Precise adenine base editors that exhibit minimized cytosine catalysis

You Kyeong Jeong

Hanyang University <https://orcid.org/0000-0002-8788-3605>

SeokHoon Lee

Hanyang University

Gue-Ho Hwang

Hanyang University <https://orcid.org/0000-0002-4201-0974>

Sung-Ah Hong

Hanyang University <https://orcid.org/0000-0002-3280-5677>

Se-eun Park

Hanyang University

Jin-Soo Kim

Institute for Basic Science <https://orcid.org/0000-0003-4847-1306>

Jae-Sung Woo

Korea University

Sangsu Bae (✉ sangsubae@hanyang.ac.kr)

Hanyang University <https://orcid.org/0000-0003-3615-8566>

Letter

Keywords: Adenine base editor, Protein engineering, Rational design, off-target editing, high-fidelity editing

Posted Date: October 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-84912/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Biotechnology on July 1st, 2021.
See the published version at <https://doi.org/10.1038/s41587-021-00943-2>.

Abstract

Adenine base editors (ABEs) promise specific A-to-G conversions at genomic sites of interest. However, ABEs also induce cytosine deamination at the target DNA site and exhibit transcriptome-wide off-target RNA editing. To alleviate the ABE-mediated cytosine editing activity, here we engineered the commonly-used version of adenosine deaminase, TadA7.10, to contain rationally designed mutations. We ultimately found that ABE7.10 with a D108Q mutation in TadA7.10 exhibited greatly reduced cytosine deamination activity, and conversely, ABE7.10 containing a P48R mutation displayed increased cytosine deamination activity rather than adenine editing. We found that the D108Q mutation also reduces cytosine deamination activity in two recently-developed versions of ABE, ABE8e and ABE8s, and has a synergistic effect with V106W, a key mutation that reduces off-target RNA editing. On the other hand, by incorporating the P48R mutation into ABE7.10, we demonstrated TC-specific base editing tools that enable either TC-to-TT or TC-to-TG conversions, broadening the utility of base editors.

Main Text

ABE is an effective gene editing tool that can convert an A/T pair to a G/C pair without generating a DNA double-strand break (DSB) or requiring a donor DNA template. An initial practical version of ABE (ABE7.10) was composed of three fused elements: a partially inactive Cas nuclease (Cas nickase or nCas) and a pair of adenosine deaminases [a wild-type tRNA-specific adenosine deaminase, TadA from *Escherichia coli* (wtTadA), and an engineered TadA (eTadA), TadA7.10, which was created by directed evolution so that it would operate on DNA instead of RNA]¹. Recently, several groups have reported ABE shortcomings², including genome-wide single-guide RNA (sgRNA)-dependent off-target DNA editing^{3,4}, transcriptome-wide sgRNA-independent off-target RNA editing⁵⁻⁷, and ABE-mediated cytosine deamination at the on-target site⁸. The first problem is caused by the imperfect target specificity of Cas nucleases, but the others result from the DNA/RNA-binding properties of adenosine deaminases. Therefore, to reduce the adenosine deaminase-mediated effects, further engineering of TadA7.10 is required.

To date, a few groups have provided new versions of ABE with additional mutations to alleviate its off-target effects in RNA. The Yang group determined that the addition of an F148A mutation in both wtTadA and TadA7.10 decreased the promiscuous RNA deamination activity⁷. The Liu group and the Joung group independently found that the wtTadA in ABE7.10 is mainly responsible for the RNA deamination activity, whereas the absence of wtTadA does not affect the DNA editing activity^{5,6}. Therefore, the Liu group deactivated wtTadA by introducing an E59A mutation and found that a further mutation (V106W) in TadA7.10 reduced the RNA off-target effects without reducing the DNA on-target activity⁵. The Joung group eliminated wtTadA entirely and added several mutations to TadA7.10 (K20A/R21A or V82G), resulting in a decrease of the RNA off-target effect⁶. However, the ABE-mediated cytosine catalysis effect has not been seriously addressed thus far. Because the eTadA enzyme has a common catalytic site for both adenine and cytosine, it is proposed that additional mutations in and around the active site may

decrease the cytosine conversion activity, without reducing the adenine conversion activity. Here, by rationally designing and testing tens of TadA variants, we identified key mutations that either eliminate or enhance cytosine catalysis activities.

