Molecular Therapy
Original Article

A precise and efficient adenine base editor

Tianxiang Tu,[1,7] Zongming Song,[2,7] Xiaoyu Liu,[1,7] Shengxing Wang,[3,7] Xiaoxue He,[1,7] Haitao Xi,[4] Jiahua Wang,[1] Tong Yan,[1] Haoran Chen,[1] Zhenwu Zhang,[5] Xiujuan Lv,[1] Jineng Lv,[1] Xiu-Feng Huang,[4] Junzhao Zhao,[4] Chao-Po Lin,[5] Caixia Gao,[3] Jinwei Zhang,[6] and Feng Gu[1]

[1]School of Ophthalmology and Optometry, Eye Hospital, State Key Laboratory and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China; [2]Henan Eye Hospital, Henan Eye Institute, Henan Provincial People's Hospital and People's Hospital of Zhengzhou University and People's Hospital of Henan University, Zhengzhou, Henan, China; [3]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China; [4]The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China; [5]School of Life Science and Technology, ShanghaiTech University, Shanghai, China; [6]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

Adenine base editors (ABEs) are novel genome-editing tools, and their activity has been greatly enhanced by eight additional mutations, thus named ABE8e. However, elevated catalytic activity was concomitant with frequent generation of bystander mutations. This bystander effect precludes its safe applications required in human gene therapy. To develop next-generation ABEs that are both catalytically efficient and positionally precise, we performed combinatorial engineering of NG-ABE8e. We identify a novel variant (NG-ABE9e), which harbors nine mutations. NG-ABE9e exhibits robust and precise base-editing activity in human cells, with more than 7-fold bystander editing reduction at some sites, compared with NG-ABE8e. To demonstrate its practical utility, we used NG-ABE9e to correct the frequent T17M mutation in *Rhodopsin* for autosomal dominant retinitis pigmentosa. It reduces bystander editing by ~4-fold while maintaining comparable efficiency. NG-ABE9e possesses substantially higher activity than NG-ABEmax and significantly lower bystander editing than NG-ABE8e in rice. Therefore, this study provides a versatile and improved adenine base editor for genome editing.

INTRODUCTION

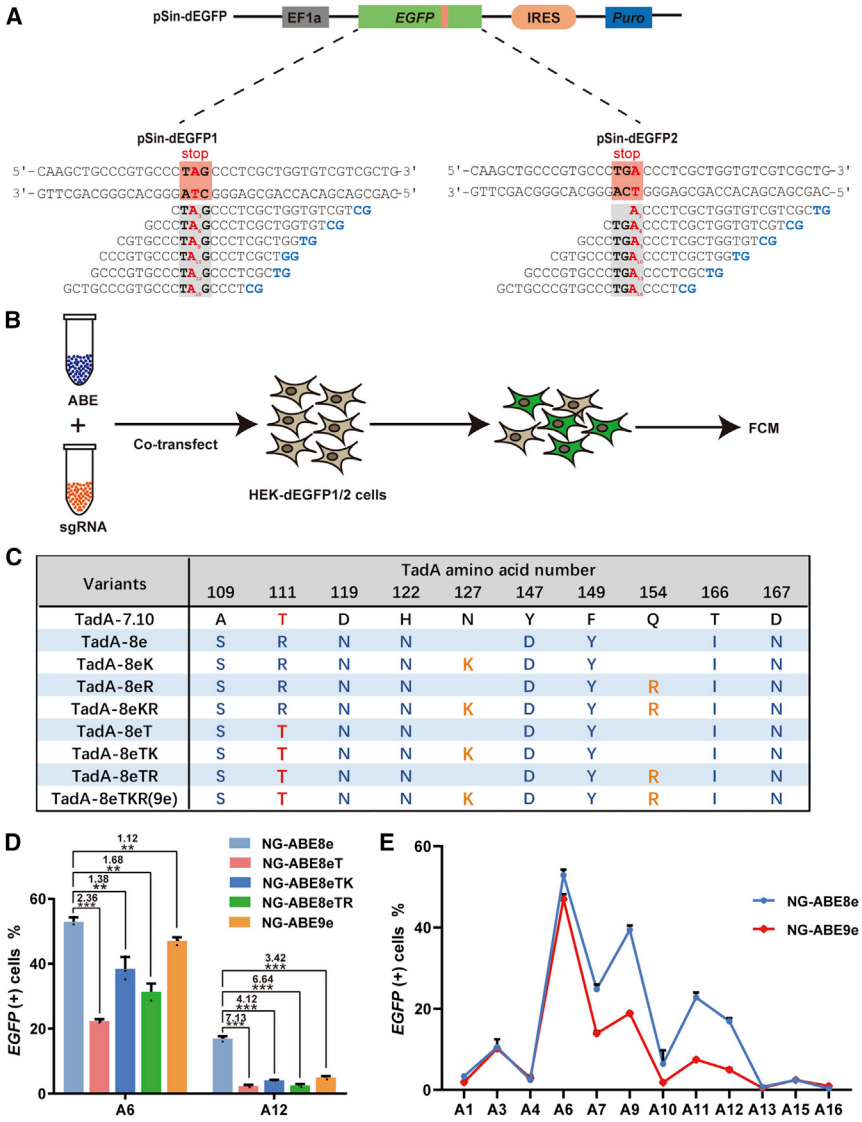
Point mutations (single substitutions) are the most prevalent disease-causing alterations in the human genome.[1] Thus, the ideal method for their correction is via precise point mutations that simply revert them back to wild type. However, this has been a challenging task to accomplish in practice. CRISPR-Cas9 provided the first generally applicable, high-efficiency genome editor. However, to perform editing, most conventional Cas9 systems must first create double-strand breaks (DSBs) that frequently lead to genome lesions and recombination. The recent advent of base editors avoids the formation of DSBs altogether and hold exceptional promise as novel editors that would be safe and reliable with further optimization.[2,3] The current base-editing technology can be classified into DNA and RNA base editors.[2–4] At least two classes of DNA base editor have been reported: adenine base editors (ABEs) and cytosine base editors (CBEs). While ABEs convert A·T base pairs to G·C pairs, CBEs convert C·G base pairs into T·A pairs.[2,3] These two base editors can mediate all four possible transition mutations (C to T, A to G, T to C, and G to A). Recently, C-to-G base editors (CGBEs) generating C-to-G transversion have also been reported.[5,6] Base editors have been successfully implemented in various organisms including human cells, animals, and plants.[7–17]

Despite increasingly wide use, base editors also suffer major limitations that currently preclude safe and reliable applications in clinical settings. The low efficiency of early ABEs severely hampered their applications until the identification of several high-activity protein variants including ABE8e, ABE8s, and, more recently, ABEmax-KR.[7,8,18] On the RNA side, chemical modifications of ABE mRNA and guide RNA also boosted editing efficiency.[19] Unfortunately, with the general elevation of the catalytic activity, the editing window of ABE8e was also greatly expanded. For instance, while the original, low-activity NG-ABEmax had an editing window spanning just four nucleotides (A4–A7 with the NG-Cas9 PAM at positions 21–22), NG-ABE8e could efficiently edit A11 in site 22 (efficiency >60%) and A12 in site 23 (efficiency>90%), approximately doubling its editing window.[7] As multiple adenosine nucleotides are typically found in reasonably close proximity in most genomic contexts, such substantial off-target effects and unscheduled generation of bystander

Received 7 December 2021; accepted 9 July 2022;
https://doi.org/10.1016/j.ymthe.2022.07.010.

