nature biotechnology

FOCUS | REVIEW ARTICLE

https://doi.org/10.1038/s41587-020-0561-9

Check for updates

# Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors

Andrew V. Anzalone[1,2,3,4], Luke W. Koblan[1,2,3,4] and David R. Liu[1,2,3] ✉

**The development of new CRISPR–Cas genome editing tools continues to drive major advances in the life sciences. Four classes of CRISPR–Cas-derived genome editing agents—nucleases, base editors, transposases/recombinases and prime editors—are currently available for modifying genomes in experimental systems. Some of these agents have also moved rapidly into the clinic. Each tool comes with its own capabilities and limitations, and major efforts have broadened their editing capabilities, expanded their targeting scope and improved editing specificity. We analyze key considerations when choosing genome editing agents and identify opportunities for future improvements and applications in basic research and therapeutics.**

The first demonstrations of programmable DNA cleavage by Cas9 nuclease[1,2] and subsequent early demonstrations of its ability to carry out targeted genome modification in living eukaryotic cells[3–7] initiated an explosive growth in the discovery, engineering and application of CRISPR–Cas genome editing tools. In this Review, we focus on CRISPR technologies that collectively bring us closer to realizing the longstanding aspiration of being able to install any genetic change at any position within the genome of any living cell with minimal unwanted genome modification or cellular perturbation. Our objectives are to present readers with an overview of the capabilities and limitations of current CRISPR technologies in a way that facilitates tool selection and to highlight opportunities for future improvement. We restrict our discussion to the targeted alteration of genomic DNA sequence using CRISPR-based tools and refer readers to excellent reviews of topics related to other CRISPR applications such as transcriptional regulation[8–12], epigenetic modifications[9–12], RNA editing[11,12] and nucleic acid detection[12].

The goal of a genome editing experiment is to convert a targeted DNA sequence into a new, desired DNA sequence (or sequences) in the native context of a cell's genome. In most cases, a single sequence product is desired in high yield, but depending on the application, heterogeneity among edited sequences may be acceptable or even preferred. In designing a strategy to achieve the intended sequence transformation, it is necessary to consider several factors.

The type of edit desired determines which classes of editing agents are suitable (Fig. 1). Common desired targeted edits include the following: (i) conversion of DNA base pairs (that is, point mutations), (ii) deletion of DNA base pairs, (iii) insertion of DNA base pairs, or (iv) a combination of the above changes (including replacement of DNA base pairs). Different classes of CRISPR–Cas editing agents mediate each of these types of changes. Other alterations, such as inversion of a DNA segment or chromosomal translocation, may also be desired. Desired edits can be further classified on the basis of the size of the sequence alteration, as well as the required efficiency and product purity (that is, what fraction of edited products must be the desired sequence, with no undesired byproducts).

The ability of the current arsenal of CRISPR–Cas proteins to engage the target DNA sequence is a second major consideration. Targeting by all DNA-targeting CRISPR–Cas systems described to date requires that a short sequence known as a protospacer-adjacent motif, or PAM, occur near the target DNA site[13]. Naturally occurring CRISPR systems provide a variety of PAM options, and still more engineered variants have been developed with broadened or altered PAM compatibility. The optimal selection of a particular tool must take into consideration the availability of PAM sequences at the target locus, while recognizing that not all CRISPR-based tools are compatible with all PAM-variant Cas protein modules. Most CRISPR editing methods further require that the number of nucleotides between the PAM and the position of the desired edit fall within a certain range.

Finally, the optimal choice of CRISPR–Cas tool and the overall editing strategy will also depend on the intended application. Factors such as cell type (for example, bacteria, yeasts, mammalian cancer cell lines or mammalian postmitotic cells), cellular environment (for example, cell culture, organoid or in vivo), form of the agent (for example, plasmid DNA, ribonucleoprotein (RNP) complex, mRNA or viral vector) and method of delivery (for example, lipid-mediated, electroporation, nucleofection or viral infection) each impose different constraints, as well as different propensities for undesired genome modification events.

In this Review, we navigate these considerations and analyze recent developments that have progressively increased the applicability and effectiveness of CRISPR-based genome editing technologies. We begin by describing the naturally occurring variants of Cas9 and Cas12 nucleases that have been characterized and detail the development of Cas9 and Cas12 nuclease variants with broadened targeting scope and specificity. Next we discuss the development and application of base editors, genome-editing agents that precisely install point mutations without requiring double-stranded DNA breaks (DSBs) or donor DNA templates. Finally, we summarize emerging CRISPR–Cas genome editing tools, including Cas transposons and recombinases, which mediate rearrangements of large segments of DNA, and prime editors, which directly copy edited sequences into target DNA sites in a manner that replaces the original DNA sequence.

## Genome editing with CRISPR–Cas nucleases

In nature, bacteria and archaea use the RNA-guided endonucleases of the various CRISPR–Cas systems to bind and cleave foreign nucleic acids as part of an adaptive immune system[13]. These systems retain a record of previously encountered pathogens by capturing nucleic acid sequences during past infections and then use these captured sequences ('spacer sequences') to direct CRISPR–Cas

[1]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [4]These authors contributed equally: Andrew V. Anzalone, Luke W. Koblan. ✉e-mail: drliu@fas.harvard.edu

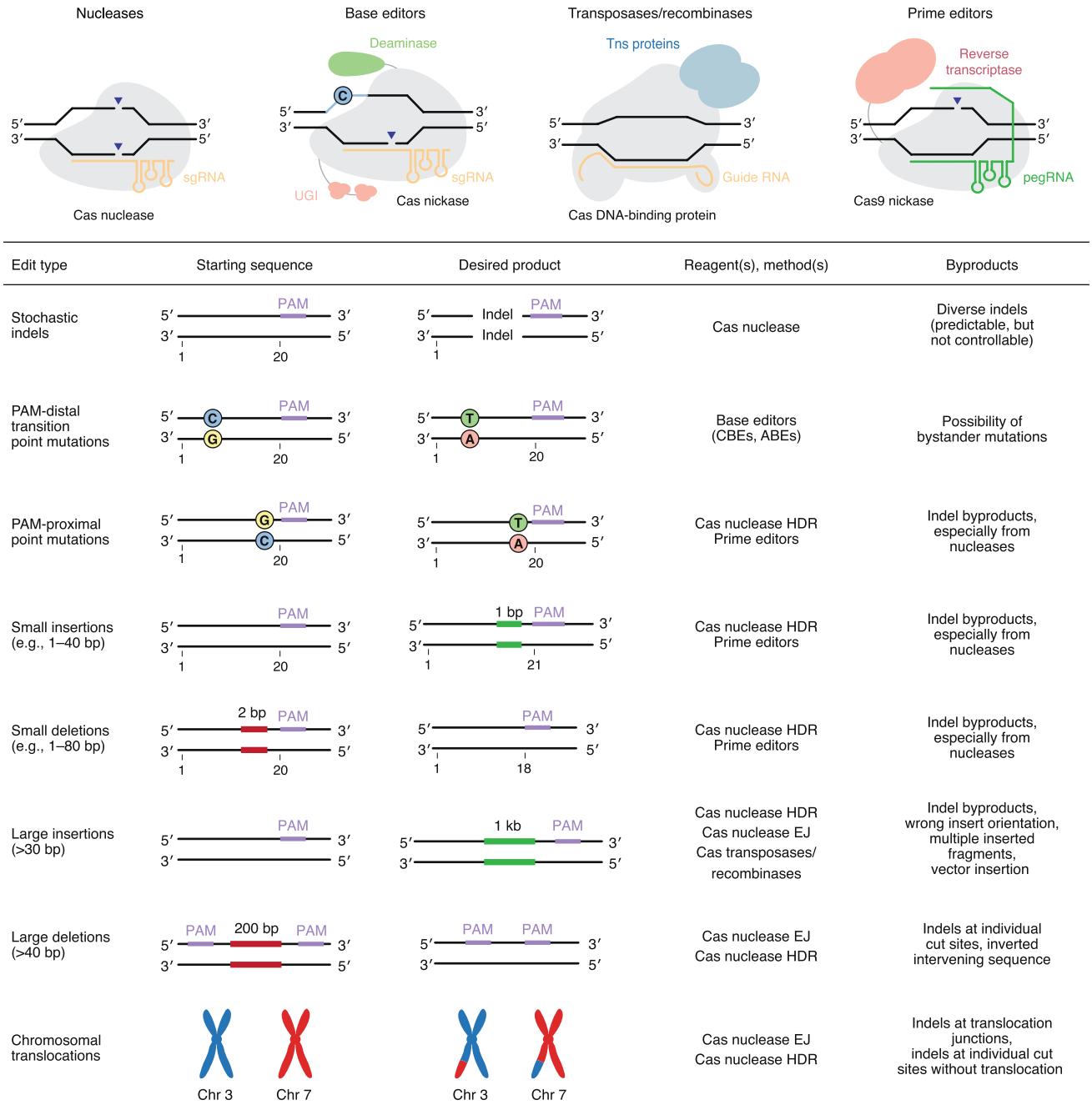



**Fig. 1 | Overview of genome editing strategies and agents.** Four classes of CRISPR-based genome editing agents are nucleases, base editors, transposases/recombinases and prime editors. Examples of starting substrates and edited products are shown, along with CRISPR–Cas agents that can be used to achieve the transformation and anticipated byproducts. Homology-directed repair (HDR); cytosine base editor (CBE); adenine base editor (ABE); end-joining (EJ); prime editor (PE); uracil glycosylase inhibitor (UGI); protospacer adjacent motif (PAM).

proteins to destroy the pathogens' DNA or RNA during future encounters. This mechanism enables CRISPR–Cas systems to be readily reprogrammed to target a wide range of DNA or RNA sequences simply by using different spacer sequences within a guide RNA molecule, provided the matching target DNA 'protospacer' sequence is positioned adjacent to a suitable PAM[1,2]. This PAM requirement protects the genomic DNA encoding these guide RNAs, which by definition must include targeted spacer sequences but which lacks adjacent PAM sequences, from being destroyed by CRISPR–Cas systems[13].

The programmability of CRISPR–Cas systems, along with the robustness of some Cas effectors across various cell types and organisms, has enabled their widespread use in the life sciences[11,12,14]. In this section, we overview naturally occurring Cas effectors, the ways in which Cas nucleases have been used for genome editing, and the development of engineered Cas nuclease variants that enable broader targeting scope and higher DNA cleavage specificity. Many of the improvements to CRISPR–Cas nucleases are also applicable to non-nuclease classes of CRISPR–Cas-based tools, described below.

Naturally occurring CRISPR–Cas immune systems have been classified into two main groups: class 1, which uses multiprotein complexes for nucleic acid cleavage; and class 2, which uses single-protein effector domains for cleavage[15,16]. Because of the advantages offered by single-protein effector domains, class 2 systems are the

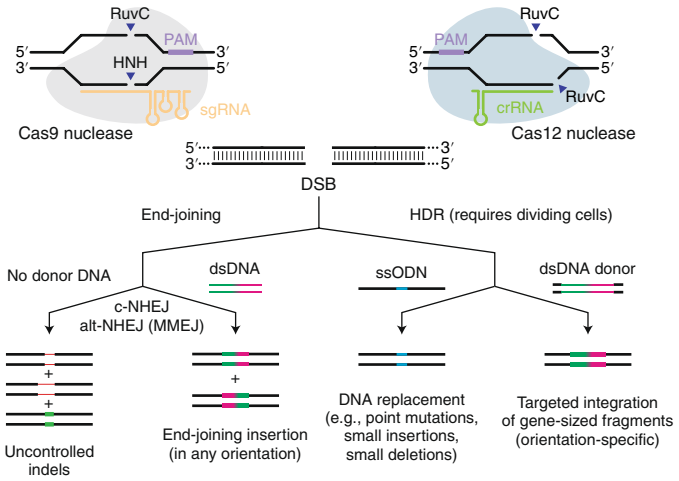



**Fig. 2 | Genome editing with Cas nucleases.** Cas9 nucleases are guided by guide RNAs to generate predominantly blunt-end DSBs using two distinct nuclease domains (RuvC and HNH). The DSBs occur within the protospacer, typically preceding three base pairs upstream of the PAM. Cas12 nucleases are guided by crRNAs to cleave both strands of DNA with a single RuvC-like nuclease domain. Cas12 cuts DNA in a staggered orientation within PAM-distal regions of the protospacer. Genome editing with CRISPR–Cas nucleases results from two major arms of DNA repair. End-joining mechanisms, such as classical nonhomologous end-joining (c-NHEJ) and microhomology-mediated end-joining (MMEJ, or alt-NHEJ), result in uncontrolled, but predictable, indels for gene disruption. In the presence of a donor DNA template, homology directed repair (HDR) is a competing (typically less efficient) pathway that occurs mostly in dividing cells and is used to install targeted mutations or to knock in larger DNA sequences. Insertion sizes with single-stranded oligonucleotide donors (ssODNs) are typically limited by synthesis capabilities[118].

most widely used CRISPR tools for biological research and translational applications. Class 2 is further subdivided into three types—II, V and VI—each of which uses a distinct type of Cas protein. Of the Cas proteins from class 2 systems, most type-II Cas9 variants and type-V Cas12 variants possess RNA-guided DNA endonuclease activity, while type-VI Cas13 variants appear to show preferential RNA-targeting and cleavage activity. As the scope of this Review is DNA editing, our discussion of naturally occurring Cas proteins is restricted mainly to Cas9 and Cas12 DNA endonucleases.

**Cas9 nucleases.** Cas9 effectors from type-II CRISPR systems are RNA-guided endonucleases that generate DSBs in target DNA sequences[17]. In their native context, Cas9 nucleases are guided by CRISPR RNAs (crRNAs), which pair with *trans*-activating crRNAs (tracrRNAs) that facilitate ribonucleoprotein complex formation[18]. However, most Cas9 genome editing applications use single guide RNAs (sgRNAs), which were engineered by fusing the crRNA and tracrRNA into a single RNA molecule[1]. The PAM sequence for all Cas9 effectors is located immediately 3′ of the protospacer on the DNA strand not complementary to the guide RNA. Cas9 nuclease predominantly makes a blunt-ended DSB 3 bp upstream of the PAM, within the protospacer sequence[19] (Fig. 2), although alternative cutting patterns have been observed for some Cas9 nucleases[20,21] and uncommon 1-bp and 2-bp staggered cuts have been inferred from DNA repair outcomes[22].

Target-site recognition begins with binding of the Cas9–guide RNA ribonucleoprotein complex to the cognate PAM sequence, followed by unwinding of the double-stranded DNA and concomitant formation of an RNA•DNA heteroduplex between the guide RNA spacer and the target DNA strand[17]. This process proceeds directionally, starting at the PAM-proximal region of the protospacer and traveling to the PAM-distal end of the protospacer[23]. As the non-target DNA strand is displaced by the guide RNA spacer, it forms a single-stranded DNA 'R-loop' that is exposed and accessible to other molecules[24,25]. The accessible and single-stranded nature of R-loops is exploited by newer genome editing technologies, including base editing and prime editing, discussed below[26–29].

After R-loop formation, Cas9 undergoes conformational changes that result in activation of its nuclease domains[25,30,31]. These conformational changes are impeded by mismatches between the target strand and guide RNA spacer, thus restricting nuclease activation to sequences that are highly complementary to the guide RNA spacer[32]. As described below, mutations in Cas9 that inhibit this conformational transition are critical to the enhanced DNA specificity of high-fidelity nucleases. Finally, after nuclease activation, the phosphodiester backbone of DNA is hydrolyzed by Cas9's two distinct nuclease domains: the HNH nuclease domain, which cleaves the guide RNA–bound target DNA strand; and the RuvC-like nuclease domain, which cleaves the PAM-containing non-target DNA strand[1]. Inactivation through mutation of either of the nuclease domains generates a Cas9 nickase (an enzyme that cleaves only one of the DNA strands), while inactivation of both nuclease domains generates a catalytically dead Cas9 (dCas9)[3,33] that can, however, still bind specific DNA sequences. Nickases are particularly useful for base editors and prime editors, which precisely edit DNA without requiring the formation of DSBs or homology-directed repair[26–29], whereas dCas9 can be exploited for various applications ranging from transcriptional regulation[8–12] to epigenetic modifications[9–12].