Previously, we showed that ABE7.10 catalyzed cytosine deamination in a narrow editing window (5th~7th from the 5' end of the Cas9 target spacer) at a preferred motif (TC*N) (**Figure 1a**) and that not only ABE7.10, but its predecessors (ABE6.3, ABE7.8, and ABE7.9) and a further optimized version (ABEmax), exhibited conserved cytosine catalysis activities⁸. Here, we investigated whether cytosine deamination activities are also displayed by newly developed ABE variants, including the versions that were generated with the aim of reducing ABE-mediated RNA off-target effects (i.e., ABEmax-F148A, ABEmax-AW, and SECURE-ABEs)⁵⁻⁷, versions that contain TadA8e variants, which show increased deamination kinetics (i.e., ABE8e and ABE8e-V106W)⁹, and a version that contains TadA8s, which displays enhanced editing activity (i.e., ABE8.17-m)¹⁰. We tested all ABE variants at two representative endogenous targets (in *FANCF* and *RNF2*), which contain both an adenine residue and a cytosine target motif within the editing window, in human HEK293T cells, and made several conclusions based on high-throughput sequencing data from bulk cell populations. (**Supplementary Tables 1 and 2**) First, all tested variants still induced cytosine editing in addition to adenine editing (**Figure 1b**), but we observed slightly decreased cytosine editing rates for ABEmax-F148A and ABEmax-AW, supporting our hypothesis that further engineering of eTadA could eliminate or minimize its cytosine deamination activity. Second, although all ABE8 variants showed adenine editing rates that were greatly increased compared with the rates of previous versions, their cytosine conversion rates were also increased. Nevertheless, it is notable that ABE8e-V106W showed mitigated cytosine editing compared to ABE8e. Third, we confirmed that deactivation or elimination of wtTadA did not hinder DNA editing activities; rather, a version of ABEmax that lacks wtTadA (ABEmax-m) showed higher DNA editing activities than intact ABEmax (**Supplementary Figure 1**).

To identify key mutations in TadA that would promote discrimination between adenine and cytosine, we first examined the amino acid sequences of TadA orthologs from various species, because some of these orthologs may have already evolved to avoid cytosine editing. Based on an amino acid sequence alignment of TadA orthologs (**Figure 1c**) and the structure of *Staphylococcus aureus* TadA (saTadA) bound to a fragment of tRNA (**Figure 1d**), we found that the identities of several residues in and around the active site vary substantially among the orthologs. For example, P48 in *E. coli* wtTadA is substituted by arginine in the majority of TadA orthologs, and D108 is changed to asparagine, glutamate, or serine in other orthologs (**Figure 1c and 1d**). In addition, the saTadA structure provided insight into what structural change in the RNA substrate is required for the deamination of cytosine, which is smaller than adenine. For adenine deamination activity, the hexagonal ring of adenine should be located deep inside the adenine-binding pocket, similar to what is shown in the saTadA structure with a purine base bound to the pocket. However, for cytosine deamination, its pyrimidine ring needs to be at the same position as that of the hexagonal ring of the purine base in the structure, and this consequently requires a shift of the sugar-phosphate backbone toward the rim of the pocket. Therefore, we chose to substitute P48 and D108 with bigger residues, which may prevent the DNA backbone from approaching the pocket rim. We also

mutated V30 and F84, located in the adenine binding pocket, into isoleucine and leucine, which are found in the corresponding positions of many TadA orthologs. In addition, we introduced mutations that have been previously tested and shown to incompletely reduce RNA editing activities, such as an R47Q mutation, which maintain DNA on-target editing activities⁵, and a D53E mutation that caused reduced RNA editing activity *in vitro*^{7,11}.

We next introduced each candidate mutation into TadA7.10 in either ABEmax or ABEmax-m (**Supplementary Table 3**) and tested the nucleotide conversion activity of each resulting ABE variant at the target sites in the *FANCF* and *RNF2* genes in HEK293T cells. When we normalized the adenine and cytosine conversion rates of the ABE variants to that of ABEmax, we found that most mutations increased or decreased both the adenine and cytosine editing activities in concert (**Figure 1e**). Nevertheless, we identified four mutations (i.e., V106W, D108Q, F148A, and F149A) that substantially lowered the cytosine editing activity but maintained or slightly decreased the adenine editing activity, resulting in higher specificity for adenine editing (**Figure 1f**). Therefore, we further tested the activities of the four ABE variants at sites in two more endogenous genes and concluded that ABEmax-m containing TadA7.10-D108Q (hereafter ABEmaxQ-m) showed the highest specificity for adenine editing (**Supplementary Figure 2** and **Supplementary Table 1**). In addition, we unexpectedly found that the P48R mutation in TadA7.10 of ABEmax-m substantially reduced adenine editing rates but increased cytosine editing rates, resulting in high specificity for cytosine editing (**Figure 1e**). In summary, we identified two TadA7.10 variants, TadA7.10-D108Q and TadA7.10-P48R, which showed enhanced selectivity for adenine or cytosine editing, respectively.