[7]These authors contributed equally

Correspondence: Caixia Gao, State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.
E-mail: cxgao@genetics.ac.cn
Correspondence: Jinwei Zhang, Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA.
E-mail: jinwei.zhang@nih.gov
Correspondence: Feng Gu, School of Ophthalmology and Optometry, Eye Hospital, State Key Laboratory and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China.
E-mail: gufenguw@gmail.com



Variants	TadA amino acid number									
	109	111	119	122	127	147	149	154	166	167
TadA-7.10	A	T	D	H	N	Y	F	Q	T	D
TadA-8e	S	R	N	N		D	Y		I	N
TadA-8eK	S	R	N	N	K	D	Y		I	N
TadA-8eR	S	R	N	N		D	Y	R	I	N
TadA-8eKR	S	R	N	N	K	D	Y	R	I	N
TadA-8eT	S	T	N	N		D	Y		I	N
TadA-8eTK	S	T	N	N	K	D	Y		I	N
TadA-8eTR	S	T	N	N		D	Y	R	I	N
TadA-8eTKR(9e)	S	T	N	N	K	D	Y	R	I	N

Figure 1. Engineered ABE variants possessing narrowed editing window

(A) *EGFP* variants and sgRNA information. Target nucleotide (A1–A16) and PAM sequence (NG) are highlighted in red. The stop codon is marked with a light red shadow. (B) Scheme of protocol for testing the efficacy of ABEs. FCM, flow cytometry analysis. (C) The mutation information of ABE variants. 111T/127K/154R residues have been highlighted in red. (D and E) Editing efficiencies of ABE variants in EGFP turn-on reporter system. Data are from the same experiment. Error bars indicate mean ± SD (n = 3 independent experiments). $^{**}p < 0.01$, $^{***}p < 0.001$ by Student's unpaired two-sided t test. (D) Editing efficiencies of A-to-G mutations at two different positions (A6 and A12) induced by ABE variants. The ratio of NG-ABE8e's editing efficiency to ABE variants is provided. (E) NG-ABE9e has a narrowed editing window with the target sequence of EGFP.

mutations by ABE8e raise important safety concerns that preclude potential clinical applications such as gene therapies. Therefore, there exists an acute need to develop next-generation ABEs with precise editing windows to minimize off-target effects while retaining sufficient editing efficiency.

In light of the non-cognate, suboptimal interface between TadA and the hairpin-like DNA substrate that it is forced to edit and the sparse TadA protein sequence space sampled so far, we reason that there is a prospect to identify combinations of TadA mutations that may produce balanced editing efficiency and precision. To this end, we performed structure-guided combinatory protein engineering incorporating TadA sites mutated in NG-ABE8e and in NG-ABEmax-KR, two independently developed high-activity variants. Through this approach we identify NG-ABE9e that harbors nine total mutations in TadA, which exhibits superior editing precision and robust catalytic efficiency, compared with NG-ABE8e. We further demonstrate the practical utility of NG-ABE9e in correcting disease mutations in human cells and in genome engineering in plants. Finally, we provide a molecular basis for NG-ABE9e's superior editing characteristics, mediated by optimization of the network of electrostatic interactions between TadA and the topologically constrained DNA substrate.

RESULTS

A high-throughput genome-integrated, *EGFP*-based screening system for ABEs

Recently, we developed a screening system based on the engineered EGFP variant system to simultaneously measure ABE activity at multiple sites.[18] Specifically, we generated 12 partially overlapping single guide RNAs (sgRNAs) that allow a single editing site to shift gradually from the 5′ end (position A1), to near the 3′ end (position A16) within a 20-nt sliding window (Figure 1A). To make this system more stable, we generated cell lines (HEK-dEGFP1 or HEK-dEGFP2) harboring the EGFP expression cassette via lentivirus infection of HEK-293 cells instead of our original episomal plasmids as editing templates (Figure S1A). To quantitatively report editing efficiencies, we used NG-ABE8e to edit either an amber stop codon (TAG, dEGFP1) or an opal stop codon (TGA, dEGFP2) back to TGG (encoding W58 of EGFP). Thus, successful editing of either A at the second or third position to G would restore full-length EGFP expression, producing a substantial increase in fluorescence (Figure 1B). Consistent with previous reports,[7] our scanning editing site analysis revealed a wide, promiscuous editing window by NG-ABE8e, with efficient editing (~45%–68% edited) occurring from A6 to A12 with little preference

(Figures S1B and S2). These results demonstrated that our scanning EGFP turn-on reporter system is effective in quantifying editing precision at multiple sites and confirms NG-ABE8e's imprecise editing behavior.

Combinatory engineering of NG-ABE8e variants led to improved editing precision

Our recent study identified NG-ABEmax-R (Q154R, Q: glutamine, R: arginine) and NG-ABEmax-K (N127K, N: asparagine, K: lysine), which have higher editing activity than the original NG-ABEmax.[18] We first asked if these two mutations alone could modulate NG-ABE8e's editing window. We generated plasmids harboring these single or double mutations and performed the editing test. The results showed that the editing efficiency was marginally increased at A4 and A7 sites but slightly decreased at A6 and A9 sites (Figure S2). Thus, compared with NG-ABE8e, the addition of either mutation (Q154R or N127K) or both produced no change in editing precision.

Although the introduction of Q154R and N127K alone had no substantial impact on the editing window, we reasoned that those combinations of the mutations with the 8 original NG-ABE8e mutations may produce new phenotypes through epistatic effects. In the creation of the NG-ABE8e, T111R (T: threonine) in TadA was a critical mutation driving a large increase in editing efficiency[7] (Figure 1C). Since T111R and Q154R/N127K have similar effects and play somewhat redundant roles, we asked if reverting R111 back to T in the context of Q154R/N127K would reduce the overall activity to achieve improved activity-precision balance.