Since the initial reports of programmed DNA cleavage by Cas9 nuclease from *Streptococcus pyogenes* (SpCas9) in vitro[1] and in mammalian cells[3,4,6,7], many more Cas9 variants have been discovered and tested for genome editing, including orthologs from *Staphylococcus aureus*[34], *Streptococcus thermophilus*[3,35,36], *Neisseria meningitidis*[35,37,38], *Campylobacter jejuni*[39] and many other organisms[40–42]. These Cas9 effectors differ in their overall size, cognate PAM sequences, guide RNA architecture, optimal spacer length, editing efficiency and editing specificity. For example, SpCas9, currently the most widely used CRISPR–Cas nuclease, contains 1,368 amino acids, recognizes a relatively common NGG PAM, can be used with either an sgRNA or crRNA/tracrRNA pair, functions optimally with 20-nt spacers, has robust DNA targeting and cleavage activity, and supports relatively high levels of off-target editing[11,14,43]. Some Cas9 variants offer particular advantages over SpCas9, such as smaller sizes (for example, SaCas9 is 1,053 amino acids)[34] or pyrimidine-rich PAMs (for example, Nme2Cas9)[38]. Supplementary Table 1 lists many Cas9 orthologs and their key properties for genome editing applications.

**Cas12 nucleases.** In the current classification scheme[16], Cas12 nucleases encompass several variants from type-V CRISPR systems. These proteins possess just a single RuvC-like nuclease domain that mediates target DNA cleavage of both strands (Fig. 2). Many Cas12 nucleases are naturally guided by a single crRNA, though some natively use an additional tracrRNA similar to that of type-II CRISPR systems. Cas12 nucleases typically generate staggered cuts within regions of the protospacer that are distal to the PAM sequence[44]. These features contrast with the PAM-proximal blunt-end cuts generated by Cas9.

Cas12a (formerly named Cpf1) is the first Cas12 nuclease that has been widely used for genome editing applications[44]. Cas12a is able to process its single crRNAs from an array using a dedicated RNase domain, facilitating multiplexed gene editing[45]. Cas12a orthologs generally use T-rich PAMs, which are orthogonal to the PAMs of most Cas9 effectors. Some Cas12 variants, such as Cas12f (formerly Cas14), robustly cleave single-stranded DNA targets[46]. Still other Cas12 variants, including Cas12b (formerly c2c1) and

Cas12i, predominantly nick double-stranded DNA[47–49], though Cas12b has been engineered[48] to more efficiently cleave both DNA strands at 37 °C. Cas12e (formerly CasX), notable for its small size (<1,000 amino acids), is active in bacterial and human cells[50,51].

Several Cas12 variants, including Cas12a, indiscriminately cleave single-stranded DNA or RNA following target site recognition and nuclease activation[49,52]. This collateral cleavage activity has been used in nucleic acid detection applications[46,52–54]. Although most Cas12 variants target and cleave DNA, other Cas12 effectors, such as Cas12g, are RNA-guided RNA-cleaving enzymes that exhibit both collateral RNase and DNase activity upon activation[49]. Supplementary Table 1 lists several Cas12 variants that have been applied for genome editing applications.

**Repair pathways.** Prior studies in mammalian cells using other targeted nucleases, including homing endonucleases (also known as meganucleases)[55–57], zinc finger nucleases[58–61] and transcription activator-like effector nucleases (TALENs)[62,63], provided an intellectual foundation for the application of Cas nucleases for genome editing by establishing how a targeted DSB can initiate DNA sequence alteration.

Multiple DNA repair mechanisms within cells repair DSBs[64]. Several factors influence which repair pathway is used by a cell, including cell type, cell state and the nature of the DSB[65]. DSB repair pathways can be divided into two main arms: those that perform re-ligation (end-joining) of the broken DNA ends, often with additional nucleotide deletions or insertions at the DSB site; and those that use DNA templates for homology-directed repair (HDR)[66]. While end-joining processes are typically efficient in most mammalian cells[67], HDR is generally active in only dividing cells and requires proteins that are expressed predominantly in the S and G2 cell-cycle phases[68–70].

CRISPR–Cas nucleases are most commonly used to efficiently and selectively disrupt target gene sequences[3–7,71]. In most mammalian cells, nuclease-induced DSBs are most often repaired by error-prone end-joining processes (Fig. 2). Perfect end-joining repair regenerates the starting sequence, which remains a substrate for subsequent nuclease cleavage, but end-joining may also result in the insertion or deletion of nucleotides around the break site that prevent subsequent recognition and re-cutting by the nuclease[72]. The mixture of insertion and deletion (indel) products that result from DSBs cannot (thus far) be controlled, but they are not random, and can be predicted by inspection or using a machine-learning model[73–77], especially when the break site lies in a region of microhomology. In some cases, end-joining produces high yields of a single desired product[74,78].

If nucleases are targeted to open reading frames, indel products after end-joining usually generate frameshift mutations in coding sequences that abrogate protein function[79–82]. Nucleases can also be used to disrupt *cis*-regulatory elements within promoters or enhancers[83], as well as to characterize non-coding RNAs[84,85]. The delivery of multiple guide RNAs that target distinct genomic sequences will typically result in the multiplex generation of targeted DSBs[3,4,71]. Pairs of guide RNA sequences that target adjacent regions of a chromosome sequence will often result in the deletion of the intervening sequence[3,4,86]. Similarly large deletions have been generated using type-I Cas3 editing systems, which represent an exciting area for future development[87–89]. End-joining has also been used to insert DNA sequences at sites of Cas9-induced DSBs[90–92] (Fig. 2).

The simultaneous introduction of two DSBs in a cell can lead to additional rearrangements such as deletions, inversions and chromosomal translocations[93] (Fig. 1). These rearrangements are of particular interest to the biomedical research community since they occur frequently during oncogenesis, particularly in hematological malignancies[94]. The efficiency and purity of this process can be enhanced by the addition of homology donors that serve as templates to stitch together the desired chromosome fragments[95]. In addition to these designed chromosomal translocations[96–98], Cas9 nuclease has been reported to induce undesired chromosomal translocations and other complex rearrangements in cells[99].

A landmark study by Jasin and co-workers[100] demonstrated that DSBs stimulate HDR at a genomic locus in human cells. HDR in theory can introduce a variety of genome edits, including point mutations, precise insertions, precise deletions and integration of gene-sized DNA fragments[101] (Fig. 1). As with other nucleases, Cas nucleases can be used to initiate HDR in the presence of double-stranded DNA or single-stranded DNA (ssDNA) donor templates[3,4,7] (Fig. 2).

Nuclease-induced DSBs are repaired more frequently by end-joining mechanisms than by the desired HDR pathway[67]. Cas9 nickases, on the other hand, do not lead to high yields of indels at most loci and can also be used to stimulate HDR[102,103]. The efficiency of HDR can be increased by suppressing the activity of proteins that mediate nonhomologous end-joining or enhancing the activity of HDR pathways using gene silencing[66], small-molecule reagents[66,104–110] or expressed proteins[66,106,111–113]. Given the importance of DNA repair proteins, however, many of these strategies are anticipated to be difficult to implement in vivo.

In addition to the DNA repair pathways described, the generation of DSBs in cells can also lead to undesired genomic alterations such as translocations and large deletions[99,114], as well as activation of p53 responses in cells[115,116]. Various improvements to HDR, such as improved designs for DNA donor templates[102,117,118], colocalization of donor DNA templates to the sites of nuclease-induced DSBs[119–121], cell cycle synchronization[70], or the use of adeno-associated virus (AAV) genomes as donor DNA templates[122], can increase editing efficiencies. Even with these improvements, in most cases indels still represent the majority of edited products with Cas nuclease-initiated HDR, especially in non-mitotic cells[14,67].

**Engineered Cas variants with expanded targeting scope.** Expanding the targeting scope of genome editing agents has been a major focus of CRISPR–Cas technology development. The primary requirement for Cas protein binding is the presence of a PAM sequence[13]. A number of Cas9 and Cas12 orthologs have been discovered that recognize various PAM sequences[123], and this natural diversity has been harnessed to enable targeting of a greater fraction of genome sequences. PAM availability, however, remains a common limitation even with the use of natural Cas protein orthologs, as only a minor fraction of total PAM space is collectively accessed by all of the natural Cas proteins shown to function in mammalian cells. Because of these challenges, many researchers have engineered or evolved Cas9 or Cas12 variants with less restrictive PAM compatibilities (see Supplementary Table 1).

The earliest reported examples of Cas9 effectors with altered PAM specificity were developed using a combination of rational design and directed evolution approaches. In 2015, Joung and co-workers reported SpCas9-EQR, SpCas9-VQR and SpCas9-VRER, mutants of SpCas9 that recognize NGAG, NGA and NGCG PAM sequences, respectively[124]. These variants were generated by first installing an R1335Q mutation, which disrupts a critical interaction between wild-type SpCas9 and the third nucleotide of its canonical PAM sequence (NGG). The PAM-interacting domain of the SpCas9(R1335Q) variant was then mutagenized into a library of variants, which was then subjected to selection in bacteria for cleavage of target sequences containing NGA PAMs. Many enriched variants contained mutations at three particular amino acid positions (D1135V/Y/N/E, R1335Q, T1337R) that restored nuclease activity on sequences containing an altered PAM. In structural studies, these mutated amino acids were later found to interact directly with the new PAM sequence[125]. Shortly thereafter, a hybrid variant

of SpCas9-VRER and SpCas9-VQR, SpCas9-VRQR, was reported to have improved NGA-targeting activity in human cells[126].

We[127] have used directed evolution to generate Cas9 proteins with broadened PAM compatibility. Phage-assisted continuous evolution (PACE) and selection in bacteria that links phage replication to broader PAM compatibility generated variants such as xCas9-3.7, which displays higher activity on non-NGG PAM sequences (especially NGT and NGA PAMs) than that of wild-type SpCas9 and greatly reduced off-target activity in human cells[127,128].

Nureki and co-workers[129] used structure-guided rational design to develop SpCas9-NG, a Cas9 variant that can target all NG PAM sequences with varying activities, in many cases with higher efficiency than xCas9-3.7. SpCas9-NG incorporates several of the mutations used in SpCas9-VQR, while also adding mutations that stabilize the Cas9–PAM interaction. Recently, an extensive comparison of SpCas9, xCas9-3.7 and SpCas9-NG activity in human cells[128] revealed that nucleotides outside of the traditional trinucleotide PAM sequence, extending to up to 5 nucleotides adjacent to the protospacer, can influence targeting activity by each of these Cas9 variants. Moreover, this study showed that the engineered variants xCas9-3.7 and SpCas9-NG were less tolerant of extended guide RNAs containing a 5′ G, which is often appended to enable efficient transcription from the commonly used U6 promoter, as compared to a standard 20-nt spacer with a matched or mismatched 5′ G[128].

More recently, our group[130] and that of Kleinstiver[131] have further broadened the range of targetable PAM sequences. We[130] used phage-assisted continuous and non-continuous evolution to generate three new SpCas9 variants (SpCas9-NRRH, SpCas9-NRCH and SpCas9-NRTH, where H is A, C or T, and R is A or G) with higher activity than xCas9-3.7 that collectively can be used to target almost any NR PAM. Kleinstiver and co-workers[131] used structure-guided engineering to develop SpG and SpRY, two SpCas9-based variants that recognize NGN and NRN/NYN (where Y is C or T) PAMs, respectively. NRN is recognized more efficiently than NYN.

An engineered chimeric Cas9 protein, Spy-mac Cas9 (and an enhanced-activity variant called iSpy-mac Cas9), was reported to recognize NAA PAM sequences[132]. Spy-mac Cas9 was constructed by replacing the PAM-interaction domain of SpCas9 with that of SmacCas9, which was computationally predicted to recognize AA dinucleotide PAMs. A similar computational strategy was used to identify an ScCas9 ortholog[40] and to subsequently engineer high-activity ScCas9 variants that recognize NNG PAM sequences[133].

SaCas9, which offers advantages for certain genome editing applications due to its smaller size[34], natively has a fairly restrictive PAM (NNGRRT), which is present on average at 1 in every ~64 genome sequence positions. Through mutagenesis of the PAM-interacting domain and bacterial selections, an engineered SaCas9-KKH variant was developed[134] that recognizes NNNRRT PAM sequences, expanding the targeting scope by ~4-fold compared to wild-type SaCas9. Additionally, FnCas9, which has a high intrinsic specificity, has been engineered using rational approaches to create effectors that recognize YG PAMs in place of its original NGG PAM[41]. The engineered FnCas9-RHA variant was designed using structural data that suggested the possibility for disrupting an interaction between Arg1556 and the third canonical PAM nucleotide (NGG). While FnCas9 was found to have low genome editing activity in human cells, targeting of a dSpCas9 to an adjacent region (the 'proxy-CRISPR' approach) facilitates FnCas9-mediated DNA cleavage[20].

The T-rich PAM sequences recognized by Cas12 proteins complement the G-rich PAMs preferred by many Cas9 orthologs. However, canonical Cas12 PAMs are often somewhat restrictive, such as the TTTV PAM (where V indicates A, C or G) recognized by LbCas12a and AsCas12a[44]. Cas12a variants have been engineered using structure-guided mutagenesis to recognize either TATV or TYCV PAM sequences[135]. Recently, Joung and co-workers[136] reported enAsCas12a, a Cas12a variant with roughly twice the activity of wild-type AsCas12a on canonical TTTV PAMs. This Cas12a variant, developed using structure-guided mutagenesis, also recognizes an expanded set of PAMs, including TTYN, VTTV and TRTV.

The set of natural and engineered Cas protein variants reported to date can collectively recognize over half of the PAM sequence space, signifying a great deal of progress in the field from its beginnings with SpCas9. Nonetheless, improving the targeting scope of Cas effectors—in particular, to recognize non-purine-containing PAM sequences—presents an important challenge. Although substantial effort has focused on expanding the targeting scope of Cas9 variants, particularly SpCas9, the discovery and engineering of Cas12 variants that recognize T-rich PAM sequences will likely continue to expand the targetability of Cas effectors to help meet these needs. Finally, we note that the robustness of engineered Cas protein variants often does not match that of their naturally occurring Cas counterparts[128]. This drawback is most evident when Cas proteins are used for more demanding applications, such as editing animal models of human genetic diseases or targeting DNA loci that are occupied by endogenous proteins. Thus, the creation of robust Cas variants that maintain high activity while accessing certain challenging PAM sequences, such as those sparsely covered or missing in Supplementary Table 1, remains an important opportunity for future advancement.

**Engineered Cas variants with higher DNA specificity.** Precise targeting by Cas nucleases depends on their ability to discriminate on-target sequences from similar off-target sequences present throughout the genome[137]. While some Cas9 and Cas12 orthologs have been found to be naturally high-fidelity effectors[38,138–142], a major research focus has been the development of engineered Cas variants that have greater discrimination for on-target sequences over off-target sequences. The development of high-fidelity enzymes has been greatly facilitated by the development of methods to sensitively detect off-target Cas nuclease activity[43].

One strategy to improve targeting precision is to require two Cas proteins to bind adjacently to induce a DSB. This goal has been achieved using two distinct strategies. Paired Cas9 nickases, together with two sgRNAs that target opposing strands of a DNA target, can induce dual nicks only when both nickases bind nearby target sequences, decreasing the likelihood of off-target DSBs, which would require that two off-target sites be near each other in the genome[143]. When using this strategy, the spacing and orientation between the two nick sites is an important determinant of DSB formation efficiency. In a second strategy, fusions of dCas9 to the catalytic domain of Fok1 have been used to increase the specificity of DSB formation[144,145]. The Fok1 nuclease catalytic domain, which has been used extensively with other programmable nucleases such as zinc finger nucleases and TALENs, is only active upon homodimerization[146]. Thus, the simultaneous recruitment of two Fok1 catalytic domain monomers at adjacent DNA sites is required for efficient and specific DNA cleavage in human cells, making the off-target editing efficiency of these systems is very low[144,145]. Alternatively, fused Cas9–Cas9 chimeras have been used to increase DNA specificity[147]. In this design, one Cas9 domain is catalytically inactivated and serves as an auxiliary DNA binding domain, while a second Cas9 domain is catalytically competent but attenuated. This strategy biases nuclease activity to sites where both Cas9 domains can bind[147]. The specificity advantages of these dual nickase systems are balanced by their greater complexity and the requirement of additional components.