Although ABEmaxQ-m exhibited greatly reduced cytosine conversion activity, its adenine conversion activity was also reduced compared to that of ABEmax. To compensate for this low editing activity, we next adapted TadA8e and TadA8.17, which were developed with the aim of enhancing editing activities^{9,10}, and tested them in place of TadA7.10. We introduced the D108Q mutation into ABE8e, ABE8e-V106W, and ABE8.17-m, thereby generating ABE8e-D108Q, ABE8e-V106W/D108Q, and ABE8.17-D108Q-m, which are referred to hereafter as ABE8eQ, ABE8eWQ, and ABE8sQ, respectively. We tested all ABE variants at a total of four endogenous sites (in *FANCF*, *RNF2*, *ABLIM3*, and *CSRNP3*). High-throughput sequencing results showed that the ABE8eQ, ABE8eWQ, and ABE8sQ variants all exhibited enhanced adenine editing activities (**Figure 2a**) with reduced cytosine editing activities (**Figure 2b**). In particular, ABE8eWQ was determined to be the most optimized version for both editing activity and specificity (**Figure 2c**), indicating that the V106W and D108Q mutations show synergistic effects (**Figure 2d**).

We next evaluated the ABE-mediated RNA off-target editing activity of all of the ABE variants. We transfected each ABE variant into HEK293T cells and measured the A-to-I conversion frequencies in four representative RNA transcripts (*CCNB1IP1*, *AARS1*, *PERP*, and *TOPRS*)^{6,7}. High-throughput sequencing results revealed that the ABE8e and ABE8.17-m versions developed earlier showed increased RNA off-target effects compared to ABEmax, and that ABE8eW showed reduced RNA off-target effects compared

to ABE8e (**Figure 2e**), as previously reported^{9,10}. To our surprise, ABE8eQ, ABE8eWQ, and ABE8sQ showed highly decreased RNA off-target effects, indicating that the D108Q mutation also effectively reduces the RNA deamination activity. D108Q could decrease the binding affinity of TadA8e for RNA, but not for DNA, because the carboxyl group of D108 forms a hydrogen bond with a 2' hydroxyl group of the bound RNA in the saTadA-RNA structure. Taken together, our results indicate that the D108Q mutation affects a key residue, decreasing both cytosine editing and RNA deamination activities, so that ABE8eQ, ABE8eWQ, and ABE8sQ are optimized versions of ABE and of these, ABE8eWQ is the best version.

Next, we sought to develop TC sequence-specific base editing tools by using the TadA7.10-P48R variant, which has increased cytosine editing activity with highly decreased adenine editing activity. To this end, we linked two copies of uracil DNA glycosylase (UGI) to the C-terminus of ABEmax-P48R as had been done in AncBE4max, an optimized cytosine base editor (CBE)¹². The addition of UGI may increase the C-to-T editing ratios rather than the C-to-G editing ratios, or vice versa. We ultimately prepared six base editing tools: AncBE4max and AncBE4max(Δ UGI) as CBEs, ABEmax and ABEmax-UGI as ABEs, and ABEmax-P48R and ABEmax-P48R-UGI as TC-specific base editors (**Figure 3a**). We tested all of them at a target site in the *CSRNP3* gene in HEK293T cells. High-throughput sequencing results showed that CBE and CBE(Δ UGI) converted all Cs (i.e., C3, C6, and C7) within the editing window with the highest rates, that ABE and ABE-UGI converted all As (i.e., A4 and A8) and C6, whereas ABE-P48R and ABE-P48R-UGI dominantly converted C6 (**Figure 3b**). We further tested these tools at three additional endogenous sites (in *FANCF*, *RNF2*, and *ABLIM3*). High-throughput sequencing results showed that editing activity tendencies were consistent with results at the *CSRNP3* site and that ABE-P48R dominantly converted C-to-G but that ABE-P48R-UGI dominantly converted C-to-T as expected (**Figure 3c**), suggesting that ABE-P48R and ABE-P48R-UGI can respectively function as specific cytosine editing tools for TC-to-TG and TC-to-TT editing without significant bystander editing effects.