First, we confirmed that the NG-ABE8e harboring the R111T revertant mutation had substantially decreased editing efficiency (Figures S3 and S4), consistent with previous reports.[7] Subsequently, when we combined the single mutation (N127K or Q154R) with the R111T revertant mutation, we also observed low editing efficiency. Remarkably, when we combined the N127K/Q154R double mutation with R111T, the resulting new triple variant exhibited not only robust editing efficiency at the A6 site but also, importantly, drastically reduced editing at the A12 bystander site (Figure 1D). To ask if the observed reduced editing at A12 is a singular occurrence at a specific site or if it represents an across-the-board reduction in the editing window, we measured editing at adjacent sites. Encouragingly, we found that bystander editing is reduced by at least 2- to 3-fold at A9 and A11 sites as well (Figures 1E and S4), suggesting a systemic change in the editing precision in this mutant. Therefore, we term this triple mutant that harbors R111T/N127K/Q154R NG-ABE9e, as it contains nine mutations compared with the original NG-ABEmax.

Besides the exogenous test sequence in the *EGFP* gene, we next asked whether NG-ABE9e would also exhibit superior editing properties at endogenous gene loci. We selected 16 representative genomic sites for testing and found that off-target bystander editing was dramatically reduced at all of them by NG-ABE9e, with up to more than 7-fold reduction at some sites (Figures 2 and S5). We then compared NG-ABE8e and NG-ABE9e in additional human cell lines, including human embryonic stem cells (hESCs), HeLa, and HT1080. Similar results were also obtained from these human cell lines (Figure S6). Collectively, these results showed that a narrower editing window and reduced bystander editing are intrinsic properties of NG-ABE9e, which are manifest in diverse genomic contexts and cellular environments.

Further characterization of NG-ABE9e

To ask whether NG-ABE9e has altered off-target effects, we amplified fragments harboring the on target and predicted the off-target target site by PCR and performed sequencing to detect potential DNA off-target editing. We tested the two potential off-target sites for HEK2 and HEK4 respectively and three reported off-target sites for *VEGFA3*.[20] Then, we calculated the on-target/off-target values. We observed decreased off-target effects at these sites with NG-ABE9e, compared with NG-ABE8e, and the ratio of on-target to off-target editing is higher than that of NG-ABE8e (Figure S7). It is reported that ABE can also induce RNA off-target effects.[21,22] Thus, we investigated the off-target RNA editing of NG-ABE8e and NG-ABE9e. Based on the literature,[21] we tested two genes (*CTNNB1* and *IP90*), which are representative genes that produce abundant mRNAs in HEK-293 cells. The results showed that RNA off-target editing of NG-ABE9e is comparable to NG-ABE8e on these genes (Figure S8A). Additionally, to investigate the cellular RNA editing by NG-ABE9e, we measured the A-to-I substitution frequency across the transcriptome using RNA sequencing (RNA-seq), which showed the relatively lower A-to-I substitution frequency of NG-ABE9e compared with Cas9(D10A) or NG-ABE8e (Figure S8B).

To ask if altered expression levels of ABEs may have contributed to the change in editing behavior, we measured the expression level of NG-ABE8e and NG-ABE9e using western blotting analysis. We found that there is slightly higher expression of NG-ABE9e compared with NG-ABE8e (Figure S9), indicating that NG-ABE9e does not achieve more precise editing by reducing the availability of the editors.

Besides all these, we compared NG-ABE9e with other high activity ABE variants. NG-ABEmax has been shown to have appreciable editing activity in some cell lines and organisms. So, we compared the activity of NG-ABE9e with NG-ABE8e and NG-ABEmax at 6 endogenous sites in HEK-293 cells. We found that they have a similar editing activity at the A5 site, but NG-ABE9e exhibited higher editing efficiency at the A7 site in human endogenous sites, compared with the low efficacy of NG-ABEmax here (Figure S11A). We also compared the trio of NG-ABEmax, NG-ABE8e, and NG-ABE9e in parallel in HEK-dEGFP1/2 cells and found that the editing efficacy of NG-ABEmax is much lower than NG-ABE8e or NG-ABE9e across all tested sites (Figure S12), highlighting the limitation of NG-ABEmax at least at certain sites.

Superior editing property of NG-ABE9e in rice

The efficiency and precision of genome editing can substantially vary depending on the properties of the host organism and cell type, since the exogenous gene editors must operate within the specific cellular