In an effort to increase the intrinsic specificity of Cas9 proteins, several high-fidelity variants have been engineered or evolved to have lower off-target editing activity[126,127,148–153]. First, Zhang and

co-workers[148] reported eSpCas9(1.1), an SpCas9 variant that was discovered through alanine scanning of positively charged residues that line the non-target-strand binding groove, with the hypothesis that interrupting interactions between these residues and the negatively charged nucleic acid backbone would decrease binding affinity. After screening mutants, the combination of K848A, K1003A and R1060A mutations was chosen, and the resulting eSpCas9(1.1) variant displayed efficient and precise genome editing in human cells. Joung and co-workers[126] used similar principles in their design of the high-fidelity Cas9 variant SpCas9-HF1, where they aimed to decrease the binding affinity between the Cas9–sgRNA complex and the DNA target by mutating Cas9 residues that make polar contacts with the phosphate backbone of the target DNA strand (N497A, R661A, Q695A, Q926A). Like eSpCas9(1.1), SpCas9-HF1 has high on-target activity and reduced off-target editing as characterized by GUIDE-seq[154]. An enhanced fidelity variant, termed HeFSpCas9, was later constructed[150] by combining mutations found in eSpCas9(1.1) and SpCas9-HF1.

Following the development of eSpCas9(1.1) and SpCas9-HF1, Doudna and co-workers[149] examined the biochemical mechanism of off-target discrimination by high-specificity Cas9 effectors. Their study revealed that, compared to wild-type Cas9, eSpCas9(1.1) and SpCas9-HF1 spend a larger fraction of time occupying the nuclease-inactive conformation when bound to off-target DNA sequences. This insight enabled the generation of HypaCas9, a high fidelity SpCas9 variant that exploits this conformational proofreading mechanism. HypaCas9 contains N692A, M694A, Q695A and H698A mutations and also demonstrates a high ratio of on-target to off-target cleavage activity[149].

Whereas eSpCas9(1.1), SpCas9-HF1 and HypaCas9 were generated using rational design approaches, Cereseto and co-workers[151] took an unbiased approach by developing a yeast-based assay to screen a library of SpCas9 effectors with mutations in the REC3 domain. Simultaneous positive and negative selection using well-established yeast auxotrophic markers allowed the direct selection of library members with both high on-target activity and low off-target activity. This resulted in the identification of evoCas9, a quadruple mutant with mutations not found in any previously reported high-fidelity Cas9 variants, which showed greater fidelity than other previously developed SpCas9 high-specificity variants[151].

Kim and co-workers[152] applied a conceptually similar positive and negative selection scheme in *Escherichia coli* to identify a more specific SpCas9 variant named Sniper–Cas9. Libraries of Cas9 variants were generated using a combination of error-prone PCR and DNA shuffling on full-length SpCas9. These libraries were then subjected to a selection in which cell survival was dependent on the Cas9 variant cutting a target sequence within a toxic plasmid encoding ccdB while also avoiding cleavage of a similar off-target sequence within the genome. The authors evaluated the fidelity of Sniper–Cas9 using Digenome-seq[155] and observed significantly lower off-target editing than with SpCas9. Moreover, Sniper–Cas9 showed wild-type-like levels of on-target activities with truncated sgRNAs or sgRNAs with 5′-G-extended mismatched spacers[152].

In general, Cas12a variants have been reported to generate few off-target editing events by comparison to many of the Cas9 orthologs[138,139]. Although an enhanced-activity AsCas12a variant (enAsCas12a) was found to produce higher levels of off-target editing than wild-type AsCas12a, a high-fidelity variant (enAsCas12a-HF1) restored the low off-target editing activity[136].

The delivery of CRISPR–Cas systems in the form of RNP complexes often results in much lower levels of off-target modification[86,139,156–158]. Recently, a single-point-mutation variant of SpCas9 (R691A), named HiFiCas9, has been shown to display even lower off-target editing when used with RNP delivery platforms[153]. Of note, the previously reported high fidelity variants eSpCas9(1.1), SpCas9-HF1 and HypaCas9 demonstrated lower on-target editing as RNPs in these systems.

Engineered sgRNAs have also been used to reduce off-target activity[159,160]. Truncated sgRNAs, referred to as tru-gRNAs[159], use spacer sequences with <20 nucleotides complementary to the protospacer target. Tru-gRNAs were shown to reduce off-target editing efficiency by up to three orders of magnitude while maintaining high levels of on-target editing. As an alternative sgRNA platform for increased editing specificity, hp-sgRNAs were designed to contain an extension on the 5′ end of the spacer that has the potential to form an RNA hairpin with the spacer sequence[160]. These hp-sgRNAs likely reduce the binding stability of Cas9 nuclease to its DNA target in a manner that more strongly impacts off-target nuclease activity than on-target activity.

Although substantial progress has been made identifying and engineering higher fidelity Cas variants that can reduce off-target editing to a substantial extent, high-throughput evaluation of these variants has shown that they generally support lower on-target activity as well[43,161]. In addition, most high-fidelity Cas9 variants do not tolerate changes to the canonical guide RNA, including 5′ G additions or mismatches[126,128,162]. These factors limit their usefulness in practice, especially when combined with other activity-reducing modifications such as those that alter PAM compatibility. Thus, identifying CRISPR–Cas genome editing systems that have reduced off-target editing activity while maintaining robust on-target editing efficiency and compatibility with commonly used guide RNA variants continues to present opportunities to advance the DNA specificity of genome-editing agents.

## Genome editing with base editors

Base editors precisely install targeted point mutations without requiring DSBs or donor DNA templates, and without reliance on HDR[14,26–28,163–166]. Current base editors contain a catalytically impaired CRISPR–Cas nuclease (that cannot make DSBs) fused to a single-stranded DNA deaminase enzyme and, in some cases, to proteins that manipulate DNA repair machinery. Two main classes of base editors have been developed to date: cytosine base editors (CBEs), which catalyze the conversion of C•G base pairs to T•A base pairs; and adenine base editors (ABEs), which catalyze A•T-to-G•C conversions[26,27]. CBEs and ABEs can efficiently mediate all four possible transition mutations (C→T, A→G, T→C, G→A), which represent approximately 30% of currently annotated human pathogenic variants[167]. Base editors have been applied in a variety of cell types and organisms, including animal models of human genetic diseases, to install or revert transition point mutations[28,164,165] using CBEs[168–206] and ABEs[180,193,205–217].

In both CBEs and ABEs, the catalytically impaired Cas nuclease domain localizes a ssDNA deaminase enzyme to a genomic target sequence of interest. Upon Cas binding, hybridization of the guide RNA spacer to the target DNA strand causes displacement of the PAM-containing genomic DNA strand to form a ssDNA R-loop[17,33]. Although PAM-proximal nucleotides (typically the 6 ± 2 nt adjacent to the PAM) are occluded by the Cas effector domain, PAM-distal nucleotides are exposed as ssDNA and are accessible to the deaminase domain of the base editors[17,26,163]. CBEs use cytidine deaminases to convert cytosines within the R-loop to uracils, which are read by polymerases as thymines[26,218]. Similarly, ABEs use laboratory-evolved TadA* deoxyadenosine deaminases to convert adenosines within the R-loop to inosines, which are read as guanines by polymerases[27]. Base editing of target nucleotides within the R-loop is dependent on productive interactions between the deaminase and substrate nucleotides, and those nucleotide positions within the R-loop that support efficient base editing outcomes define the base editing 'activity window'. For canonical CBEs and ABEs that use SpCas9, this activity window spans approximately protospacer positions 4–8 (where position 1 is the first nucleotide

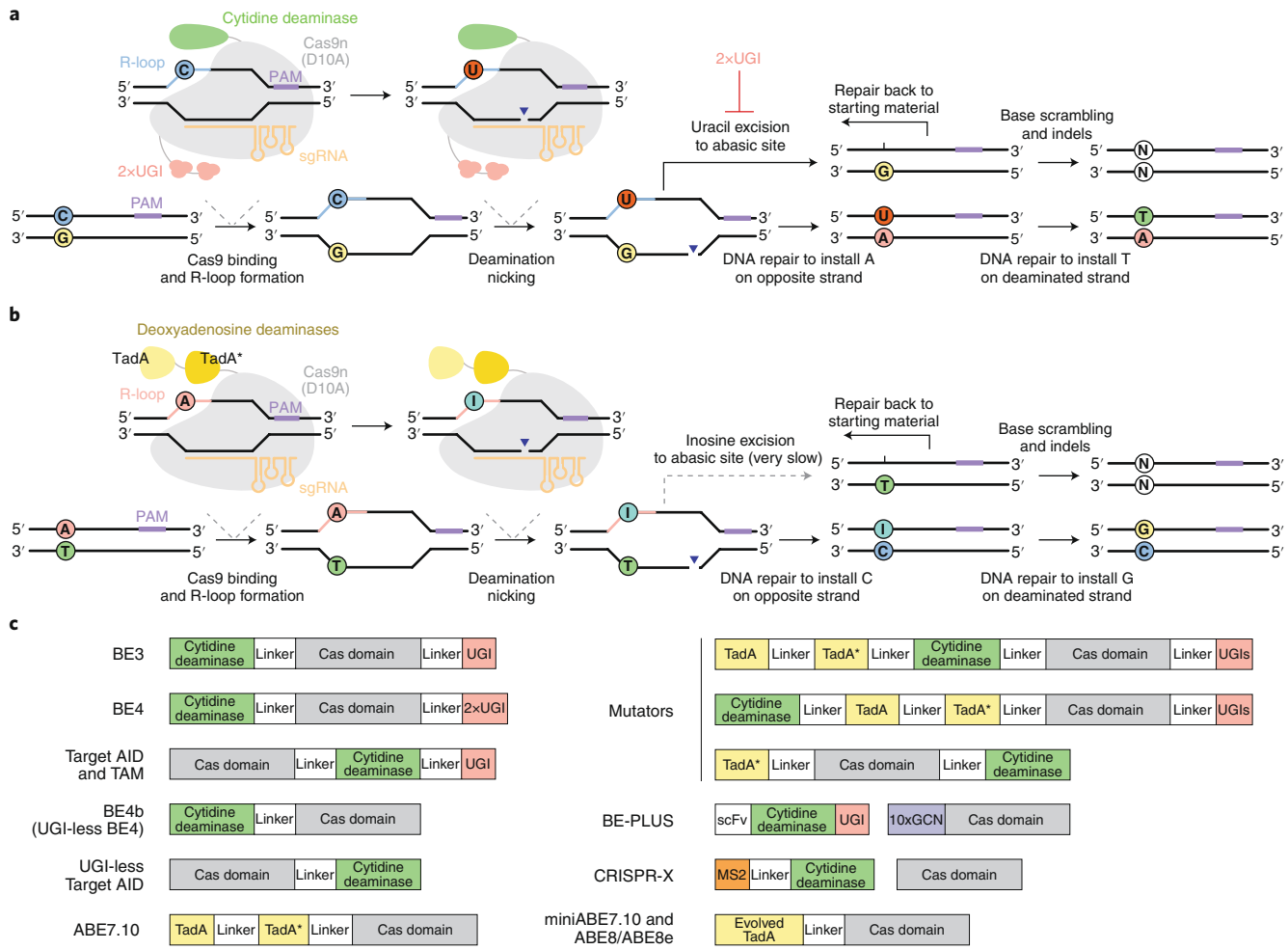



**Fig. 3 | Installing transition point mutations with CRISPR–Cas base editors.** Base editing with CBEs and ABEs is shown using Cas9 effectors as examples. **a**, CBEs install C•G-to-T•A point mutations using Cas9 nickases or dCas9 fused to cytidine deaminases and uracil glycosylase inhibitor domains (UGI). CBEs bind to a target DNA sequence and form a single-stranded R-loop. Cytosines within the R-loop are substrates for the cytidine deaminase domain, which catalyzes the hydrolytic deamination of cytosines to uracils. UGIs inhibit the activity of uracil glycosylases, which excise uracil bases in DNA to form abasic sites. Abasic sites promote base scrambling and indel formation. The Cas9 domain in most base editors nicks the non-deaminated strand, which directs DNA repair to install an adenine opposite the uracil as the nicked strand is remade. **b**, ABEs install A•T-to-G•C mutations using a fusion of Cas9 nickase (or dCas9) and evolved TadA* deoxyadenosine deaminase. In the example shown, one TadA domain is wild-type and the other TadA* domain is evolved to accept deoxyadenosine substrates. ABE8 variants use a single evolved TadA* monomer. ABEs deaminate target deoxyadenosines to deoxyinosines and also nick the non-deaminated strand, directing DNA repair to incorporate a cytosine (cytidine nucleoside) across from hypoxanthine (inosine nucleoside) on the opposite strand. As with uracil, hypoxanthine is also a substrate for excision by cellular glycosylases, which can generate abasic sites that lead to base scrambling or indels.

of the protospacer and the PAM is at positions 21–23), but can be influenced by differences in DNA state, such as chromatin architecture, that may vary by locus or cell type[26,27].

Deamination of substrate nucleotides within the editing window initially generates uridine and inosine, creating a mismatched DNA base pair with the G or T, respectively, on the opposite, non-deaminated strand (Fig. 3). Stable base editing outcomes require replacement of the unedited strand to install the corresponding A and C complementary nucleotides opposite the uridine or inosine, respectively. However, uracil and inosine intermediates are mutagenic, and DNA repair pathways have evolved in most organisms to remove these bases from genomic DNA[219]. Uracil is rapidly excised from genomic DNA by uracil DNA N-glycosylase (UNG2)[220–223]. If uracil excision occurs before installation of the complementary G-to-A conversion on the non-deaminated strand, the resulting abasic site will often revert to the original sequence (or an undesired transversion mutation) through the base excision repair pathway (Fig. 3a). To increase the half-life of uracil at the target locus and consequently increase editing efficiency and purity, CBEs typically include uracil glycosylase inhibitor proteins (UGIs) that substantially increase editing yield and product purity[26,163,218,224]. For ABEs, inhibition of MPG, the glycosylase thought to excise inosine from genomic DNA in eukaryotic cells, did not further increase already very high editing product purities, suggesting that inosine excision is much less efficient in mammalian cells than uracil excision[27,225].

To further improve base editing efficiency, most base editors, such as BE3 (the first described CBE) and ABE7.10, use a Cas nickase, rather than a dCas DNA-binding protein, to nick the non-deaminated DNA strand (Fig. 3). The resulting nick stimulates cellular repair of the non-deaminated strand, using the deaminated strand as a template for resynthesizing the nicked strand. Deamination of one strand and resynthesis of the complementary strand therefore results in editing of both target DNA strands to

yield stable conversion of the target base pair[26,218]. Additional improvements such as linker optimization and the fusion of a second UGI domain further improved CBE activity, resulting in the BE4 architecture (Fig. 3c)[218]. Nuclear localization sequence and codon optimization yielded the BE4max and ABEmax variants, which offer substantially increased editing efficiencies in mammalian cells and in vivo[169,226]. Subsequent improvements, summarized below, have dramatically expanded the genome targeting scope, efficiency and product purity of base editors.

**Maximizing base-editor targeting scope.** Base editors use RNA programmable CRISPR–Cas DNA binding proteins to enable the targeted deamination of single nucleotides that fall within a particular window in the target protospacer. As such, PAM availability is an important determinant of whether a target sequence can be modified with a base editor. The first CBE and ABE variants were developed with SpCas9, which has an NGG PAM requirement. NGG PAM CBEs could in theory correct approximately 26% of annotated pathogenic T•A-to-C•G mutations in ClinVar, while NGG ABEs could in theory correct 28% of annotated pathogenic G•C-to-A•T point mutations[26,27,127,167].

Early efforts to expand the targeting scope of base editors used Cas orthologs or engineered Cas variants that were developed to recognize alternative PAMs (summarized above; Supplementary Table 2). In addition to Cas9 effector variants, CBEs have also been constructed using various Cas12a homologs (Supplementary Table 2). By contrast, ABE7.10 has proven to be less tolerant of Cas domains beyond wild-type SpCas9 than CBEs, although several ABE7.10 Cas variants have been reported with at least modest activity in mammalian cells (Supplementary Table 2). Recently, the incompatibility of ABEs with some Cas domains (most notably, Cas12 domains) was overcome through the evolution of ABE8 deaminases, which offer up to 590-fold higher activity than ABE7.10 and enable efficient A•T-to-G•C editing with all Cas9 and Cas12 domains tested, including LbCas12a and enAsCas12a[227,228]. The editor architectures and editing windows for many of these constructs are shown in Fig. 3c and Fig. 4[28]. Notably, ABE8 variants support high-efficiency editing with a smaller architecture that uses a single evolved TadA* monomer[229], removing ~500 bp of coding sequence from the editor. We anticipate that further evolution and computational analysis may help describe the contributions each mutation makes to the overall activity of these enzymes[230], enabling future engineering efforts. Collectively, these Cas variants expand the genome targeting breadth of CBEs and ABEs to ~95% of pathogenic transition mutations in ClinVar[167].