To show the potential of ABE-P48R and ABE-P48R-UGI for treating genetic diseases, we inspected all targetable variations registered in the ClinVar database *in silico*. A total of 36,153 T>C mutations causing pathological phenotypes in the database are located in a canonical cytosine base editing window (4th ~ 8th) (**Figure 3d**). Among them, 3,874 mutations are associated with a cytosine target motif within the ABE-P48R-UGI editing window. In addition, 3,248 of the 23,237 G>C mutations in the database can be targeted by ABE-P48R. In the case of a missense mutation in the *TUBB6* gene (causing an F394S change in the protein) that is associated with congenital facial palsy, bilateral ptosis, and velopharyngeal dysfunction¹³, a TC sequence should be corrected to TT. We first established a cell line containing the appropriate mutation in the genome to mimic the disease situation (**Supplementary Figure 3**), after which we transfected a CBE (AncBE4max) or ABE-P48R-UGI into the cell line. High-throughput sequencing results showed that relative to ABE-P48R-UGI, CBE generated higher total cytosine conversion rates but lower rates of exact corrections (<1% compared to 4.1% for ABE-P48R-UGI) (**Figure 3e**). It is notable that ABE-P48R-UGI generated negligible bystander effects, whereas CBE generated abundant bystander effects, which suggests that ABE-P48R-UGI would be a useful TC-to-TT editing tool.

On the other hand, for a missense mutation in the *TPO* gene (causing a Q660E change) that is found in both nontoxic and toxic goiter patients¹⁴, a TC sequence should be corrected to TG. As before, we generated a cell line with the appropriate genomic mutation to mimic the disease condition (**Supplementary Figure 3**), and then transfected the CBE (AncBE4max) or ABE-P48R into the cell line. High-throughput sequencing results showed that compared to ABE-P48R, CBE generated higher total cytosine conversion rates but lower rates of exact corrections (<1% compared to 3.1% for ABE-P48R) (**Figure 3f**). It is notable that like ABE-P48R-UGI, ABE-P48R generated negligible bystander effects, whereas CBE generated abundant bystander effects, which suggests that ABE-P48R could be a useful TC-to-TG editing tool.

In this study, through rational design we determined a key mutation in eTadA, D108Q, that is responsible for decreasing cytosine catalysis activity. Interestingly, in contrast, D108F and D108W mutations dramatically decreased both adenine and cytosine editing activities (**Figure 1e**), suggesting that amino acids larger than glutamine may limit the accessibility of the sugar-phosphate backbone. D108E completely abolished the activity, probably due to strong charge repulsion between the carboxyl group of glutamate and the backbone phosphate. However, D108K caused only a mild decrease, suggesting that the lysine-phosphate interaction may not have a significant effect on the conformational dynamics of the DNA backbone. Although the size of methionine is similar to that of glutamine, D108M decreased the editing rates much more than D108Q, suggesting that direct polar interactions or water-mediated interactions between glutamine and the sugar-phosphate backbone may be crucial for the catalytic activity.

Conversely, we further identified a key eTadA mutation, P48R, that increases cytosine catalysis activity while reducing adenine editing activity. Because other mutations that affected P48 also similarly decreased the ratio of adenine editing to cytosine editing (**Figure 1e**), the cytosine specificity can likely be attributed to a change in the conformation of the main chain that consequently affects the orientations of neighboring residues. For example, a slight conformational change of N46 could have a dramatic effect on the editing activity because the residue closely contacts the purine hexagonal ring in the saTadA-RNA structure. Fortunately, in the case of P48K and P48R, the editing activity seemed to be restored by the interactions of the residue at position 48 with the backbone phosphate; the arginine-phosphate interaction might have an additional effect on the cytosine specificity by stabilizing a backbone conformation favorable to cytosine binding.

Recently, novel forms of DNA base editing such as a C>G editing¹⁵⁻¹⁷, and simultaneous C and A editing¹⁸⁻²¹ have been suggested, which would substantially expand the utilities of DNA base editors. Along with the intense efforts to improve DNA base editors, our suggested tools, high-fidelity ABE variants that exhibit minimized cytosine catalysis and reduced off-target RNA editing, and TC-specific base editors with negligible bystander effects, should make DNA base editing tools a more attractive alternative for gene editing in many research areas, such as disease therapy development, gene regulation, and plant transformation.