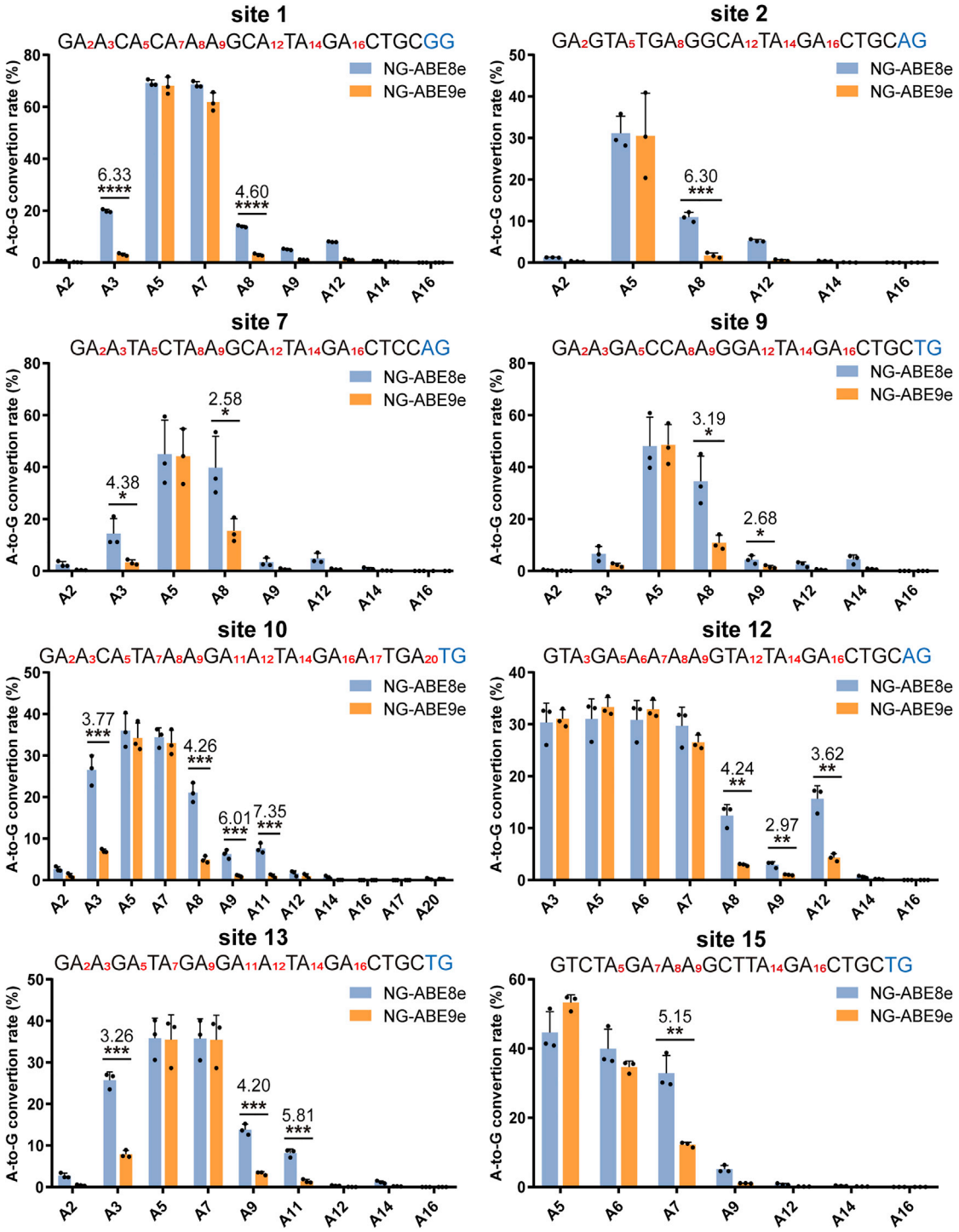


Figure 2. NG-ABE9e possessing high fidelity to edit human endogenous genes

The editing efficiency of NG-ABE8e and NG-ABE9e in human cells. Each A base is highlighted with red, and PAM sequences are in blue. The ratio of NG-ABE8e's editing efficiency to NG-ABE9e is provided. The target sequence information is provided in the supplemental information. Error bars indicate mean ± SD (n = 3 independent experiments). *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 by Student's unpaired two-sided t test.

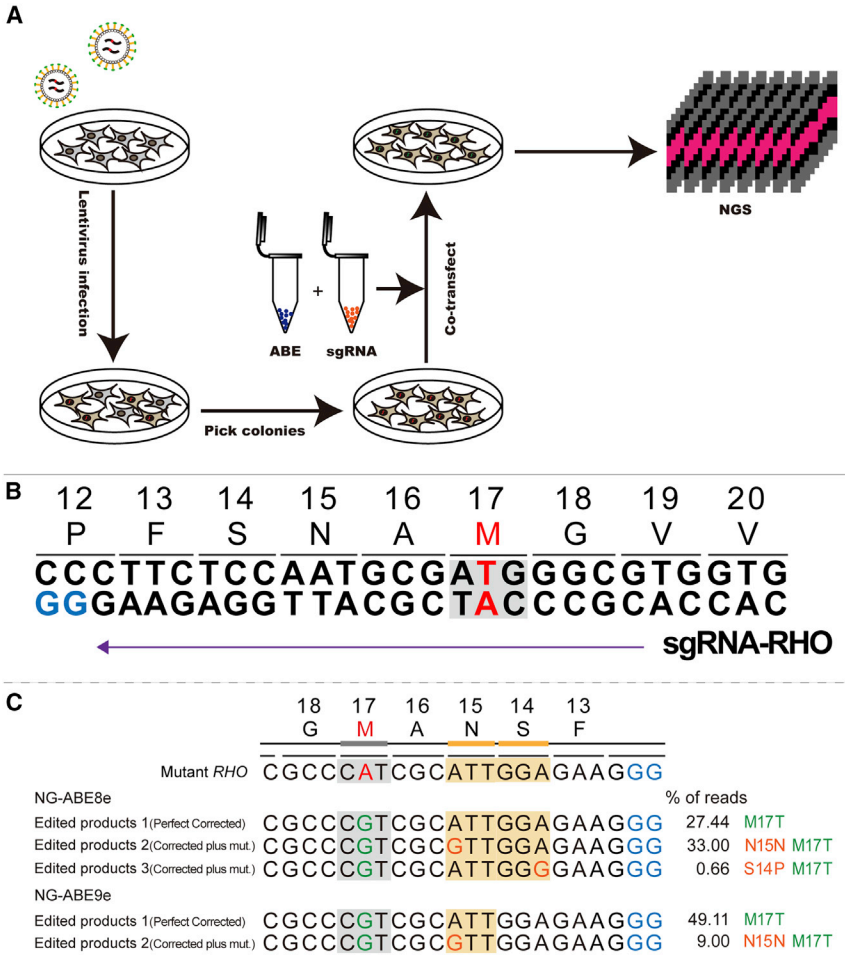


Figure 3. Superior editing precision of NG-ABE9e for gene correction

(A) Schematic of generation of HEK-293 cell line harboring *RHO* pathogenic mutation (T17M) and correction. The lentivirus with the T17M target sequences used to infect HEK-293 cell. Then, the single colony cell line with T17M mutation was selected with puromycin. (B) The pathogenic mutation and sgRNA information (in purple) are provided. The pathogenic mutation (T17M) is highlighted in red. (C) Editing products triggered with NG-ABE8e and NG-ABE9e with sgRNA-RHO, respectively. The parental and corrected nucleotides are in red and green, respectively. The bystander mutations are in orange. Data are shown as the mean (n = 3 independent experiments).

contexts including numerous host protein and RNA machineries and diverse environments such as nuclear and chromatin structures and intracellular osmolarity. To ask if the superior editing properties of NG-ABE9e are generally applicable in other cellular hosts, we also compared several ABE variants, including ABE7.10 (labeled as PABE7.10), ABEmax (PABEmax), NG-ABE8e (PABE8e), and NG-ABE9e (PABE9e), in editing nine genomic sites in rice. We found that although PABE9e exhibited slightly reduced overall editing efficiency compared with PABE8e, it is substantially more precise, driving up to 5-fold reduction in bystander editing (Figure S10). Notably, PABE9e also exhibited substantially high editing efficiency compared with PABEmax and PABE7.10 in activity editing window in rice (Figure S11B). Surprisingly, as to the performance of activity and bystander editing, NG-ABE9e is much better in rice than that in human cells.

Taken together, based on the results of editing in multiple human cell lines and in plant cells, we conclude that NG-ABE9e is broadly applicable across multiple cellular hosts and types and exhibits a consistent, inherent ability to drive high-precision, high-efficiency base editing.