In addition to increasing genome targeting scope, the continued development of base editors with different Cas domains is also enabling in other ways. Genes encoding smaller Cas proteins such as SaCas9[34] and newer Cas12 variants[46,231], for example, are more readily deliverable to target tissues using size-constrained delivery modalities such as AAVs. Recent work has reported dual AAV split base editor systems that enable efficient in vivo base editing[170,180,207]. Further optimization of Cas domains and deaminases may enable single-AAV editing approaches. Finally, because the on-target efficiencies of all genome editing tools vary by site (including exact protospacer and PAM position), cell type and cell state, a diverse toolbox that provides multiple base editor options for a given target of interest will facilitate many applications, including functional screens[191,192,204,232–234].

**Modifying R-loop–deaminase interactions.** The R-loop that is formed after the Cas domain binds its target provides a single-stranded DNA substrate for the deaminase domain of a base editor. Base editing efficiency within this R-loop is determined by productive interactions between R-loop substrate nucleotides and the deaminase enzyme[235]. The efficiency with which target bases are edited typically peaks around the most accessible nucleotides within the R-loop[26,27,235]. The base editing activity window is defined as those protospacer positions that support a certain fraction (typically 50%) or higher of average peak editing efficiency (Fig. 4). The location of the editing window can change when a different Cas domain is used or when the deaminase domain changes[227,236,237]. Broader editing windows increase the frequency of bystander editing—editing

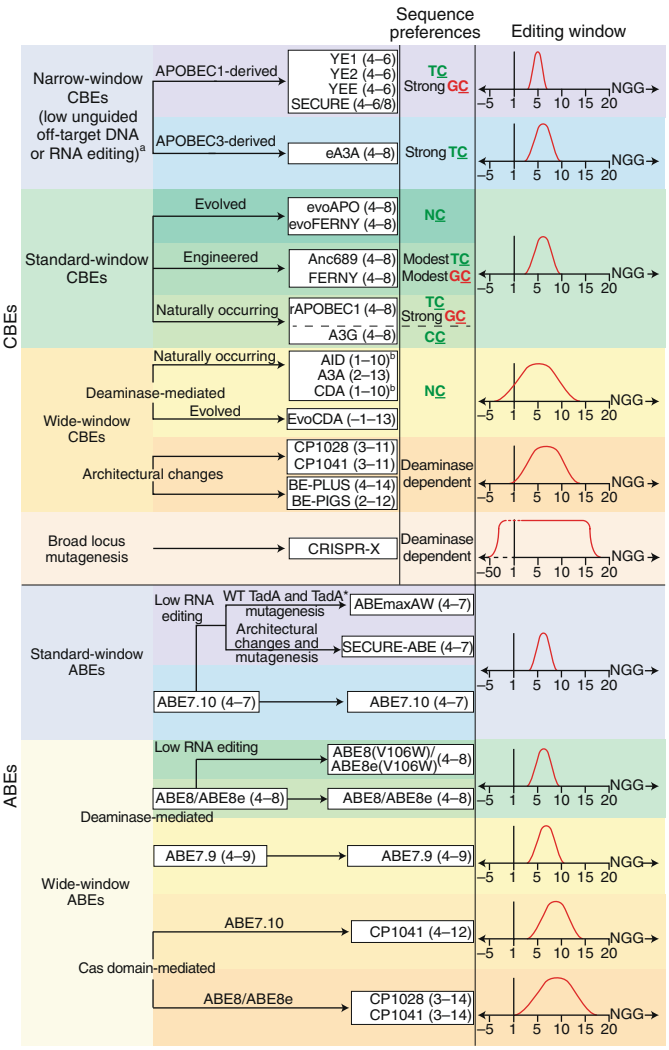



**Fig. 4 | CRISPR–Cas base editors and variants thereof.** CBEs and ABEs classified by editing characteristics and origin. SpCas9 CBEs and ABEs are shown to illustrate the array of editors available; in most cases, other Cas domains can also be used. CBEs are available as narrow-window editors, standard-window editors, wide-window editors and broad locus (random mutagenic) editors. ABEs are available as standard-window editors and wide-window editors. White boxes contain editor names followed by the approximate editing window size in parentheses. Window numbering is based on the 5′ position of the protospacer numbered 1. Deaminase sequence preferences, if known, are listed. Favored sequence contexts are shown in green and disfavored contexts in red. Editing window sizes are shown as representative diagrams but vary based on sequence. References for editing windows are provided in Supplementary Tables 2 and 3. [a]CBE variants with decreased DNA and RNA off-target editing have recently been developed with PpAPOBEC1, AmAPOBEC1, RrA3F and SsAPOBEC3B. Engineered mutants of these variants with decreased off-target DNA and RNA editing include PpAPOBEC1 H122A, PpAPOBEC1 R33A and SsAPOBEC3B R54Q. The editing window boundaries of these new CBEs are not yet known. [b]Low RNA editing.

within the protospacer at a position other than the target position—of nearby C or A nucleotides, which in some cases may lead to undesired effects.

Naturally occurring and engineered Cas variants offer different editing windows. For example, SaCas9 typically supports a broader editing window (typically protospacer positions ~3–12 for CBEs and ~4–12 for ABEs)[236] than SpCas9 (positions ~4–8 for CBEs and ~4–7 for ABEs), while Cas12a-derived CBEs and ABEs generally offer an editing window of 8–13 (where position 1 is the first nucleotide after the 5′ PAM)[170,238]. Researchers have also engineered SpCas9 domains that modify the location of the base editing window. Modifying the Cas–deaminase linker, for example, can restrict editing to particular locations within the editing window of some CBEs[239,240]. Circular permutation (CP) of SpCas9[241] changes the position of the deaminase relative to the R-loop, enabling more efficient conversion of target nucleotides positioned at the edge of the editing window[237]. Embedding the deaminase domain within a loop of the Cas9 PAM-interacting domain, which predictably changes the spatial relationship between the deaminase and the R-loop, has also led to editors with broadened editing windows[242]. Recruiting multiple deaminase copies to the R-loop to increase the exposure of available nucleotides to high concentrations of deaminase enzymes can also broaden the size of the editing window[243–245].

Varying the deaminase domain of base editors has also provided base editors with different editing window features. CBEs were first developed[26] with APOBEC1, a deaminase with a standard editing window (protospacer positions 4–8), and were subsequently reported with different APOBEC1 family members[246], as well as other deaminases such as CDA[129,163,171,179,181,189,194,195,218,235,239,247–253], AID[204,206,218,244,248,252–257] and APOBEC3 family members (A3A, A3B, A3C, A3D, A3F, A3G, A3H, A3I)[185,235,246,253,258–265]. Deaminases vary in their kinetic parameters and nucleotide substrate preferences, giving rise to different editing window widths and activities on different substrate sequences. For example, rat APOBEC1 prefers to deaminate cytosines within TC motifs, and it strongly disfavors deamination of cytosines within a GC context. In contrast, A3G prefers editing CC motifs[26,260]. Base editors with CDA, AID and A3A deaminases typically have wider windows, likely due to the higher activities of these deaminases[163,253,258,265]. Supplementary Table 3 summarizes the deaminase domains and Cas domains used to construct base editors to date.

Engineered deaminase domains have also been used in base editors with specific editing characteristics. For example, editors with narrowed windows have been developed by structure-based mutagenesis of APOBEC1 to impair its catalytic activity[236,266]. Ancestral reconstruction of APOBEC family members uncovered variants with unique sequence preferences and led to increased editor expression levels[226]. Engineering of APOBEC3A led to eA3A (APOBEC3A N57G), a deaminase with strong TC motif preference[258]. Truncated A3B variants that remove the N-terminal RNA-binding domain show higher editing activity than wild-type A3B deaminases, while truncated forms of APOBEC1, A3A and CDA have also been shown to support efficient base editing activity[240,261,267]. Other engineered AID, A3A and A3B deaminase variants have also been observed to support base editing activity, particularly for targeted random mutagenesis[206,244,252,254,257].

PACE of CBEs has also yielded base editors with increased editing activity, broader sequence context compatibility surrounding the target cytosine (for example, evoAPOBEC1-BE4), and wider editing windows (for example, evoCDA-BE4)[235]. Other evolution campaigns have generated deaminases derived from ancestrally reconstructed APOBEC lineages that are ~30% smaller than other APOBEC1-based deaminases and show little sequence context preference (for example, evoFERNY-BE4)[235]. The development of diverse CBEs with natural, engineered, and evolved deaminases has thus expanded the targeting scope, precision and effectiveness of CBEs.

Combining the rapidly expanding suite of Cas effectors with the growing set of deaminase domains has substantially advanced the capabilities and utility of base editors. The rapid increase in the number of known CBE- and ABE-compatible Cas domains with distinct PAM preferences, truncation or extension of the sgRNA to alter editing outcomes[152,207], as well as the increase in deaminase domains with different target sequence compatibilities and editing window positions, greatly increases the probability that a base editor exists that can convert a given target nucleotide of interest. At the theoretical extreme, when sufficient base editors exist such that every genome position lies within at least one base editor's activity window, ideal base editors would edit only a single nucleotide position within a single sequence context to minimize the possibility of bystander or off-target editing. The ever-increasing collective targeting scope of base editors therefore creates opportunities to develop more sequence-specific and narrow-window editors that minimize bystander or other unwanted base editing events.

**Minimizing undesired base editing.** Undesired byproducts of base editing can be classified into those that occur at the target site and those that occur at off-targets. Three types of undesired editing byproducts occur at on-target sites: transversion mutations at the target nucleotide (that is, C•G-to-[A•T or G•C] or A•T-to-[C•G or T•A]), bystander editing and indels.

Transversions occur when target nucleotides are converted to unintended nucleotides, and are observed as rare byproducts of CBEs that vary in frequency by target locus. Transversions likely arise from error-prone repair of abasic sites that are generated when cellular DNA glycosylases hydrolyze the glycosidic bond linking deaminated nucleobases such as uracil to the deoxyribose backbone[218]. Architectural changes from BE3 to BE4 decrease the frequency of transversions for CBEs by providing a second UGI copy, while increasing average editing efficiency[218]. Overexpression of UGI *in trans* or by P2A linkage also demonstrated improved editing purity[224]. Transversions are highly site-dependent for CBEs. ABEs do not display significant base randomization[27], presumably due to the much slower rate of inosine excision compared to uracil excision.

All base editors can generate bystander edits when multiple C or A nucleotides are present within the editing window. Because of degeneracy in the genetic code, when editing protein-coding genes with canonical CBEs or ABEs, ~53% of the time, bystander edits will not occur or will result in only silent mutations[28]. Some non-silent bystander base edits result in conservative mutations, such as Ile→Val, that may be inconsequential. In other cases, however, bystander editing is problematic. Researchers have reduced or avoided bystander editing by engineering deaminases with narrowed editing windows (YE1, YE2, YEE, EE and YFE)[236,266] or with high sequence context dependence (eA3A, which preferentially edits TC motifs)[229]. CBEs with highly active deaminases and broad editing windows, such as those containing CDA, AID, A3A or evoCDA deaminases[28,235] or that use circularly permuted Cas9 domains[237], intrinsically have a higher propensity to produce bystander edits. If a target site lacks an ideally positioned PAM, however, or if base editors are sought for large-scale mutation and screening, then wide-window editors may be required to achieve acceptable levels of on-target editing[235,243,244].

Sequence context-specific and narrowed-window ABEs have yet to be developed, but would be quite useful given the broad Cas9 and Cas12a PAM variant compatibility of high-activity ABEs such as ABE8[228] or ABE8e[227].

Indel byproducts at target loci can arise from the use of CBEs and, to a lesser extent, ABEs. Because base editors use nickase or dead Cas effectors, they do not directly generate DSBs and therefore generate far fewer indel products than Cas nucleases. Nonetheless, the excision of deaminated bases can initiate base excision repair,

which transiently breaks the DNA backbone of the deaminated strand[219,268]. If base excision repair occurs while the non-deaminated strand is nicked by a nicking-competent base editor or by endogenous DNA repair processes, then a DSB can result. Fusion of Mu Gam, a protein that binds to the ends of DSBs and protects them from degradation, can decrease indel formation in CBEs[218]. ABEs typically induce lower indel frequencies than CBEs[27], consistent with inefficient base excision of inosine and/or the slower kinetics of the ABE7.10 deaminase. Non-nicking dCas9 variants can be used to markedly reduce indel frequencies as well[26,228], but typically have lower base editing activity.

In addition to byproducts generated at the target site, undesired deamination at off-target DNA or RNA sites can also occur following the use of base editors in cells. Off-target DNA base editing can occur in a Cas-dependent (guided) or a Cas-independent (unguided) manner. Cas-dependent off-target base editing is caused by the Cas domain and guide RNA binding to sequences that are similar to the on-target locus[137,269,270]. Assays for detecting the genome-wide Cas-dependent off-targets of base editors have detected Cas9-dependent off-target base editing for both CBEs and ABEs, which primarily occur at a subset of Cas9 nuclease off-target sites[271–273]. Productive base editing requires additional criteria that affect the off-target profile, such as the presence of a cytosine or adenine within the editing window, nucleotide sequence context, and R-loop accessibility by deaminases, that are not satisfied for all Cas nuclease-dependent off-target sites. Cas9-mediated R-loop formation drives off-target editing events, making the use of highly sensitive Cas nuclease off-target detection methods such as CIRCLE-seq[274,275] an effective assay for identifying Cas-dependent off-target base editing candidate loci. The transient formation of R-loops by Cas9 melting of off-target DNA may also lead to off-target base editing at sites that are not bona fide nuclease off-target loci, although unbiased base editing off-target studies suggest such sites are rare[271,272].

The use of high-fidelity Cas domain variants can greatly reduce the frequency of Cas-dependent off-targets events[127,158]. While only a subset of the many high-fidelity Cas domains have been tested as base editors, increasing the DNA specificity of the Cas domain indeed minimizes Cas-dependent off-target DNA base editing of both CBEs and ABEs (Supplementary Table 2)[127,152,158,184,247,251,258,273,276]. In addition to using Cas domains with enhanced DNA specificity, sgRNA truncations have been reported to limit off-target base editing[152]. Limiting exposure of cells to base editors by delivering base editor RNPs, instead of DNA constructs, also greatly decreases off-target base editing while maintaining comparable on-target editing, likely due to more rapid action of base editors at on-target loci than off-target loci[158]. Finally, catalytically impaired or sequence-specific deaminases (for example, those in YE-BE4 and similar variants)[236,258,266] have also been shown to improve DNA specificity by decreasing off-target base editing more than on-target editing[236,277].

Base editors can also induce Cas-independent DNA off-target editing, which occurs from the long-term expression of deaminases that can randomly deaminate transiently accessible nucleotides at a low level across the genome. This phenomenon has been observed for CBEs in both mammalian and plant cells, but was not detected for ABE7.10[278–280]. Detecting Cas-independent off-target base editing is especially challenging because their low frequency ($\sim 5 \times 10^{-8}$ per base pair in mouse embryos injected with high levels of CBE mRNA[278]) is below most reported rates of spontaneous somatic cell mutation, and their random locations preclude the use of targeted amplicon sequencing to amplify off-target editing signals.

Recently, Liu and co-workers[277] and Gaudelli and co-workers[246] independently generated a series of bacterial and mammalian cell assays that do not require whole genome sequencing and sensitively detect the propensity of different CBEs to induce Cas9-independent off-target editing[246,277]. These efforts led to a broad characterization of the Cas-independent off-target activity for many reported cytosine base editor variants. Our group[277] developed YE1-BE4 and newly engineered YE1-BE4 variants with greater targeting scope were found to offer on-target editing efficiencies similar to those of BE4max, but with minimal Cas-independent off-target DNA editing activity. Gaudelli and co-workers[246] tested 153 deaminase domains and identified four APOBEC deaminase variants (PpABOBEC1, RrA3F, AmAPOBEC1 and SsAPOBEC3b) with decreased Cas-independent off-target DNA editing activity[246]. Additional engineering of these four deaminases led to the development of 'hifi' versions of these deaminases that further decreased the Cas9-independent off-target DNA editing (PpAPOBEC1 H122A, PpAPOBEC1 R33A, RrA3F F130L and SsAPOBEC3B R54Q). The relationship between on-target editing efficiency and off-target effects must be weighed for each editing application. Highly sequence-specific or narrow-window deaminases can be combined with the ever-growing set of Cas domains to create editors with targeting breadth, efficiency and DNA specificity properties that are well-suited for most base editing applications[246,277].

