<td>Engineered hA3A</td>
<td>TTTV</td>
<td>++</td>
<td>DNA: +++, RNA: +</td>
<td>[75]</td>
</tr>
<tr>
<td>BEACON2</td>
<td>dLbCas12a/gRNA</td>
<td>Engineered hA3A</td>
<td>TTTV</td>
<td>++</td>
<td>DNA: +++, RNA: –</td>
<td>[75]</td>
</tr>
<tr>
<td>enAsBE</td>
<td>denAsCas12a/gRNA</td>
<td>rA1</td>
<td>VTTV, TTTT, TTCN/TATV</td>
<td>+</td>
<td>DNA: +++, RNA: +</td>
<td>[74]</td>
</tr>
<tr>
<td>PBE</td>
<td>nCas9/gRNA</td>
<td>rA1</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[61]</td>
</tr>
<tr>
<td>A3A-PBE</td>
<td>nCas9/gRNA</td>
<td>hA3A</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[64]</td>
</tr>
<tr>
<td>ABE7.10</td>
<td>nCas9/gRNA</td>
<td>TadA-TadA*</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: –, RNA: –</td>
<td>[17]</td>
</tr>
<tr>
<td>ABE8e</td>
<td>nCas9/gRNA</td>
<td>TadA-TadA-8e</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: –, RNA: –</td>
<td>[63]</td>
</tr>
<tr>
<td>ABE8e-V106W</td>
<td>nCas9/gRNA</td>
<td>TadA-TadA-8e-V106W</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: –, RNA: –</td>
<td>[83]</td>
</tr>
<tr>
<td>LbABE9e</td>
<td>dLbCas12a/gRNA</td>
<td>TadA-TadA-8e</td>
<td>TTTV</td>
<td>++</td>
<td>DNA: +++, RNA: +</td>
<td>[83]</td>
</tr>
<tr>
<td>STEME-1</td>
<td>nCas9/gRNA</td>
<td>hA3A-TadA-TadA*</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: +++, RNA: +</td>
<td>[63]</td>
</tr>
<tr>
<td>ABE-P1</td>
<td>nCas9/gRNA</td>
<td>TadA-TadA*</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: –, RNA: –</td>
<td>[92]</td>
</tr>
<tr>
<td>ABE-P2</td>
<td>nSaCas9/gRNA</td>
<td>TadA-TadA*</td>
<td>NNGRRRT</td>
<td>+++</td>
<td>DNA: –, RNA: –</td>
<td>[92]</td>
</tr>
<tr>
<td>rBE14</td>
<td>nCas9/gRNA</td>
<td>TadA-TadA*</td>
<td>NGG</td>
<td>+++</td>
<td>DNA: –, RNA: +</td>
<td>[93]</td>
</tr>
<tr>
<td>PE1</td>
<td>dCas9/gRNA</td>
<td>M-MLV RTase</td>
<td>NGG</td>
<td>+</td>
<td>DNA: ?, RNA: ?</td>
<td>[18]</td>
</tr>
</tbody>
</table>
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 then be 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

Table 1. (continued)

<table>
<thead>
<tr>
<th>Genome editor</th>
<th>Locator</th>
<th>Effector</th>
<th>PAM</th>
<th>Locator-dependent OT effects</th>
<th>Locator-independent OT effects</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE2</td>
<td>nCas9/gRNA</td>
<td>M-MLV RTase</td>
<td>NGG</td>
<td>+</td>
<td>DNA: ?, RNA: ?</td>
<td>[18]</td>
</tr>
<tr>
<td>PE3</td>
<td>nCas9/gRNA</td>
<td>M-MLV RTase</td>
<td>NGG</td>
<td>+</td>
<td>DNA: ?, RNA: ?</td>
<td>[18]</td>
</tr>
<tr>
<td>PPE</td>
<td>nCas9/gRNA</td>
<td>M-MLV RTase</td>
<td>NGG</td>
<td>+</td>
<td>DNA: ?, RNA: ?</td>
<td>[94]</td>
</tr>
</tbody>
</table>

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, 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.
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, eaA3A 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 deamination activity in ssDNA [17]. ABEs were then constructed by fusing a TadA-TadA* heterodimer 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 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
Conjugating Nucleobase Deaminases or Reverse Transcriptases with CRISPR/Cas Proteins to Achieve Precise Genome Editing at Single Nucleotide Resolution

Key Figure

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 [34]. 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.
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/Cas-independent 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 editing activity than that of plasmid DNA. 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.

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 clinical 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.), and 81872305 (J.C.) from National Natural Science Foundation of China.

References


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?
Trends in Biochemical Sciences

49. Wang, Y. et al. (2018) Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. Genome Biol. 20, 218
64. Li, J. et al. (2019) Efficient base editing in G/C-rich regions to model androgen insensitivity syndrome. Cell 176, 174–176


Trends in Biochemical Sciences, Month 2020, Vol. xx, No. xx