Superior editing precision of NG-ABE9e for human gene therapy

To evaluate the performance of NG-ABE9e in precisely correcting disease-causing mutations in practical applications, we used it to revert a vision loss disease mutation, T17M in rhodopsin (*RHO*). This mutation causes autosomal dominant retinitis pigmentosa (adRP), and ultimately blindness, by compromising the folding and thermostability of the Rhodopsin protein, an essential photoreceptor protein required for image-forming vision and photoreceptor cell viability.[23,24] Since adRP is an autosomal dominant disease caused by gain-of-function mutations, direct reversion of the point mutation is expected to be an effective therapeutic strategy to treat this disease. Hence, we generated a HEK-293 cell line harboring the pathogenic T17M mutation in the *RHO* gene and designed a sgRNA to introduce the revertant M17T mutation (A6, CAT to CGT, editing site underlined) using NG-ABE8e or NG-ABE9e (Figures 3A and 3B). We observed that both NG-ABE8e and NG-ABE9e had relatively high and comparable A6 to G6 editing efficiencies of 61% and 58%, respectively (Figure 3C). However, their editing precision was drastically different. Specifically, NG-ABE8e achieved about 27% "perfect editing," meaning correcting the pathogenic mutation without any bystander mutation. Remarkably, under the same conditions and using the same sgRNA, NG-ABE9e achieved 49% perfect editing, nearly twice that of NG-ABE8e (Figure 3C). Strikingly, NG-ABE9e produced only 9% bystander mutations, about a quarter of the 33% by NG-ABE8e (Figure 3C). These findings suggest that NG-ABE9e is substantially more precise than NG-ABE8e in safely and reliably correcting disease-causing mutations.

Molecular basis of the enhanced editing-site selectivity of ABE9e

To understand how NG-ABE9e achieves higher editing-site selectivity while maintaining robust activity, we performed structural

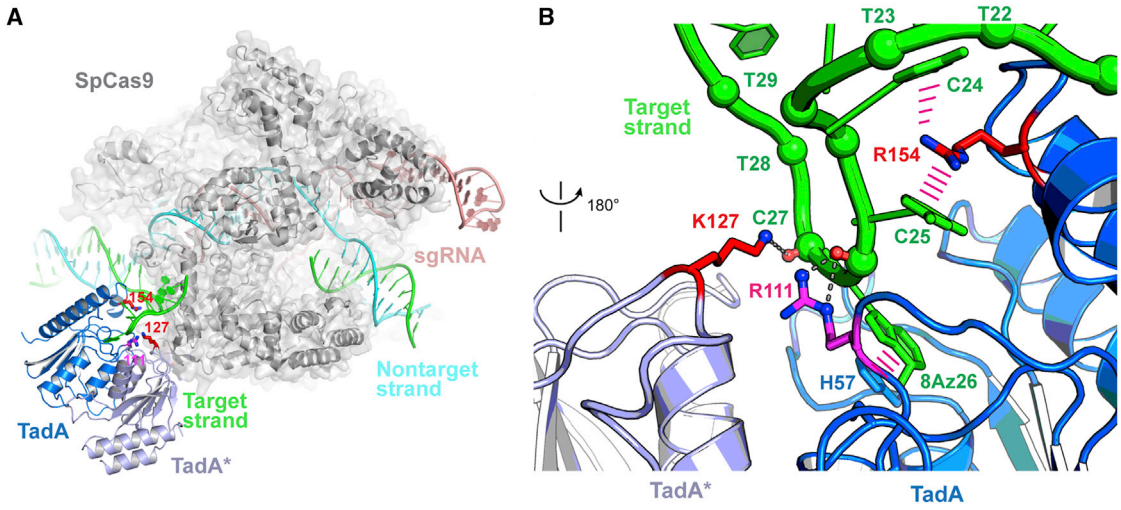


Figure 4. Structural analysis and modeling of ABE variants

(A) Overall structural model of ABE8e-KR showing SpCas9 (gray), target-strand DNA (TS; green), nontarget-strand DNA (NTS; cyan), single-target RNA (sgRNA; pink), TadA (blue), and TadA* (light blue). The side chains of N127K of TadA* and Q154R of TadA are shown as red sticks. The side chain of R111 is shown as magenta sticks. (B) Modeled interactions between key residues of ABE8e-KR (R111, K127, R154) with TS DNA, colored as in (A). Gray dotted lines denote proposed hydrogen bonds; parallel magenta lines indicate cation-π or π-π interactions.

modeling analysis based on a cryoelectron microscopy (cryo-EM) structure of ABE8e bound to its DNA target.[25] The single-stranded DNA substrate assumes a conformationally constrained, inside-out hairpin-like structure that is recognized by both subunits of TadA near the dimer interface (Figures 4A and 4B). The original TadA homodimer evolved to recognize the anticodon stem loop (ASL) structure of tRNAs.[26] Comparing the DNA structure with that of the tRNA ASL reveals that this DNA structure only locally tracks and mimics the cognate tRNA substrate in the immediate vicinity of the catalytic center.[25] Beyond this immediate region, the overall trajectory of the DNA is distinctly different from the tRNA ASL and makes dissimilar contacts with TadA. This observation suggests ample opportunities for further mutagenesis and optimization of the TadA-DNA interface to elicit desirable traits in base editors.

Indeed, congruent with the modeling analysis, the double mutant of N127K/Q154R (NG-ABEmax-KR) we recently isolated through directed evolution substantially improved the catalytic efficiency of TadA on DNA.[18] This phenotype is most likely driven by enhancing the electrostatic interactions to the backbone by K127 on one side of the DNA, as well as the addition of new cation-π interactions between R154 and nucleobases of C24 and C25 on the opposite side (Figure 4B). Interestingly, the guanidinium side chain of R111 is located only ~4 Å away from the epsilon-amine of K127, both of which are in range to contact the backbone non-bridging phosphate oxygens of DNA. In fact, R111 is seen to makes two hydrogen bonds to the backbone near the apex of the DNA hairpin and the catalytic site (Figure 4B). Thus, R111 and K127, neither original residues of TadA, appear functionally redundant. Their combined positive charge may over-stabilize the strained conformation of DNA substrate, thus promoting promiscuous deamination at undesired bystander sites. Elimination of the excess positive charge on R111 via its reversion to threonine (T) reduces the overall strength of the Tad-DNA interactions, thereby simultaneously achieving high editing-site specificity and maintaining high catalytic efficiency. Together, structural modeling analyses provide a feasible molecular basis for the observed improved editing characteristics of NG-ABE9e under study and suggest that structure-guided rational engineering and optimization of the network of TadA-DNA electrostatic contacts can tune the catalytic behavior of ABEs in predictable manners.