Finally, the deaminase domains of CBEs and ABEs, which are derived from enzymes that natively deaminate RNA, can also induce Cas-independent off-target RNA base editing when overexpressed at high levels in mammalian cells[229,263,281,282]. Like Cas-independent off-target DNA editing, off-target RNA base editing occurs at low levels but in a random and widespread manner. For example, overexpression of ABE7.10 induces ~0.22–0.24% A-to-I deamination across the transcriptome, compared with ~0.14–0.19% A-to-I deamination from endogenous cellular adenosine deaminases[227,282]. Although the overall transcriptome-wide frequency of C-to-U deamination with and without CBE overexpression has not yet been reported, Joung and co-workers[281] observed low-level but widespread cytosine deamination among cellular RNAs in cells highly enriched for maximal CBE overexpression. RNA editing from base editors is transient, given the rapid rate of RNA turnover, and is only expected to occur while base editors are at high levels (for example, ~48 h following transient transfection of overexpression plasmids).

Rational engineering of deaminase domains has generated CBEs and ABEs with minimal RNA editing activity. For both CBEs and ABEs there are now several engineered variants with decreased Cas9-independent RNA editing activity[229,246,263,281,282]. Some of these improvements—namely, the V106W mutant[282]—have been combined with the recently evolved ABE8 variants to generate editors with high on-target activity and minimal Cas9-independent DNA and RNA off-target activity[227,228]. Successes minimizing Cas-independent off-target DNA and RNA editing reflect the feasibility of modifying deaminases to be more dependent on Cas domain-mediated DNA binding—for example, by elevating the Michaelis constant ($K_m$) of the deaminase for DNA or RNA to minimize nucleic acid binding without Cas domain assistance.

**Considerations for the use of base editors.** Researchers have now generated a large suite of base editors, each with different characteristics. Although the diversity of base editors increases the likelihood of their successful application, the many available choices can make selecting the most promising base editor for a given application challenging. In this section, we discuss key considerations, including PAM availability, nucleotide context, editing window, product purity and DNA specificity, each of which guides the appropriate selection of a base editor for a given target sequence and application. Figure 5 shows a strategic flow chart summarizing these considerations. Recent efforts to develop machine learning models of genome editing outcomes[73,74,283–285] are further facilitating and automating this process.

If the genome editing application calls for a C•G-to-T•A transition point mutation, CBEs will be the appropriate editors, while for A•T-to-G•C transition point mutations, ABEs will be required. For random mutagenesis applications, base editors

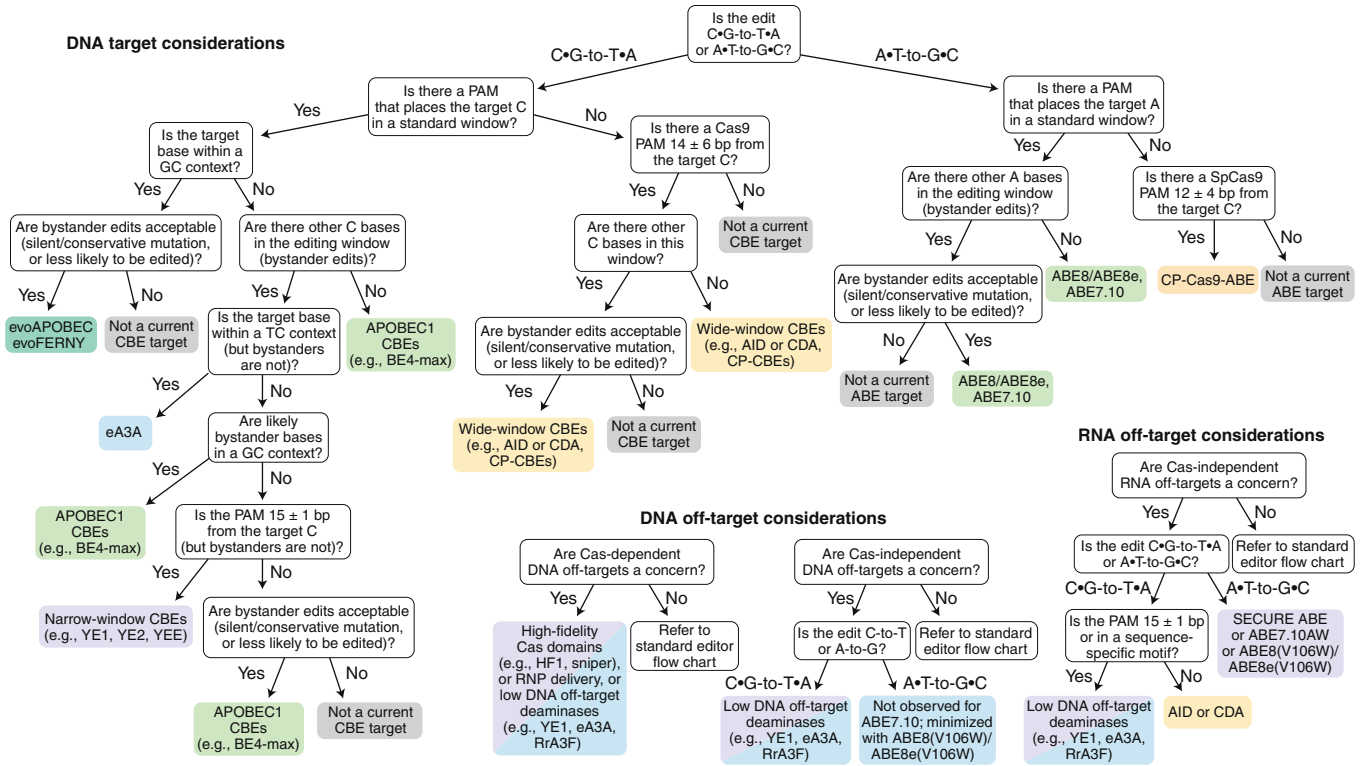



**Fig. 5 | Decision trees for choosing base editors.** Base editors can be chosen on the basis of several criteria, including the desired edit (C•G-to-T•A or A•T-to-G•C), the availability of suitable PAMs within the target sequence, sequence motifs surrounding the target nucleotide, the position of the target nucleotide within the protospacer, the possibility of undesired bystander mutations and the need to minimize off-target DNA or RNA editing. The flow chart can guide identification of candidate base editors. Depending on the application, some criteria will take priority over others, and trade-offs between efficiency and specificity may be necessary. The standard editing window for SpCas9 CBEs is typically protospacer positions 4–8, for SaCas9 CBEs is positions 3–12 and for Cas12a CBEs is 8–13. The standard editing window for SpCas9 ABEs is typically protospacer positions 4–7, for SaCas9 ABEs is positions 6–12 and for Cas12a ABEs is 8–13.

containing both cytidine deaminases and deoxyadenosine deaminases domains[286–289], CBE variants with no UGI domains (for targeted base scrambling)[163,218,244,250,254,257,265], and wide-window base editor variants[243,244] are often useful.

After selecting the appropriate class of editor for an application of interest, the next consideration is the identification of a suitable PAM. For C•G-to-T•A transitions, a Cas9- or Cas12-based CBE can be used if a suitable PAM exists 3′ (for Cas9-containing editors) or 5′ (for Cas12-containing editors) of the target C nucleotide on the same strand. A•T-to-G•C transitions can similarly be achieved with Cas9- or Cas12-based ABEs. Optimal editing occurs when the target base falls in the center of the editing window, typically protospacer positions ~4–8 for SpCas9-based CBEs and ABEs and positions ~8–13 for Cas12a-based editors. SaCas9 variants typically show a broader editing window (positions ~4–12) for both CBEs and ABEs (Fig. 5 and Supplementary Table 2).

Next, the sequence context of the target nucleotide should be analyzed. For CBEs, some of the component deaminases have strong sequence preferences (for example, the APOBEC1 family does not efficiently edit some cytosines in GC contexts). When needed, CBEs harboring deaminases with broader sequence context compatibility may be used, such as CBEs with evoAPOBEC1, evoFERNY, AID, CDA or A3A deaminases (Fig. 5). It is possible that additional C nucleotides, beyond the target C, will also be present in the editing window, especially when using broad-window editors. If the resulting bystander edits are predicted to occur with high efficiency, and if they are unacceptable for the application (for example, if they introduce non-silent or non-conservative substitutions in a gene correction effort), then other PAM sequences, narrow-window editors or sequence-constrained deaminase editors should be considered (Fig. 5).

If the target base is not in a GC context and there is low concern for bystander editing, then standard window editors such as APOBEC1, Anc689 or FERNY can be used. If the potential bystander edit is in a GC context, non-evolved APOBEC1 editors are excellent candidates because GC bystander nucleotides are less likely to be deaminated by APOBEC1 deaminases[226,235]. If bystander edits are possible and the target C falls within a TC motif while other non-target bases do not, the eA3A CBE variant can be used[258]. APOBEC3G (D316R, D317R) can be used to efficiently target cytosines in CC motifs[260] and eAID N51G prefers to deaminate GC motifs in a standard AID editing window[255]. If specificity cannot be attained on the basis of sequence context, narrow-window editors such as YE1, YE2, YEE, EE or YFE can be used if the target base exists in protospacer positions 4–6. Notably, these deaminases share sequence context preferences with APOBEC1[236,266]. Changes that rigidify the Cas–deaminase linker have also led to window-narrowing effects[239] (Fig. 5).

If PAM availability does not allow placement of the target C within a standard editing window, then expanded-window editors should be considered (Fig. 5), although using CBEs with wider editing windows increases the likelihood of bystander mutations. While deaminases with broader editing windows have been used to expand the editing scope of base editing, architectural changes to the SpCas9 domain have also broadened the editing window of base editors. When PAM availability permits, using broad-window deaminases with other Cas domains that show different standard editing windows may also expand the editing window. Expanded-window

editors should only be used when the target nucleotide does not exist within a standard base editing window and when bystander edits do not exist or are acceptable for the application. Recruiting multiple deaminase domains to the target locus can also increase the width of the editing window[243–245].

If the desired edit is an A•T-to-G•C transition, similar considerations apply for ABE use. If a candidate PAM has been identified that places the target A base in a standard ABE editing window (protospacer positions ~4–7), the potential for bystander edits should be considered. If no problematic bystander editing is possible, then ABE7.10[27], ABE8.20-m[228] or ABE8e[227] can be used. ABE8 variants show substantially higher activity that supports a much broader set of Cas domains compared to ABE7.10, although the higher activity of ABE8 variants may necessitate the use of Cas domain variants that decrease Cas-dependent off-target DNA editing or the use of deaminase variants described above that reduce Cas-independent off-target DNA and RNA editing (for example, ABE8e V106W or ABE8.17-m V106W)[227,228]. If there are bystander edits, the narrow-window ABE7.10 F148A may be considered[263]. If the target nucleotide is more PAM-proximal, ABE7.9 should be considered[215]. If the target A falls outside of the standard editing window but within protospacer positions 3–11, CP-1041 ABE can be used[237] (Fig. 5). Broadened ABE editing windows have been observed with ABE8 variants containing circularly permuted Cas9 domains that further extend this window (Fig. 5 and Supplementary Table 2)[227].

If Cas-dependent off-target DNA editing is a concern when using either CBEs or ABEs on the basis of experimental or computational analysis of the target protospacer, then high-fidelity Cas9 variants with enhanced DNA specificity should be considered (Supplementary Table 1). Although only some high-fidelity Cas domains have been tested as base editors, if the wild-type Cas domain is known to be compatible with the base editor, then corresponding well-characterized high-fidelity variants of that Cas domain may retain base editor activity (Supplementary Table 2)[127,152,158,184,247,251,273,276]. Truncations of the sgRNA can also help mitigate off-target effects with minor window-narrowing effects at the on-target locus[152]. Importantly, most high-fidelity Cas domains show lower tolerance for non-canonical guide RNAs (see above).

If Cas-independent DNA or RNA off-target editing cannot be tolerated for a particular application, both CBE and ABE deaminases have been developed to mitigate these issues. To minimize Cas-independent editing, narrow-window CBEs such as YE1, or variants of YE1 that have wider editing windows or expanded PAM compatibility, can be used[277]. Additional APOBEC variants with decreased Cas9-independent DNA off-target editing have been further engineered to decrease RNA off-target editing[246]. If off-target RNA editing is a concern and a standard editing window is desired, several variants have been developed to suit these needs[229,246,263,277,281]. If a broad editing window is required while minimizing RNA editing, CDA and AID can be considered[229]. Among the ABEs, ABE7.10 has not been reported to show Cas-independent DNA editing[279], while ABE8e shows modest increases in this form of off-target editing that are reduced to ABE7.10 levels by introducing the V106W mutation[227]. Off-target RNA editing activity from ABEs has been mitigated by the development of the ABE7.10-AW[282] and SECURE-ABE editors[229], as well as by introducing the V106W mutation to ABE8 and ABE8e, as ABE8 V106W and ABE8e V106W show similar RNA editing profiles to that of ABE7.10[227]. These constructs show significantly decreased RNA editing activity, often to levels similar to background levels from endogenous cellular RNA deamination. If these editors do not support sufficient on-target editing, ABE8.17m V106W[228] or ABE8e V106W[227] should be considered (Fig. 5).

General strategies for limiting cellular exposure to base editors beyond the duration necessary to achieve desired on-target editing levels—for example, by delivery of base editor RNPs rather than plasmids or viruses[158,173,190] or by using small-molecule-controlled editing agents—has been shown to greatly reduce off-target base editing, likely because on-target sequences are edited at much faster rates than off-target sequences[158,190,290–293]. The substantial improvements in DNA specificity that result from transient exposure of cells to editing agents create opportunities for the development of additional strategies to control editing agent activity at the post-translational level.

**Targeted random mutagenesis with base editors.** In some instances, it is desirable to use base editors to randomly mutagenize a locus of interest rather than to install a specific point mutation. Variants of CBEs and ABEs have been used to enable effective target locus random mutagenesis in bacterial[250], plant[254,287] and mammalian cells[244,257]. One strategy for targeted random mutagenesis uses Cas proteins fused to high-activity cytidine deaminases without UGI fusion or coexpression. Since abasic site generation likely mediates base scrambling following cytosine deamination and uracil excision, UGI inhibits base randomization. UGI-less CBEs using APOBEC1, A3A or AIDx (AID without a putative nuclear export sequence) have been shown to lead to higher rates of base scrambling at targeted cytosines[163,218,244,250,254,257,265]. This strategy likely can be applied to other cytidine deaminases as well, whereas it is unlikely to be effective with ABEs, which do not give rise to significant amounts of target base scrambling even when known inosine glycosylases are inhibited[27].

An alternative targeted random mutagenesis strategy has been to create base editors with both cytidine and adenosine deaminases fused to a single Cas domain[286–289] or tandem fusion of multiple cytidine deaminases[294]. While mutators of this type were reported with AID or CDA and the evolved ABE7.10 deaminase in tandem, other combinations from the set of the deaminases in Supplementary Table 3 are likely to work as well. These combination base editors generate both C→T and A→G edits, and some variants have removed the UGI domain to increase the base randomization incurred by cytosine deamination. Another mechanism for colocalizing multiple editing domains to a locus of interest uses the MS2 coat protein–aptamer pair (CRISPR-X)[244] or the Suntag epitope–monobody pair (BE-PLUS)[243], both of which drive multimeric protein recruitment and high levels of mutagenesis.

Base editor variants that introduce random mutations at targeted sites complement a recently developed mutagenesis tool, EvolvR, that consists of a fusion of a mutagenic DNA polymerase to a Cas9 nickase[295]. EvolvR has been shown to introduce random mutations within a window of up to 350 bp adjacent to targeted regions of the *E. coli* genome[295], although application of this tool for directed mutagenesis in mammalian cells has not yet been reported.

**Next steps for base editors.** Building on successive developments from many researchers, current-generation CBEs and ABEs are capable of efficiently installing precise transition mutations at targeted loci throughout a genome, in a variety of sequence contexts and in a wide range of cell types and organisms, including mammals. As base editors continue to advance toward clinical applications, their continued optimization to maximize their efficiency, specificity and ability to be delivered in vivo will remain an important priority. Developing additional strategies to control base editor activity temporally[158,173,190] or using exogenous small molecules[296,297] could further increase their utility as research tools and potential therapeutics.

Recent improvements in base editing efficiency and targeting capabilities make their application to install multiple different targeted point mutations, to randomize target bases in high-throughput genetic screens, or to record cellular signals and lineages especially promising. Given the unique mechanism of base editing—direct, self-contained conversion of one DNA base to another—and

its established compatibility with many types of dividing and non-dividing cells in vitro and in vivo, the development of additional base editors capable of installing transversion point mutations remains an opportunity to further expand the capabilities of genome editing in a manner that complements nuclease-initiated HDR and prime editing approaches.

## Transposases and recombinases

General methods for targeted integration in living cells have been a long-standing challenge in genome editing. While nuclease- and nickase-mediated HDR can insert genetic payloads of interest, these approaches are limited to actively dividing cells, constraining their usefulness for many applications. Recently described natural CRISPR-associated transposases and engineered Cas-domain-fused transposase systems can integrate genomic cargos in vitro and into bacterial genomes[298–300]. Engineered Cas-fused recombinases, which can in theory insert, delete, invert or replace target DNA, have also been reported to modify plasmid substrates in mammalian cells and to delete targeted genomic DNA in human cells, albeit so far with low efficiency and substantial target sequence restrictions[301].

**CRISPR-associated transposases.** Exploring the biodiversity of CRISPR systems led to the recent discovery of CRISPR-associated transposases that function in bacterial cells[298,299,302,303]. Computational analysis identified CRISPR loci containing CRISPR RNA array elements, Cas genes, and transposase-specific genes[15,302–307]. These loci appear to be the product of transposable genetic elements co-opting CRISPR–Cas machinery. Many of these loci, however, lack nuclease-active Cas genes. For example, some type I loci lack the essential Cas3 helicase–nuclease domain[303]. Type V loci include Cas12 variants predicted to be unable to function as nucleases[302]. Additional prophage-derived type IV systems, as well as subtype I-E variants, have been identified as candidate transposase-associated CRISPR–Cas systems with similar characteristics[302]. The putative lack of functional enzyme domains suggests that the Cas domains of these systems serve as genome-targeting modules for other nucleic acid effectors, such as the transposases associated with these loci. Some of these Tn7-like transposon variants have been reconstituted to perform CRISPR-associated transposon-mediated genomic integration events in bacteria[298,299].

Understanding the components of the Tn7 transposase and their function helps illuminate these new tools. Tn7-like transposons typically contain *tnsA*, *tnsB*, *tnsC*, *tnsD* and *tnsE* genes, as well as other genetic cargo. Typically, TnsA and TnsB form a TnsAB complex that specifically recognizes the left end (LE) and right end (RE) motifs that flank the transposon and catalyze the excision of the transposon from the donor locus. TnsB is a member of the retroviral integrase superfamily and catalyzes the 3′ cleavage of the phosphodiester backbone, while TnsA, a FokI-like protein, subsequently catalyzes cleavage of the 5′ ends, although this activity is not essential to all transposases[308,309]. TnsB then joins the free 3′ ends of the transposon DNA to a target DNA substrate determined by the TnsD or TnsE substrate-partitioning domains, leaving small gaps at the 5′ junctions. Repair of these gaps leads to a characteristic 5-bp target site duplication (Fig. 6a). TnsC is an ATPase that interacts with the TnsAB complex, duplex DNA, and one of TnsD or TnsE. CRISPR-associated transposons use many of these same components but replace the *tnsD* or *E* genes with a Cas domain for DNA targeting.

A type I-F locus isolated from *Vibrio cholerae* (Tn6677) was recently developed into a tool that inserts a genetic cargo of interest into *E. coli* genomes[298,310]. This system uses a combination of three plasmids to install cargo at target loci in the *E. coli* genome. The helper plasmid provides the transposase operon expressing TnsA, TnsB and TnsC. TnsB and TnsC were reported to be essential in catalyzing transposition, while a system with catalytically inactive TnsA still supported product formation, as this system is predicted to perform replicative transposition as opposed to cut-and-paste transposition[298,311]. A second plasmid, termed pQCascade, encodes the CRISPR–Cas associated machinery, as well as TniQ, a TnsD homolog. The CRISPR-associated components include the CRISPR array (a targeting RNA and spacer sequence, termed a guide RNA) as well as the *cas6*, *cas7* and *cas8–cas5* fusion genes. The third plasmid contains the donor LE–cargo–RE transposase DNA substrate that will be integrated into the target genomic locus. These components together form R-loop structures and perform transposition on these substrates in a coordinated manner that cannot be replicated by simple R-loop formation using dCas9[298]. The Tn6677 system uses a CC PAM motif that lies 5′ to a 32-bp protospacer sequence. Transposition cargo insertions occur 47 to 51 bp downstream of the end of the protospacer, with the integration distance being somewhat heterogenous and locus-dependent. Extending the guide RNA also modulates the integration distance[298]. Optimal cargo size was determined to be ~775 bp, although successful transposition of cargos up to 10 kb in length was observed. Cargos were found to integrate in either orientation. Product alleles generated by Tn6677 transposition contain the original PAM and protospacer, a 47- to 51-bp gap, the inserted sequence, the 5 bp downstream of the insertion site, the LE–cargo–RE, and a duplication of the 5-bp sequence downstream of the insertion site. Genome-wide off-target sites were not reproducibly detected for these tools in the *E. coli* genome[298]. Continued exploration of cargo compatibility, PAM preference, guide RNA sequences and locus requirements should improve our understanding of this system. Engineering or further exploring these systems to develop variants that can perform targeted transposition in mammalian cells would add an exciting capability to the eukaryotic genome editing toolbox.

Type V-K Tn7-like CRISPR-associated transposases have also recently been reported to support cargo transposition in *E. coli*[299]. This system, termed CAST for CRISPR-associated transposase, uses a helper plasmid that bears *tnsB*, *tnsC*, *tniQ*, *cas12k* and the Cas12k guide RNA array. This type V-K system lacks the *tnsA* gene. Cas12k is a catalytically inactive Cas12 variant, suggesting that this domain was solely a DNA-binding and R-loop-generating domain. Transposition was not observed when Cas12k was replaced with Cas9, suggesting that Cas12k or its specific interactions with the other CAST components facilitate product formation[299]. Two CAST variants were characterized: *Scytonema hofmannii* (ShCAST) and *Anabaena cylindrica* (AcCAST). ShCAST and AcCAST both reportedly require a 5′-NGTN-3′ PAM. Insertions for ShCAST occur predominantly between 60 and 66 bp 3′ of the PAM, while AcCAST cargos are inserted 49–56 bp 3′ of the PAM. CAST cargo integration was only observed in the forward 5′-LE–cargo–RE-3′ orientation. ShCAST exhibits a modest preference for W (A or T) 3 nt upstream of the insertion site (Fig. 6a). ShCAST supports insertion of cargo DNA up to 10 kb. Of 48 tested target loci in the *E. coli* genome, 29 (60%) showed detectable transposition. Off-target transposition was detected at multiple loci for ShCAST, unlike with the Tn6677 system[298]. As with the Tn6677 system, the activity of the CAST system in mammalian cells has not yet been reported.

**dCas9–transposase and recombinase fusions.** Fusions to dCas9 have also been used to facilitate locus-specific transposition and recombination. Recent efforts to develop CRISPR–Cas transposases inspired the development of a Himar transposase–dCas9 fusion[300]. This transposase is highly active as a dimer and does not require additional cellular factors to function. Himar can insert genomic cargos >7 kb into TA motifs in the genome. Wang and co-workers demonstrate that fusion of this highly sequence-permissive Himar monomer with dCas9 enables locus-enriched cargo integration into plasmids in *E. coli* using paired sgRNAs. The spacing from the insertion site to the protospacer can alter transposition

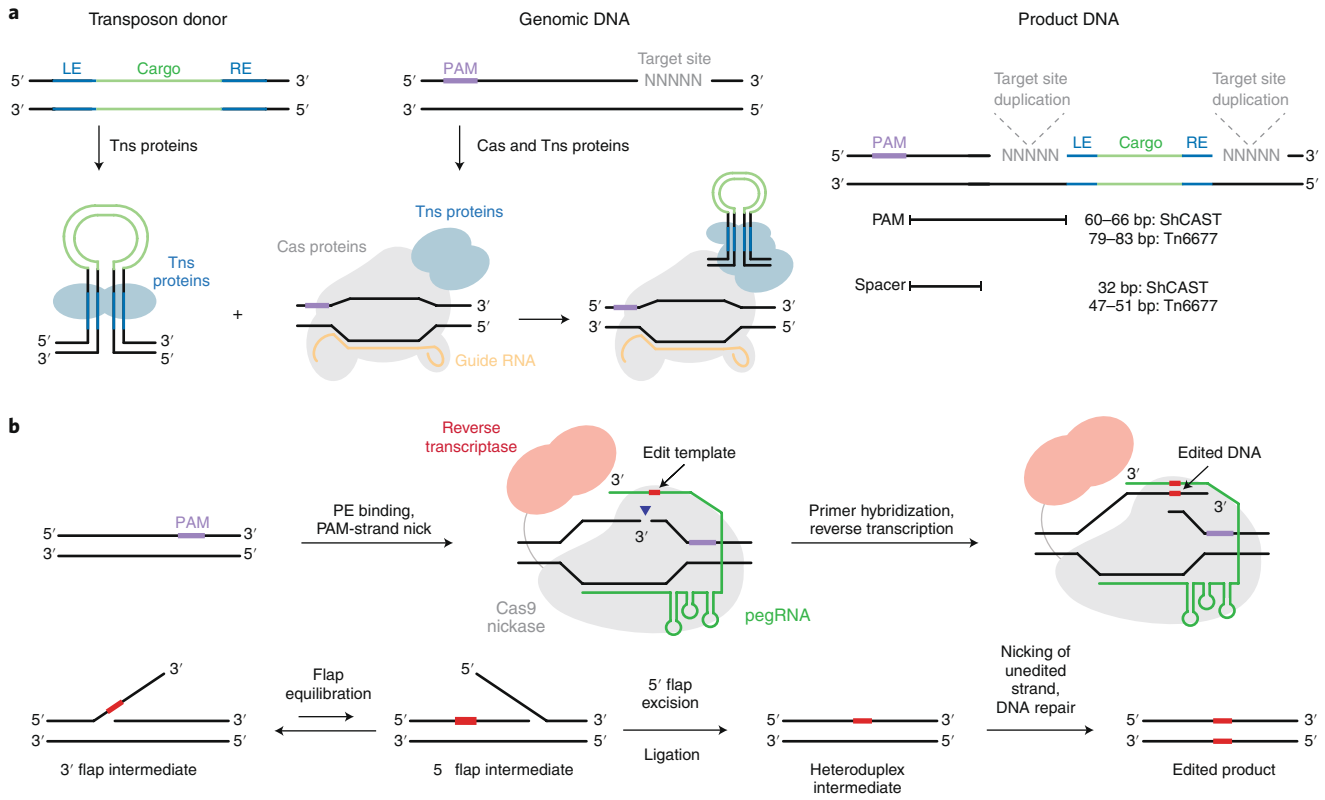



**Fig. 6 | Emerging CRISPR–Cas genome editing technologies: transposases and prime editors. a**, Cas transposases include both Cas proteins and transposase-associated components. Cargo DNA is identified by its left end (LE) and right end (RE) sequences and bound by transposase proteins (Tns). Cas proteins are directed to the target locus of interest in a PAM-dependent, RNA-directed manner. Cas binding localizes transposase proteins to the target sequence of interest and facilitates cargo integration at the target site. The target site is duplicated and flanks the integrated LE–cargo–RE sequence. Each Cas-transposase complex has a particular guide RNA length and preferred integration distance 3′ of the PAM. ShCAST, *S. hofmannii* CRISPR-associated transposase; Tn6677, CRISPR-associated transposase isolated from a *V. cholerae* type I-F locus. **b**, Prime editors consist of a Cas9 nickase domain fused to a reverse transcriptase domain. The spacer sequence of an engineered prime editing guide RNA (pegRNA) guides the prime editor to its genomic DNA target and also encodes the desired edit within an extension. After nicking the PAM-containing strand, the newly released genomic DNA 3′ end hybridizes to the pegRNA extension to form a primer–template complex. The reverse transcriptase domain then copies the template from the pegRNA extension into the genomic DNA directly, adding the edited sequence to the target locus. The product of reverse transcription, an edited 3′ flap, can then incorporate into the DNA duplex by competing with the original and redundant 5′ flap sequence. After 5′ flap excision and ligation of the edited strand, the non-edited complementary strand is replaced by DNA repair using the edited strand as a template. The latter step is facilitated by the adding an sgRNA to programs the prime editor protein to nick the non-edited strand away from the initial nick.

frequency, and sgRNA positions and orientations relative to the target TA sequence show modest effects on transposition outcomes. The high frequency of off-target cargo integration, as well as the restriction of the method to bacterial cells, limits its utility as a targeted genome-editing technique. These off-targets likely arise from *in trans* formation of dimers that can transpose at TA motifs across the genome.

Recombinases have also been explored as tools for targeted genomic modification. The stringent sequence requirements for these tools often limits their usefulness; however, efforts to engineer and evolve these agents have led to the development of non-programmable recombinases that can excise target genetic cargo of interest, such as the HIV genome[312,313]. Fusions of dCas9 with recombinase domains have also been explored in an attempt to overcome the sequence constraints of these tools. These highly restrictive sequence requirements limit off-target integration concerns. Liu and co-workers sought to overcome these high sequence constraints using Ginβ–dCas9 fusions[301]. Because Ginβ recombinase functions as an obligate dimer, two sgRNAs were used to adjacently localize two Ginβ monomers fused to separate dCas9 domains at the locus of interest. This ‘recCas9’ system showed modest efficiency on plasmid substrates and could mediate with low efficiency a large genomic deletion in mammalian cells. Although the sequence constraints of known recombinase domains, including Ginβ, limit their use, the ability of recombinases to perform a wide diversity of genome-modifying activities including insertions, deletions, replacements and inversions make them exciting targets for continued development.

**Next steps for CRISPR transposases and recombinases.** Although CRISPR-targeted transposases have been used to modify DNA substrates in vitro and in bacterial cells with impressive outcomes[298–300], activity in mammalian cells has yet to be reported. A limited number of the computationally predicted CAST systems have been characterized thus far. Investigating the properties of other nodes and comprehensive characterization of the genome targeting breadth and cargo limitations, as well as the continued study of potential off-target activities of these new tools, remain important research focuses. Enabling eukaryotic genome editing with CRISPR-targeted transposases, especially in human cells, would also mark a major milestone in the continued development of these technologies.

Further development of CRISPR-targeted recombinases is an exciting but challenging area of development in genome editing. The activity of known recombinases is highly sequence dependent, and the development of recombinases compatible with a wide range of sequences that could in principle be fused to a Cas targeting domain has proven difficult. Recent recombinase DNA profiling methods may facilitate these efforts[314]. Together, identifying, engineering and further developing CRISPR-targeted transposases and recombinases represent exciting opportunities in genome editing that may enable precise rearrangements of large DNA sequences of interest.

## Prime editing

Many genome editing applications, particularly those relating to the installation or correction of pathogenic mutations in mammalian genomes, require the introduction of precise point mutations, small insertions or small deletions[167]. As discussed above, CRISPR–Cas nucleases generate DSBs and can be used to disrupt, insert or delete DNA sequences using end-joining processes or to install precise changes using HDR in amenable cell types with high levels of accompanying indel byproducts. Base editors can install transition point mutations without generating DBSs, but cannot currently install transversion point mutations (C•G-to-A•T, C•G-to-G•C, T•A-to-A•T and T•A-to-G•C), precise insertions or precise deletions. Moreover, base editors can generate undesired bystander mutations when multiple target nucleotides exist within the base editing window, and PAM availability sometimes prevents targeting particular C or A bases[28].

Prime editing is a recent genome editing technology that can introduce all 12 possible types of point mutations (that is, all 6 possible base pair conversions), small insertions and small deletions in a precise and targeted manner with favorable editing to indel ratios[29]. Prime editors are fusion proteins between a Cas9 nickase domain (inactivated HNH nuclease) and an engineered reverse transcriptase domain. The prime editor protein, exemplified by PE2, is targeted to the editing site by an engineered prime editing guide RNA (pegRNA), which not only specifies the target site in its spacer sequence, but also encodes the desired edit in an extension that is typically at the 3′ end of the pegRNA. Upon target binding, the Cas9 RuvC nuclease domain nicks the PAM-containing DNA strand. The prime editor then uses the newly liberated 3′ end at the target DNA site to prime reverse transcription using the extension in the pegRNA as a template. Successful priming requires that the extension in the pegRNA contain a primer binding sequence (PBS) that can hybridize with the 3′ end of the nicked target DNA strand to form a primer•template complex. In addition, pegRNAs contain a reverse transcription template that directs the synthesis of the edited DNA strand onto the 3′ end of the target DNA strand. The reverse transcription template contains the desired DNA sequence change(s), as well as a region of homology to the target site to facilitate DNA repair (Fig. 6b).