DISCUSSION

It is reported that a considerable fraction (47%) of disease-associated point mutations are conversions of a G/C pair to an A/T pair in nature. Thus, ABEs that enzymatically revert such conversions are particularly promising tools for human gene therapy.[16,27] The deaminases of ABEs enable A-to-G conversion within a predefined window of activity on the DNA substrate.[2] Recently, several ABE variants (ABE8e, ABE8s, and ABEmax-KR) with high activity have been identified,[7,8,18] which greatly augmented ABE-mediated genome editing. However, these high-activity ABE variants also have a high tendency to introduce bystander mutations in the target sequence, which preclude its applications, especially clinical treatments where safety is paramount. To address this conundrum, here we developed a high-fidelity, high-activity ABE with minimized bystander off-target effects. The effectiveness of NG-ABE9e also provides a proof of principle that the activity and specificity of ABEs, and likely of other base editors, are separable traits that can be engineered independently, allowing multiple desirable traits to be achieved simultaneously. This is largely enabled by the vast sequence space available to even a relatively small protein such as TadA. Current selection and screening technologies can only sample

a small fraction of amino acid sequence space and are thus unable to thoroughly explore epistatic effects when multiple mutations occur simultaneously. This deficiency leaves ample opportunities to rationally engineer protein and protein-DNA/RNA interfaces guided by structural modeling and molecular dynamics simulations to create new combination variants with multiple desirable traits. Another enabling factor for directed ABE evolution and further optimization is the suboptimal, non-cognate interface between TadA and the DNA substrate and the conformational strain imposed on the DNA to mimic a cognate RNA hairpin structure (Figures 4A and 4B). From this perspective, other base-editing enzymes that do not require specific DNA distortions may provide more efficient and predictable editing than the transplanted tRNA deaminase TadA.

In this study, we generated different variants harboring the mutations from the NG-ABE8e and our recently identified NG-ABEmax-KR variants and took advantage of an *EGFP*-based scanning editing reporter system to select ABE variants with narrowed editing windows (Figure S1). Using structure-guided combinatory protein engineering, we identified one triple mutant (NG-ABE9e) possessing both high precision and robust activity. Based on our results from human cells (including HEK-293, hESCs, HeLa, and HT1080) and rice, we found the canonical editing window for NG-ABE8e to be A3–A12, indicating that multiple A bases within the broad window would be edited. By contrast, NG-ABE9e exhibited substantially reduced activity at A3 and A8–A12, which minimized off-target editing at these flanking positions (Figures 1D, 1E, 2, 3, S5, S6, and S10). Of note, as to the editing efficiency in the central, targeted area, A4–A7, only very slight reduction in editing activity was observed with NG-ABE9e (Figures 1E, 2, S5, and S6). To demonstrate NG-ABE9e's utility in practical gene-therapy applications, we showed that NG-ABE9e effectively reverted a disease-causing mutation in *RHO* with much higher precision than NG-ABE8e, generating much fewer unscheduled nucleotide change (Figure 3C). We note that the major bystander mutations in *RHO* happen to be synonymous codon changes that do not change the protein sequence. Nonetheless, these types of mutations could still be pathogenic due to altered mRNA splicing, codon usage, or RNA structure leading to abnormal protein expression.[16,28–30] To ask how NG-ABE9e may impact DNA and RNA off-target effects, we also characterized the general fidelity of NG-ABE9e and observed lower DNA off-target effects at the predicted sites (Figure S7) and relatively lower RNA off-target editing (Figure S8), congruent with its more precise nature.

So far, the identification of a perfect base editor is still a challenge. Our study showed that NG-ABE9e possesses improved editing precision compared with NG-ABE8e. However, we note that at the A5 site in certain target sequences, the editing efficacy is saturated with NG-ABEmax, NG-ABE8e, or NG-ABE9e. At these particular target sequences harboring A5, NG-ABEmax may generate fewer bystander mutations than NG-ABE8e and NG-ABE9e. However, the editing efficiency of NG-ABEmax is unacceptably low at certain sequences (site 17; Figure S11). By contrast, NG-ABE9e exhibited drastically improved editing efficiencies than NG-ABEmax at nearly all A7 sites. For comparison, NG-ABE8e edited the wide A3–A12 window, while NG-ABE9e only edited A4–A7. Thus, when disease mutation occurs at A7, NG-ABE9e is generally the preferred base editor.

During the combination engineering of the NG-ABE8e variants (Figure 1D), we found NG-ABE8eTK possessing slightly lower activity at A6 and slightly higher activity at A9 and thus chose NG-ABE9e for further experiments. NG-ABE8eTK remains a promising candidate for precise editing for future analyses.

Mirroring our and others' efforts to improve the precision of ABEs, different strategies have also been employed to create higher-fidelity CBEs, as CBEs induce genome-wide off-target mutations.[31] These include the introduction of various mutations in the deaminase (YEE-BE3 and YFE-BE4max),[32,33] engineering human APOBEC3A (eA3A-BE3),[34] and exploiting a shorter linker sequence plus removing a non-essential part from CDA1 (nCDA1-BE3).[35] Despite substantial improvement in precision, the editing window of NG-ABE9e is still not completely satisfactory, with the goal being single base correction only. Conceivably, these aforementioned strategies that improved CBEs could also be adopted to further engineer ABEs to completely eliminate bystander mutations. In reciprocation, our strategies in improving ABEs are also transferrable to CBEs and CGBEs.

Interestingly, it is reported that ABEs can also induce cytosine deamination at the target site. Indeed, a D108Q mutation in TadA in ABE7.10 reduced cytosine deamination activity by 10-fold.[36] In our study, we did not observe any bystander cytosines edits, so it is unknown if NG-ABE9e would also reduce bystander cytosine editing at sequence contexts that presumably drive higher cytosine editing by ABEs. Indeed, machine-learning methods recently used to predict the sequence context effects on CGBE[37] can be similarly applied to ABEs and CBEs to provide additional guides to execute a precise gene-editing application.

Taken together, our study identifies, characterizes, and validates NG-ABE9e, a novel base editor with both robust editing efficiency and precise genome editing in human and plant cells. Thus, NG-ABE9e provides a practical and versatile alternative for genome editing to NG-ABEmax and NG-ABE8e.