After reverse transcription, the newly synthesized edited DNA strand exists as a 3′ DNA flap that is redundant with a 5′ flap containing the original, unedited DNA sequence. Cellular DNA repair processes are thought to excise the 5′ flap, allowing the edited 3′ DNA flap to be incorporated into the target site to generate heteroduplex DNA containing one edited and one non-edited strand (Fig. 6b). Finally, permanent installation of the edit occurs through replacement of the non-edited strand by DNA repair of the non-edited strand, which can be promoted by using a simple sgRNA to direct PE2 to nick the non-edited strand. This additional nick stimulates resynthesis of the non-edited strand using the edited strand as a template, resulting in a fully edited duplex (Fig. 6b).

Three versions of the prime editor system have been characterized. PE1 contains a fusion of Cas9 nickase to the wild-type Moloney murine leukemia virus (M-MLV) reverse transcriptase (RT). PE2 substitutes for the wild-type M-MLV reverse transcriptase an engineered pentamutant M-MLV RT that increases editing efficiency by about threefold. Finally, PE3 combines the PE2 fusion protein and pegRNA with an additional sgRNA that targets the non-edited strand for nicking, further increasing editing efficiency two- to fourfold. A variant of the PE3 system called PE3b uses a nicking sgRNA that targets only the edited sequence, resulting in decreased levels of indel products by preventing nicking of the non-edited DNA strand until the other strand has been converted to the edited sequence.

A major determinant of prime editing efficiency is the design of the pegRNA[29]. These pegRNAs contain 3′ extensions with a PBS and an RT template, each of which can be chosen from many plausible candidates. The combinatorial matrix of possible PBS and RT template pairings presents many possible designs to optimize the efficiency of a given desired edit, and typically only a small fraction of PBS and RT template combinations will support optimal prime editing efficiencies. In general, efficient PBSs will fall in a range between 8 and 15 nt, whereas RT templates are often optimally 10–20 nt in length. An exception is made for larger insertions and deletions (>10 nt), which appear to be more efficient with longer RT templates that incorporate additional homology to the region downstream of the edit[29]. Although optimal PBSs and RT templates are currently determined empirically, systematic insights into the factors that govern optimal pegRNA designs—such as GC content, primary sequence motifs and secondary structures within pegRNA 3′ extensions—would greatly facilitate the design of efficient prime editing agents for new targets.

Prime editors are able to install point mutations at distances far (>30 bp) from the site of Cas9 nicking, which offers greater targeting flexibility than nuclease-mediated HDR with ssDNA donor templates, which typically are unable to introduce edits efficiently more than ~10 bp from the cut site[117,315]. In principle, this feature also makes PAM availability less restrictive for prime editing. Cas9-dependent off-target prime editing was found to be much lower than Cas9 nuclease off-target indel generation at known Cas9 off-target sites[29], likely because two nucleic acid hybridization steps in addition to conventional protospacer–spacer annealing (that is, nicked target strand–PBS hybridization and 3′ flap–target strand hybridization) are required for productive prime editing. Each of these three hybridization steps provide an opportunity to reject off-target loci. More studies will be required to determine whether prime editors generate other forms of off-target genomic changes, such as those that might arise from incorporation of reverse-transcribed DNA away from the target site.

Prime editing has been tested in multiple human cell lines[29], postmitotic mouse cortical neurons[29], human induced pluripotent stem cells[316] and mouse embryos[317]. While multiple mammalian cell types support prime editing, they do so with varying efficiencies[29]. In addition, prime editors have been shown to support all types of nucleotide substitutions, as well as targeted insertions and deletions, in rice and wheat protoplasts, which can generate edited plants[318–321]. While the cellular factors that result in variable prime editing efficiencies across cell types has not yet been elucidated, differences in the expression levels of prime editing components or differences in the repertoire of expressed DNA repair proteins involved in prime editing could explain these observations. An understanding of how these factors affect prime editing efficiency could prove useful to further advance prime editing capabilities[322].

Compared with current-generation base editors, prime editors as initially reported are typically less efficient and generate higher levels of indels, although both editing modalities offer favorable editing/indel ratios. Prime editors can mediate all types of local edits, however, including those inaccessible to base editors; may be more easily targeted than other precision CRISPR editing methods because of greater flexibility in the distance between the PAM

and the edit; and also appear to be free of bystander editing since the sequence of the RT template determines the sequence of the edited DNA.

While prime editing brings a highly versatile new approach to precision genome editing, several key issues remain to be addressed for prime editing to achieve the widespread applicability and therapeutic potential demonstrated by CRISPR nucleases and base editors. These issues include illuminating the cell-state or cell-type determinants of prime editing efficiency, understanding the DNA repair mechanisms that result in productive or unproductive prime editing, and developing delivery strategies for in vivo applications that require delivery of the prime editor protein and pegRNA. Manipulation of DNA repair to favor the replacement of the 5′ flap by the edited 3′ flap, or to favor the replacement of the non-edited strand over the edited strand after successful incorporation of the 3′ flap, could improve prime editing efficiency and broaden its compatibility with less amenable cell types[322]. Moreover, in vivo delivery of nucleic acids encoding prime editors could be made more efficient by using smaller reverse transcriptase enzymes. These and other developments will expand the applicability of this new genome editing strategy in diverse research and therapeutic contexts.

## Conclusions

The progression from the initial discovery and characterization of CRISPR–Cas systems to their development into sophisticated, precise and versatile genome editing agents has occurred at a breathtaking pace. These tools have transformed the life sciences, enabling many advances in basic research, and established a promising foundation for the development of a new generation of human therapeutics that can treat—or even potentially cure—many human diseases with a genetic component. In less than 8 years, increasingly precise and versatile tools have been engineered and evolved that bring us nearer to the ultimate aspiration of being able to make any desired sequence alteration to the genome of any living cell without undesired editing byproducts.

The rapid evolution of genome-editing technologies has led us into a new era—an era in which we can now edit our own genomes, as well as the genomes of many other organisms that affect our communities. We are at the fragile beginnings of this new era. Continued efforts to improve editing capabilities, to understand all the consequences of editing our genomes, to innovate new ways to deliver editing agents into cells, and to fully engage scientists, doctors, ethicists, governments and other stakeholders will be crucial to guide our next steps and to ensure that these scientific advances can realize their full potential to benefit society.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0561-9.

Received: 3 February 2020; Accepted: 15 May 2020;
Published online: 22 June 2020

## References

1. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
2. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
3. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
4. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
5. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
6. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
7. Jinek, M. et al. RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
8. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
9. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).
10. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13**, 127–137 (2016).
11. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
12. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
13. Hille, F. et al. The biology of CRISPR-Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).
14. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
15. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
16. Makarova, K. S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
17. Jiang, F. & Doudna, J. A. CRISPR–Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
18. Deltcheva, E. et al. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
19. Garneau, J. E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
20. Chen, F. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.* **8**, 14958 (2017).
21. Schmid-Burgk, J. L. et al. Highly parallel profiling of Cas9 variant specificity. *Mol. Cell* **78**, 794–800.e8 (2020).
22. Shou, J., Li, J., Liu, Y. & Wu, Q. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell* **71**, 498–509.e4 (2018).
23. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
24. Szczelkun, M. D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl Acad. Sci. USA* **111**, 9798–9803 (2014).
25. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
26. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
27. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
28. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
29. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
30. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
31. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
32. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113 (2015).
33. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
34. Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
35. Esvelt, K. M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
36. Müller, M. et al. *Streptococcus thermophilus* CRISPR-Cas9 systems enable specific editing of the human genome. *Mol. Ther.* **24**, 636–644 (2016).
37. Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* **110**, 15644–15649 (2013).
38. Edraki, A. et al. A compact, high-accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. *Mol. Cell* **73**, 714–726.e4 (2019).

39. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* **8**, 14500 (2017).
40. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci. Adv.* **4**, eaau0766 (2018).
41. Hirano, H. et al. Structure and engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).
42. Harrington, L. B. et al. A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1424 (2017).
43. Kim, D., Luk, K., Wolfe, S. A. & Kim, J.-S. Evaluating and enhancing target specificity of gene-editing nucleases and deaminases. *Annu. Rev. Biochem.* **88**, 191–220 (2019).
44. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
45. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
46. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).
47. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* **60**, 385–397 (2015).
48. Strecker, J. et al. Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).
49. Yan, W. X. et al. Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91 (2019).
50. Burstein, D. et al. New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017).
51. Liu, J.-J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
52. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
53. Gootenberg, J. S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438–442 (2017).
54. Gootenberg, J. S. et al. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360**, 439–444 (2018).
55. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
56. Epinat, J.-C. et al. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* **31**, 2952–2962 (2003).
57. Stoddard, B. L. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19**, 7–15 (2011).
58. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**, 1169–1175 (2002).
59. Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
60. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
61. Carroll, D. Genome engineering with zinc-finger nucleases. *Genetics* **188**, 773–782 (2011).
62. Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
63. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
64. Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40**, 179–204 (2010).
65. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* **47**, 497–510 (2012).
66. Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21**, 1468–1478 (2019).
67. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
68. Heyer, W.-D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
69. Moynahan, M. E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **11**, 196–207 (2010).
70. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).
71. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
72. Brinkman, E. K. et al. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol. Cell* **70**, 801–813.e6 (2018).
73. van Overbeek, M. et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646 (2016).
74. Shen, M. W. et al. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
75. Allen, F. et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–72 (2018).
76. Chen, W. et al. Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res.* **47**, 7989–8003 (2019).
77. Leenay, R. T. et al. Large dataset enables prediction of repair after CRISPR-Cas9 editing in primary T cells. *Nat. Biotechnol.* **37**, 1034–1037 (2019).
78. Iyer, S. et al. Precise therapeutic gene correction by a simple nuclease-induced double-stranded break. *Nature* **568**, 561–565 (2019).
79. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
80. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
81. Doench, J. G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
82. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
83. Korkmaz, G. et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* **34**, 192–198 (2016).
84. Zhu, S. et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat. Biotechnol.* **34**, 1279–1286 (2016).
85. Montalbano, A., Canver, M. C. & Sanjana, N. E. High-throughput approaches to pinpoint function within the noncoding genome. *Mol. Cell* **68**, 44–59 (2017).
86. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
87. Dolan, A. E. et al. Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-Cas. *Mol. Cell* **74**, 936–950.e5 (2019).
88. Morisaka, H. et al. CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.* **10**, 5302 (2019).
89. Cameron, P. et al. Harnessing type I CRISPR-Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477 (2019).
90. Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24**, 142–153 (2014).
91. Nakade, S. et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* **5**, 5560 (2014).
92. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
93. Mani, R.-S. & Chinnaiyan, A. M. Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nat. Rev. Genet.* **11**, 819–829 (2010).
94. Rabbitts, T. H. Chromosomal translocations in human cancer. *Nature* **372**, 143–149 (1994).
95. Vanoli, F. et al. CRISPR-Cas9-guided oncogenic chromosomal translocations with conditional fusion protein expression in human mesenchymal cells. *Proc. Natl Acad. Sci. USA* **114**, 3696–3701 (2017).
96. Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat. Commun.* **5**, 3728 (2014).
97. Torres, R. et al. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* **5**, 3964 (2014).
98. Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423–427 (2014).
99. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
100. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl Acad. Sci. USA* **91**, 6064–6068 (1994).
101. Jasin, M. & Rothstein, R. Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.* **5**, a012740 (2013).
102. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).

103. Rees, H. A., Yeh, W.-H. & Liu, D. R. Development of hRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* **10**, 2212 (2019).
104. Srivastava, M. et al. An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* **151**, 1474–1487 (2012).
105. Robert, F., Barbeau, M., Éthier, S., Dostie, J. & Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **7**, 93 (2015).
106. Chu, V. T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).
107. Yu, C. et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **16**, 142–147 (2015).
108. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).
109. Pinder, J., Salsman, J. & Dellaire, G. Nuclear domain 'knock-in' screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic Acids Res.* **43**, 9379–9392 (2015).
110. Song, J. et al. RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat. Commun.* **7**, 10548 (2016).
111. Canny, M. D. et al. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **36**, 95–102 (2018).
112. Charpentier, M. et al. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* **9**, 1133 (2018).
113. Nambiar, T. S. et al. Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nat. Commun.* **10**, 3395 (2019).
114. Cullot, G. et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.* **10**, 1136 (2019).
115. Ihry, R. J. et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).
116. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
117. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
118. Renaud, J.-B. et al. Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Rep.* **14**, 2263–2272 (2016).
119. Savic, N. et al. Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife* **7**, e33761 (2018).
120. Aird, E. J., Lovendahl, K. N., St Martin, A., Harris, R. S. & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. Commun. *Biol* **1**, 54 (2018).
121. Carlson-Stevermer, J. et al. Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* **8**, 1711 (2017).
122. Nishiyama, J., Mikuni, T. & Yasuda, R. Virus-mediated genome editing via homology-directed repair in mitotic and postmitotic cells in mammalian brain. *Neuron* **96**, 755–768.e5 (2017).
123. Cebrian-Serrano, A. & Davies, B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* **28**, 247–261 (2017).
124. Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
125. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. *Mol. Cell* **61**, 886–894 (2016).
126. Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
127. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
128. Kim, H. K. et al. High-throughput analysis of the activities of xCas9, SpCas9-NG and SpCas9 at matched and mismatched target sequences in human cells. *Nat. Biomed. Eng.* **4**, 111–124 (2020).
129. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
130. Miller, S. M. et al. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat. Biotechnol.* **38**, 471–481 (2020).
131. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
132. Chatterjee, P. et al. A Cas9 with PAM recognition for adenine dinucleotides. *Nat. Commun.* **11**, 2474 (2020).
133. Chatterjee, P. et al. An engineered ScCas9 with broad PAM range and high specificity and activity. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0517-0 (2020).
134. Kleinstiver, B. P. et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
135. Gao, L. et al. Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792 (2017).
136. Kleinstiver, B. P. et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
137. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
138. Kleinstiver, B. P. et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).
139. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
140. Yan, W. X. et al. BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
141. Kim, H. K. et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods* **14**, 153–159 (2017).
142. Amrani, N. et al. NmeCas9 is an intrinsically high-fidelity genome-editing platform. *Genome Biol.* **19**, 214 (2018).
143. Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
144. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582 (2014).
145. Tsai, S. Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576 (2014).
146. Bitinaite, J., Wah, D. A., Aggarwal, A. K. & Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA* **95**, 10570–10575 (1998).
147. Bolukbasi, M. F. et al. Orthogonal Cas9–Cas9 chimeras provide a versatile platform for genome editing. *Nat. Commun.* **9**, 4856 (2018).
148. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
149. Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
150. Kulcsár, P. I. et al. Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol.* **18**, 190 (2017).
151. Casini, A. et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).
152. Lee, J. K. et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).
153. Vakulskas, C. A. et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **24**, 1216–1224 (2018).
154. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
155. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
156. Zuris, J. A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).
157. Wang, M. et al. Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proc. Natl Acad. Sci. USA* **113**, 2868–2873 (2016).
158. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
159. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
160. Kocak, D. D. et al. Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat. Biotechnol.* **37**, 657–666 (2019).
161. Kim, N. et al. Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0537-9 (2020).
162. Kim, S., Bae, T., Hwang, J. & Kim, J.-S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5′ nucleotides. *Genome Biol.* **18**, 218 (2017).
163. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
164. Molla, K. A. & Yang, Y. CRISPR/Cas-mediated base editing: technical considerations and practical applications. *Trends Biotechnol.* **37**, 1121–1142 (2019).