MATERIAL AND METHODS

Plasmid construction

The original plasmid NG-ABE8e was obtained from Addgene (Addgene plasmid #138491). The plasmid pSin-EGFP containing a mutant *EGFP* gene as well as mutant *RHO*, *IRES*, and *Puromycin* genes were generated as previously described.[38] The sgRNA expression cassettes were generated by standard protocol. Desired point mutations were introduced into the coding sequence of *EGFP* by PCR to generate *EGFP* variant, *dEGFP1*, and *dEGFP2*.[18]

To generate PABE-8e and PABE-9e expression vectors, the tRNA editing deaminase ecTadA-8e, 32aa linker, and nSpCas9 (D10A)

sequences were codon optimized for cereal plants and synthesized (GenScript, Nanjing, China). These fusion protein sequences were cloned into the pJIT163 backbone using a ClonExpressII One Step Cloning Kit (Vazyme, Nanjing, China), generating PABE-8e. Point mutations (R111T, N127K, and Q154R) were introduced into the ecTadA-8e coding sequence of PABE-8e, generating PABE-9e. The coding sequences of the two above constructs are listed in the supplemental information. The constructs pOsU3-sgRNA were generated by standard protocol. All plasmids were confirmed by Sanger sequencing. The oligonucleotide sequences used for plasmid construction and PCR in this study are listed in Tables S1–S7.

Cells and cell culture

HEK-293 cells were obtained from ATCC (CAT#CRL-1573). HEK-293 cells were grown at 37°C in 5% CO_2 in Dulbecco's modified Eagle's medium (Life Technologies, Carlsbad, CA, USA) supplemented with 10% heat-inactivated fetal bovine serum and penicillin/streptomycin. HeLa and HT1080 cells were obtained from ATCC (CAT#CCL-2 and CCL-121) and cultured as the condition of HEK-293.

Transfection protocol, genomic DNA extraction, images and flow cytometry/Western blotting analysis

For ABE variant experiments, HEK-293 cells were seeded at 0.9×10^5 cells per well on 24-well plates in Dulbecco's modified Eagle's medium (DMEM). Twenty-four h after seeding, cells were co-transfected with 200 ng ABE variant plasmids, 100 ng sgRNA expression plasmids, 200 ng dEGFP1 plasmids, and 1.5 μL TurboFect Transfection Reagent (Thermo Fisher Scientific). Genomic DNA extraction of cells was subsequently performed using genome extraction kit (Vazyme). Images were obtained at 48 h post-transfection, and cells were collected at 48 h post-transfection. Flow cytometry analysis was performed with FACSAria II (BD Biosciences). Western blotting analysis was performed as previously described.[38]

Next-generation sequencing (NGS)

The genomic region flanking the sgRNA target site was amplified with TransStart FastPfu DNA polymerase (TransGen Biotech, Beijing, China) using site-specific primers (Tables S5–S7) in the first-round PCR. For the second-step PCR amplification, we fixed the barcodes, index, and adaptor sequences to the first-step PCR amplification products (Tables S6 and S7). The second-step PCR amplification products were purified and pooled and subsequently subjected to paired-end read sequencing using the Hiseq-PE150 strategy at Novogene (Nanjing, China). Finally, open-sourced "CRISPResso" software (v.1.0.10) was used to analyze the status of base editing. The oligonucleotide sequences for NGS are listed in Tables S1–S7.

Rice genome editing

We used the *Japonica rice* variety Zhonghua11 to prepare protoplasts. Protoplast isolation and transformation were performed as described.[7] PABE-8e and PABE-9e plasmids were co-transformed into protoplasts with sgRNA plasmids (1:1, 5 μg per construct), respectively, and three biologically independent experiments were performed for each target site. Transformed protoplasts were incubated at 26°C for 48 h and then collected for genomic DNA extraction. Genomic DNA was extracted from protoplast by the CTAB method.

Molecular modeling

Structural modeling was performed on wild-type ABE8e structure (PDB: 6VPC [25]) since the corresponding NG-ABE8e structure used in this study was not available. Single amino acid substitutions were generated *in silico* using Coot. Side-chain orientations were then geometrically optimized in Coot. Molecular graphics were prepared using MacPyMOL (Schrödinger).

Statistics

Statistical analysis and graphing were performed using GraphPad Prism 8. All data are expressed as mean ± SD. Statistical significance were determined by two-tailed Student's t test or Mann-Whitney test between two groups. The criterion for statistical significance was $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

Data availability

The data that support the findings of this study are available in the supplemental information. For sequence data, OsAAT (LOC_Os01g55540), OsACC (LOC_Os05g22940), OsALS (LOC_Os03g54790), OsCDC48 (LOC_Os03g05730), OsDEP1 (LOC_Os09g26999), OsOD (LOC_Os02g11010), and OsNRT1.1B (LOC_Os10g40600) are from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu). NGS data have been deposited in the NCBI Sequence Read Archive database (accession nos. PRJNA752490, PRJNA755472, and PRJNA752490). Pre-processed data and relevant plasmid constructs are available upon request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ymthe.2022.07.010.

ACKNOWLEDGMENTS

We thank our group members for technical assistance. This work was supported by grants from National Natural Science Foundation of China (81201181), National Key R&D Program of China (2018YFA0107304), Science and Technology Project of Zhejiang Province (2017C37176), Wenzhou city grant (H20210011), Basic research project of Henan Eye Hospital (20JCZD001), 23456 talent project of Henan Provincial People's Hospital, and by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (ZIADK075136).