165. Yang, B., Yang, L. & Chen, J. Development and application of base editors. *CRISPR J.* **2**, 91–104 (2019).
166. Hess, G. T., Tycko, J., Yao, D. & Bassik, M. C. Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. *Mol. Cell* **68**, 26–43 (2017).
167. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016). (D1).
168. Liu, Z. et al. Highly efficient RNA-guided base editing in rabbit. *Nat. Commun.* **9**, 2717 (2018).
169. Zafra, M. P. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
170. Villiger, L. et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* **24**, 1519–1525 (2018).
171. Zhang, Y. et al. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* **8**, 118 (2017).
172. Li, Q. et al. CRISPR-Cas9-mediated base-editing screening in mice identifies DND1 amino acids that are critical for primordial germ cell development. *Nat. Cell Biol.* **20**, 1315–1325 (2018).
173. Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* **9**, 2184 (2018).
174. Zeng, Y. et al. Correction of the Marfan syndrome pathogenic FBN1 mutation by base editing in human cells and heterozygous embryos. *Mol. Ther.* **26**, 2631–2637 (2018).
175. Li, G. et al. Highly efficient and precise base editing in discarded human tripronuclear embryos. *Protein Cell* **8**, 776–779 (2017).
176. Zhou, C. et al. Highly efficient base editing in human tripronuclear zygotes. *Protein Cell* **8**, 772–775 (2017).
177. Liang, P. et al. Correction of β-thalassemia mutant by base editor in human embryos. *Protein Cell* **8**, 811–822 (2017).
178. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
179. Shimatani, Z. et al. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
180. Levy, J. M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat. Biomed. Eng.* **4**, 97–110 (2020).
181. Sasaguri, H. et al. Introduction of pathogenic mutations into the mouse Psen1 gene by Base Editor and Target-AID. *Nat. Commun.* **9**, 2892 (2018).
182. Lu, Y. & Zhu, J.-K. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 523–525 (2017).
183. Park, D.-S. et al. Targeted base editing via RNA-guided cytidine deaminases in *Xenopus laevis* embryos. *Mol. Cells* **40**, 823–827 (2017).
184. Liang, P. et al. Effective gene editing by high-fidelity base editor 2 in mouse zygotes. *Protein Cell* **8**, 601–611 (2017).
185. Xie, J. et al. Efficient base editing for multiple genes and loci in pigs using base editors. *Nat. Commun.* **10**, 2852 (2019).
186. Chadwick, A. C., Wang, X. & Musunuru, K. 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 (2017).
187. Rossidis, A. C. et al. In utero CRISPR-mediated therapeutic editing of metabolic genes. *Nat. Med.* **24**, 1513–1518 (2018).
188. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).
189. Farzadfard, F. et al. Single-nucleotide-resolution computing and memory in living cells. *Mol. Cell* **75**, 769–780.e4 (2019).
190. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
191. Kuscu, C. et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
192. Billon, P. et al. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of stop codons. *Mol. Cell* **67**, 1068–1079.e4 (2017).
193. Liu, Z. et al. Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat. Commun.* **9**, 2338 (2018).
194. Tanaka, S. et al. In vivo targeted single-nucleotide editing in zebrafish. *Sci. Rep.* **8**, 11423 (2018).
195. Wu, Y. et al. Increasing cytosine base editing scope and efficiency with engineered Cas9-PmCDA1 fusions and the modified sgRNA in rice. *Front. Genet.* **10**, 379 (2019).
196. Qin, L. et al. High-efficient and precise base editing of C•G to T•A in the allotetraploid cotton (*Gossypium hirsutum*) genome using a modified CRISPR/Cas9 system. *Plant Biotechnol. J.* **18**, 45–56 (2020).
197. Ren, B. et al. A CRISPR/Cas9 toolkit for efficient targeted base editing to induce genetic variations in rice. *Sci. China Life Sci.* **60**, 516–519 (2017).
198. Chen, W. et al. CRISPR/Cas9-based genome editing in *Pseudomonas aeruginosa* and cytidine deaminase-mediated base editing in Pseudomonas species. *iScience* **6**, 222–231 (2018).
199. Chen, Y. et al. CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in *Arabidopsis*. *Sci. China Life Sci.* **60**, 520–523 (2017).
200. Li, J., Sun, Y., Du, J., Zhao, Y. & Xia, L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 526–529 (2017).
201. Lee, H. K. et al. Simultaneous targeting of linked loci in mouse embryos using base editing. *Sci. Rep.* **9**, 1662 (2019).
202. Li, Y. et al. Programmable single and multiplex base-editing in *Bombyx mori* using RNA-guided cytidine deaminases. *G3 (Bethesda)* **8**, 1701–1709 (2018).
203. Gu, T. et al. Highly efficient base editing in *Staphylococcus aureus* using an engineered CRISPR RNA-guided cytidine deaminase. *Chem. Sci.* **9**, 3248–3253 (2018).
204. Yuan, J. et al. Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell* **72**, 380–394.e7 (2018).
205. Hua, K., Tao, X. & Zhu, J.-K. Expanding the base editing scope in rice by using Cas9 variants. *Plant Biotechnol. J.* **17**, 499–504 (2019).
206. Ren, B. et al. Cas9-NG greatly expands the targeting scope of the genome-editing toolkit by recognizing NG and other atypical PAMs in rice. *Mol. Plant* **12**, 1015–1026 (2019).
207. Ryu, S.-M. et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* **36**, 536–539 (2018).
208. Qin, W. et al. Precise A•T to G•C base editing in the zebrafish genome. *BMC Biol.* **16**, 139 (2018).
209. Li, C. et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* **19**, 59 (2018).
210. Liang, P. et al. Effective and precise adenine base editing in mouse zygotes. *Protein Cell* **9**, 808–813 (2018).
211. Ma, Y. et al. Highly efficient and precise base editing by engineered dCas9-guide tRNA adenosine deaminase in rats. *Cell Discov.* **4**, 39 (2018).
212. Kang, B.-C. et al. Precision genome engineering through adenine base editing in plants. *Nat. Plants* **4**, 427–431 (2018).
213. Yang, L. et al. Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell* **9**, 814–819 (2018).
214. Hua, K., Tao, X., Yuan, F., Wang, D. & Zhu, J.-K. Precise A·T to G·C base editing in the rice genome. *Mol. Plant* **11**, 627–630 (2018).
215. Song, C.-Q. et al. Adenine base editing in an adult mouse model of tyrosinaemia. *Nat. Biomed. Eng.* **4**, 125–130 (2020).
216. Lee, C. et al. CRISPR-Pass: gene rescue of nonsense mutations using adenine base editors. *Mol. Ther.* **27**, 1364–1371 (2019).
217. Yan, F. et al. Highly efficient A·T to G·C base editing by Cas9n-guided tRNA adenosine deaminase in rice. *Mol. Plant* **11**, 631–634 (2018).
218. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T: Abase editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
219. Wood, R. D. DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**, 135–167 (1996).
220. Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
221. Radany, E. H. et al. Increased spontaneous mutation frequency in human cells expressing the phage PBS2-encoded inhibitor of uracil-DNA glycosylase. *Mutat. Res.* **461**, 41–58 (2000).
222. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487–493 (1995).
223. Mol, C. D. et al. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80**, 869–878 (1995).
224. Wang, L. et al. Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res.* **27**, 1289–1292 (2017).
225. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl Acad. Sci. USA* **97**, 13573–13578 (2000).
226. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
227. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0453-z (2020).
228. Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0491-6 (2020).
229. Grünewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).

230. Rallapalli, K. L., Komor, A. C. & Paesani, F. Computer simulations explain mutation-induced effects on the DNA editing by adenine base editors. *Sci. Adv.* **6**, eaaz2309 (2020).
231. Karvelis, T. et al. PAM recognition by miniature CRISPR-Cas12f nuclease triggers programmable double-stranded DNA cleavage. *Nuc. Acids. Res.* **48**, 5016–5023 (2019).
232. Kweon, J. et al. A CRISPR-based base-editing screen for the functional assessment of BRCA1 variants. *Oncogene* **39**, 30–35 (2020).
233. Jun, S., Lim, H., Chun, H., Lee, J. H. & Bang, D. Single-cell analysis of a mutant library generated using CRISPR-guided deaminase in human melanoma cells. Commun. *Biol* **3**, 154 (2020).
234. Després, P. C., Dubé, A. K., Seki, M., Yachie, N. & Landry, C. R. Perturbing proteomes at single residue resolution using base editing. *Nat. Commun.* **11**, 1871 (2020).
235. Thuronyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
236. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
237. Huang, T. P. et al. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* **37**, 626–631 (2019).
238. Li, X. et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
239. Tan, J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun.* **10**, 439 (2019).
240. Tan, J., Zhang, F., Karcher, D. & Bock, R. Expanding the genome-targeting scope and the site selectivity of high-precision base editors. *Nat. Commun.* **11**, 629 (2020).
241. Oakes, B. L. et al. CRISPR-Cas9 circular permutants as programmable scaffolds for genome modification. *Cell* **176**, 254–267.e16 (2019).
242. Wang, Y., Zhou, L., Liu, N. & Yao, S. BE-PIGS: a base-editing tool with deaminases inlaid into Cas9 PI domain significantly expanded the editing scope. *Signal Transduct. Target. Ther.* **4**, 36 (2019).
243. Jiang, W. et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res.* **28**, 855–861 (2018).
244. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
245. McCann, J. L., Salamango, D. J., Law, E. K., Brown, W. L. & Harris, R. S. MagnEdit—interacting factors that recruit DNA-editing enzymes to single base targets. *Life Sci. Alliance* **3**, e201900606 (2020).
246. Yu, Y. et al. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, 2052 (2020).
247. Zhong, Z. et al. Improving plant genome editing with high-fidelity xCas9 and non-canonical PAM-targeting Cas9-NG. *Mol. Plant* **12**, 1027–1036 (2019).
248. Endo, M. et al. Genome editing in plants by engineered CRISPR-Cas9 recognizing NG PAM. *Nat. Plants* **5**, 14–17 (2019).
249. Shevidi, S., Uchida, A., Schudrowitz, N., Wessel, G. M. & Yajima, M. Single nucleotide editing without DNA cleavage using CRISPR/Cas9-deaminase in the sea urchin embryo. *Dev. Dyn.* **246**, 1036–1046 (2017).
250. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nat. Microbiol.* **3**, 423–429 (2018).
251. Xu, W. et al. Multiplex nucleotide editing by high-fidelity Cas9 variants with improved efficiency in rice. *BMC Plant Biol.* **19**, 511 (2019).
252. Liu, L. D. et al. Intrinsic nucleotide preference of diversifying base editors guides antibody ex vivo affinity maturation. *Cell Rep.* **25**, 884–892.e3 (2018).
253. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
254. Ren, B. et al. Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-guided hyperactive hAID mutant. *Mol. Plant* **11**, 623–626 (2018).
255. Liu, Z. et al. Improved base editor for efficient editing in GC contexts in rabbits with an optimized AID-Cas9 fusion. *FASEB J.* **33**, 9210–9219 (2019).
256. Wang, Y. et al. MACBETH: multiplex automated *Corynebacterium glutamicum* base editing method. *Metab. Eng.* **47**, 200–210 (2018).
257. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
258. Gehrke, J. M. et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
259. Coelho, M. A. et al. BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC Biol.* **16**, 150 (2018).
260. Liu, Z. et al. Highly precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. Preprint at *bioRxiv* https://doi.org/10.1101/658351 (2019).
261. Martin, A. S. et al. A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Sci. Rep.* **9**, 497 (2019).
262. St Martin, A. et al. A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC-Cas9 or cleavage by Cas9 in living cells. *Nucleic Acids Res.* **46**, e84 (2018).
263. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
264. Liu, Z. et al. Efficient and precise base editing in rabbits using human APOBEC3A-nCas9 fusions. *Cell Discov.* **5**, 31 (2019).
265. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).
266. Liu, Z. et al. Efficient base editing with high precision in rabbits using YFE-BE4max. *Cell Death Dis.* **11**, 36 (2020).
267. Zuo, E. et al. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* https://doi.org/10.1038/s41592-020-0832-x (2020).
268. Sancar, A. DNA excision repair. *Annu. Rev. Biochem.* **65**, 43–81 (1996).
269. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
270. Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
271. Kim, D. et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
272. Liang, P. et al. Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).
273. Kim, D., Kim, D. E., Lee, G., Cho, S.-I. & Kim, J.-S. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).
274. Tsai, S. Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).
275. Akcakaya, P. et al. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* **561**, 416–419 (2018).
276. Hong, R., Ma, S. & Wang, F. Improving the specificity of adenine base editor using high-fidelity Cas9. Preprint at *bioRxiv* https://doi.org/10.1101/712109 (2019).
277. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).
278. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
279. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
280. McGrath, E. et al. Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nat. Commun.* **10**, 5353 (2019).
281. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
282. Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **5**, eaax5717 (2019).
283. Hwang, G.-H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinforma.* **19**, 542 (2018).
284. Dandage, R., Després, P. C., Yachie, N. & Landry, C. R. beditor: a computational workflow for designing libraries of guide RNAs for CRISPR-mediated base editing. *Genetics* **212**, 377–385 (2019).
285. Arbab, M. et al. Determinants of base editing outcomes from target library analysis and machine learning. *Cell* (in the press).
286. Sakata, R. C. et al. Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0509-0 (2020).
287. Li, C. et al. Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-019-0393-7 (2020).
288. Zhang, X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0527-y (2020).
289. Grünewald, J. et al. A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0535-y (2020).
290. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).
291. Gao, X. et al. Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).
292. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).

293. Liu, K. I. et al. A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12**, 980–987 (2016).
294. Veillet, F. et al. Expanding the CRISPR toolbox in *P. patens* using SpCas9-NG variant and application for gene and base editing in *Solanaceae* crops. *Int. J. Mol. Sci.* **21**, 1024 (2020).
295. Halperin, S. O. et al. CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252 (2018).
296. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).
297. Kundert, K. et al. Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat. Commun.* **10**, 2127 (2019).
298. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).
299. Strecker, J. et al. RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48–53 (2019).
300. Chen, S. P. & Wang, H. H. An engineered Cas-transposon system for programmable and site-directed DNA transpositions. *CRISPR J.* **2**, 376–394 (2019).
301. Chaikind, B., Bessen, J. L., Thompson, D. B., Hu, J. H. & Liu, D. R. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* **44**, 9758–9770 (2016).
302. Faure, G. et al. CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* **17**, 513–525 (2019).
303. Peters, J. E., Makarova, K. S., Shmakov, S. & Koonin, E. V. Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl Acad. Sci. USA* **114**, E7358–E7366 (2017).
304. Koonin, E. V. & Makarova, K. S. Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol. Evol.* **9**, 2812–2825 (2017).
305. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
306. Krupovic, M., Béguin, P. & Koonin, E. V. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* **38**, 36–43 (2017).
307. Makarova, K. S. et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
308. Haren, L., Ton-Hoang, B. & Chandler, M. Integrating DNA: transposases and retroviral integrases. *Annu. Rev. Microbiol.* **53**, 245–281 (1999).
309. Peters, J. E. & Craig, N. L. Tn7: smarter than we thought. *Nat. Rev. Mol. Cell Biol.* **2**, 806–814 (2001).
310. Halpin-Healy, T. S., Klompe, S. E., Sternberg, S. H. & Fernández, I. S. Structural basis of DNA targeting by a transposon-encoded CRISPR-Cas system. *Nature* **577**, 271–274 (2020).
311. May, E. W. & Craig, N. L. Switching from cut-and-paste to replicative Tn7 transposition. *Science* **272**, 401–404 (1996).
312. Sarkar, I., Hauber, I., Hauber, J. & Buchholz, F. HIV-1 proviral DNA excision using an evolved recombinase. *Science* **316**, 1912–1915 (2007).
313. Karpinski, J. et al. Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat. Biotechnol.* **34**, 401–409 (2016).
314. Bessen, J. L. et al. High-resolution specificity profiling and off-target prediction for site-specific DNA recombinases. *Nat. Commun.* **10**, 1937 (2019).
315. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* **241**, 136–146 (2017).
316. Sürün, D. et al. Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. *Genes (Basel)* **11**, E511 (2020).
317. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell Discov.* **6**, 27 (2020).
318. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
319. Tang, X. et al. Plant prime editors enable precise gene editing in rice cells. *Mol. Plant* **13**, 667–670 (2020).
320. Li, H., Li, J., Chen, J., Yan, L. & Xia, L. Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* **13**, 671–674 (2020).
321. Xu, W. et al. Versatile nucleotides substitution in plant using an improved prime editing system. *Mol. Plant* **13**, 675–678 (2020).
322. Yan, J., Cirincione, A. & Adamson, B. Prime editing: precision genome editing by reverse transcription. *Mol. Cell* **77**, 210–212 (2020).


## Acknowledgements

This work was supported by the Merkin Institute of Transformative Technologies in Healthcare, US NIH grants U01AI142756, RM1HG009490, R01EB022376 and R35GM118062, and the HHMI. A.V.A. acknowledges a Jane Coffin Childs postdoctoral fellowship. L.W.K. acknowledges an NSF graduate fellowship.

## Competing interests

Authors through the Broad Institute and Harvard University have filed patent applications on genome editing technologies, including base editing and prime editing. D.R.L. is a consultant and cofounder of Prime Medicine, Beam Therapeutics, Pairwise Plants and Editas Medicine, companies that use genome editing.

## Additional information

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41587-020-0561-9.

**Correspondence** should be addressed to D.R.L.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature America, Inc. 2020