AUTHOR CONTRIBUTIONS

F.G., J. Zhang, and C.G. designed this study; T.T., Z.S., X. Liu, S.W., X. He, H.X., J.W., T.Y., H.C., Z.Z., X. Lv, J.L., and J. Zhao performed research; T.T., Z.S., X. Liu, S.W., X. Huang, C.L., J. Zhao, and F.G. performed data analyses; and F.G. and J. Zhao wrote the manuscript. All authors have read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Antonarakis, S.E., Krawczak, M., and Cooper, D.N. (2000). Disease-causing mutations in the human genome. Eur. J. Pediatr. *159*, S173–S178.
2. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature *551*, 464–471.
3. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420–424.
4. Fukuda, M., Umeno, H., Nose, K., Nishitarumizu, A., Noguchi, R., and Nakagawa, H. (2017). Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. Sci. Rep. *7*, 41478.
5. Koblan, L.W., Arbab, M., Shen, M.W., Hussmann, J.A., Anzalone, A.V., Doman, J.L., et al. (2021). Efficient C*G-to-G*C base editors developed using CRISPRi screens, target-library analysis, and machine learning. Nat. Biotechnol. *39*, 1414–1425.
6. Kurt, I.C., Zhou, R., Iyer, S., Garcia, S.P., Miller, B.R., Langner, L.M., Grünewald, J., and Joung, J.K. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat. Biotechnol. *39*, 41–46.
7. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat. Biotechnol. *38*, 883–891.
8. Gaudelli, N.M., Lam, D.K., Rees, H.A., Solá-Esteves, N.M., Barrera, L.A., Born, D.A., Edwards, A., Gehrke, J.M., Lee, S.J., Liquori, A.J., et al. (2020). Directed evolution of adenine base editors with increased activity and therapeutic application. Nat. Biotechnol. *38*, 892–900.
9. Geurts, M.H., de Poel, E., Amatngalim, G.D., Oka, R., Meijers, F.M., Kruisselbrink, E., van Mourik, P., Berkers, G., de Winter-de Groot, K.M., Michel, S., et al. (2020). CRISPR-based adenine editors correct nonsense mutations in a cystic fibrosis organoid biobank. Cell Stem Cell *26*, 503–510.e7.
10. Lv, X., Qiu, K., Tu, T., He, X., Peng, Y., Ye, J., Fu, J., Deng, R., Wang, Y., Wu, J., et al. (2020). Development of a simple and quick method to assess base editing in human cells. Mol. Ther. Nucleic Acids *20*, 580–588.
11. Zhao, T., Li, Q., Zhou, C., Lv, X., Liu, H., Tu, T., Tang, N., Cheng, Y., Liu, X., Liu, C., et al. (2021). Small-molecule compounds boost genome-editing efficiency of cytosine base editor. Nucleic Acids Res. *49*, 8974–8986.
12. Rothgangl, T., Dennis, M.K., Lin, P.J.C., Oka, R., Witzigmann, D., Villiger, L., Qi, W., Hruzova, M., Kissling, L., Lenggenhager, D., et al. (2021). In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. Nat. Biotechnol. *39*, 949–957.
13. Zhang, W., Aida, T., Del Rosario, R.C.H., Wilde, J.J., Ding, C., Zhang, X., Baloch, Z., Huang, Y., Tang, Y., Li, D., et al. (2020). Multiplex precise base editing in cynomolgus monkeys. Nat. Commun. *11*, 2325.
14. Song, C.Q., Jiang, T., Richter, M., Rhym, L.H., Koblan, L.W., Zafra, M.P., Schatoff, E.M., Doman, J.L., Cao, Y., Dow, L.E., et al. (2020). Adenine base editing in an adult mouse model of tyrosinaemia. Nat. Biomed. Eng. *4*, 125–130.
15. Yeh, W.H., Shubina-Oleinik, O., Levy, J.M., Pan, B., Newby, G.A., Wornow, M., Burt, R., Chen, J.C., Holt, J.R., and Liu, D.R. (2020). In vivo base editing restores sensory transduction and transiently improves auditory function in a mouse model of recessive deafness. Sci. Transl. Med. *12*, eaay9101.
16. Koblan, L.W., Erdos, M.R., Wilson, C., Cabral, W.A., Levy, J.M., Xiong, Z.M., Tavarez, U.L., Davison, L.M., Gete, Y.G., Mao, X., et al. (2021). In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. Nature *589*, 608–614.
17. Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., Qiu, J.L., Chen, Y.H., Li, J., and Gao, C. (2020). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. Nat. Biotechnol. *38*, 875–882.
18. Fu, J., Li, Q., Liu, X., Tu, T., Lv, X., Yin, X., Lv, J., Song, Z., Qu, J., Zhang, J., et al. (2021). Human cell based directed evolution of adenine base editors with improved efficiency. Nat. Commun. *12*, 5897.
19. Jiang, T., Henderson, J.M., Coote, K., Cheng, Y., Valley, H.C., Zhang, X.O., et al. (2020). Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope. Nat. Commun. *11*, 1979.
20. Osborn, M.J., Newby, G.A., McElroy, A.N., Knipping, F., Nielsen, S.C., Riddle, M.J., Xia, L., Chen, W., Eide, C.R., Webber, B.R., et al. (2020). Base editor correction of COL7A1 in recessive dystrophic epidermolysis bullosa patient-derived fibroblasts and iPSCs. J. Invest. Dermatol. *140*, 338–347.e5.
21. Rees, H.A., Wilson, C., Doman, J.L., and Liu, D.R. (2019). Analysis and minimization of cellular RNA editing by DNA adenine base editors. Sci. Adv. *5*, eaax5717.
22. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., et al. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature *571*, 275–278.
23. Li, T., Sandberg, M.A., Pawlyk, B.S., Rosner, B., Hayes, K.C., Dryja, T.P., and Berson, E.L. (1998). Effect of vitamin A supplementation on rhodopsin mutants threonine-17–> methionine and proline-347–> serine in transgenic mice and in cell cultures. Proc. Natl. Acad. Sci. USA *95*, 11933–11938.
24. Daiger, S.P., Bowne, S.J., and Sullivan, L.S. (2014). Genes and mutations causing autosomal dominant retinitis pigmentosa. Cold Spring Harb. Perspect. Med. a017129.
25. Lapinaite, A., Knott, G.J., Palumbo, C.M., Lin-Shiao, E., Richter, M.F., Zhao, K.T., Beal, P.A., Liu, D.R., and Doudna, J.A. (2020). DNA capture by a CRISPR-Cas9-guided adenine base editor. Science *369*, 566–571.
26. Wolf, J., Gerber, A.P., and Keller, W. (2002). tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. EMBO J. *21*, 3841–3851.
27. Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. *19*, 770–788.
28. Chevance, F.F.V., Le Guyon, S., and Hughes, K.T. (2014). The effects of codon context on in vivo translation speed. PLoS Genet. *10*, e1004392.
29. Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S.V., and Gottesman, M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525–528.
30. Todorova, A., Halliger-Keller, B., Walter, M.C., Dabauvalle, M.C., Lochmüller, H., and Müller, C.R. (2003). A synonymous codon change in the LMNA gene alters mRNA splicing and causes limb girdle muscular dystrophy type 1B. J. Med. Genet. *40*, e115.
31. Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.L., Zhang, F., and Gao, C. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science *364*, 292–295.
32. Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., Lai, L., and Li, Z. (2020). Efficient base editing with high precision in rabbits using YFE-BE4max. Cell Death Dis. *11*, 36.
33. Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. *35*, 371–376.
34. Gehrke, J.M., Cervantes, O., Clement, M.K., Wu, Y., Zeng, J., Bauer, D.E., Pinello, L., and Joung, J.K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat. Biotechnol. *36*, 977–982.
35. Tan, J., Zhang, F., Karcher, D., and Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat. Commun. *10*, 439.
36. Jeong, Y.K., Lee, S., Hwang, G.H., Hong, S.A., Park, S.E., Kim, J.S., Woo, J.S., and Bae, S. (2021). Adenine base editor engineering reduces editing of bystander cytosines. Nat. Biotechnol. *39*, 1426–1433.
37. Yuan, T., Yan, N., Fei, T., Zheng, J., Meng, J., Li, N., Liu, J., Zhang, H., Xie, L., Ying, W., et al. (2021). Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods. Nat. Commun. *12*, 4902.
38. Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., Jin, Z.B., Qu, J., and Gu, F. (2014). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci. Rep. *4*, 5405.