Online Methods

ABE variant expression plasmid construction. We constructed the ABE variant expression plasmids based on either pCMV_ABEmax_P2A_GFP (Addgene #112101), pCMV-ABEmax(TadA E59A) (Addgene #125648), pCMV-ABEmax(TadA, eTadAE59A) (Addgene #125662), pCMV-ABEmax(TadA E59A, eTadAR47Q) (Addgene #125657), pCMV-ABEmax(TadA E59A, eTadAD108Q) (Addgene #125655), pCMV-ABEmaxAW (Addgene #125647), ABE8e (Addgene #138489), ABE8e(TadA-8e V106W) (Addgene #138495), or ABE8.17-m (Addgene #136298). To obtain the vector for ABE expression, 1 µg of pCMV_ABEmax_P2A_GFP, one unit of NotI-HF (New England Biolabs, catalog number: R3189L), and one unit of BglII (Enzymomics, catalog number: R010S) were mixed together in a final volume of 50 µl. The mixture was incubated at 37°C for 1h to allow cleavage to take place. Products were separated through electrophoresis and purified using a gel extraction kit, ExpinTM Gel SV mini (GeneAll, catalog number: 102-102). The purified product and a synthesized PCR product containing the intended mutation were mixed in a volume of 10 µl containing 2U T5 exonuclease (New England Biolabs, catalog number: M0363S), 12.5U Phusion DNA Polymerase (Thermo Fisher Scientific, catalog number: F530L), 2kU Taq DNA ligase (New England Biolabs, catalog number: M0208S), 0.2 M Tris-HCl (pH 7.5), 0.2 M MgCl₂, 2 mM dNTPs, 0.2 M dithiothreitol, 25% PEG-8000, and 1 mM NAD and incubated at 50 °C for 1h. The products were then transformed into 100 µl of DH5α competent cells. Single transformed colonies were inoculated into LB medium containing antibiotics. Plasmids were isolated from cells using a DNA prep kit (Enzymomics, EP101-200N).

ABE-UGI construction through restriction enzyme-free DNA cloning using CRISPR-Cas9. Restriction enzyme-free DNA cloning using CRISPR-Cas9 was performed as described previously, with minor modifications²². The sgRNA was designed for cleaving the sequence encoding the C-terminal end of Cas9(D10A) in pCMV_ABEmax_P2A_GFP. For *in vitro* DNA cleavage, 0.7 µg of *in vitro* transcribed sgRNA and 1 µg of recombinant NG-SpCas9 were pre-incubated for 5 minutes at room temperature, after which 1 µg of pCMV_ABEmax_P2A_GFP and DEPC-treated water were added to the SpCas9-sgRNA complexes to reach the final volume of 50 µl. The mixture was incubated at 37°C for 30 min to allow cleavage to take place. Products were purified through electrophoresis. Insert PCR products were amplified from pCMV_AncBE4max (Addgene plasmid #112094). The two fragments were ligated *via* Gibson assembly.

Transfection with reagent. HEK293T (ATCC®CRL-3216™) cells were cultivated in DMEM medium supplemented with 10% FBS, 1% ampicillin in a humidified incubator at 37°C with 5% CO₂. The cell density was estimated using a hemocytometer and microscope. Before transfection, HEK293T cells (1 × 10⁵ of cells per well) were cultivated in a 24-well plate for 24 h. A mixture of 2 µL of lipofectamine® 2000 reagent (Thermo Fisher Scientific, 11668019), 1 µg of plasmid DNA (750 ng of ABE expression plasmid and 250 ng of sgRNA expression plasmid) and serum-free medium were added to the cells. Genomic DNA was isolated 72h after transfection.

Electroporation. The ABEmax expression plasmid (500 ng) and sgRNA plasmids (170 ng) were electroporated into 2 × 10⁵ cells using a Neon™ Transfection System 10 µL Kit (Thermo Fisher Scientific,

MPK1025). Appropriate electroporation parameters (1,500V-20ms-2pulses for HEK293T) were used according to the manufacturer's protocol. Genomic DNA was isolated 72h after transfection.

Sequence alignment of TadA orthologs. Protein BLAST at the National Center for Biotechnology Information were used to search for TadA orthologs. The *E. coli* wtTadA sequence (GenBank ID: WP_001297409.1) was used as an input sequence, and 10 orthologs showing ~40% sequence identities were selected. The accession number and species for the selected sequences are as follows: vsTadA, WP_127165941.1, *Veillonella* sp. CHU732; ssTadA, WP_105128341.1, *Streptococcus suis*; asTadA, WP_067866801.1, *Acinetobacter* sp. SFB; bfTadA, WP_073388705.1, *Butyrivibrio fibrisolvens*; sxTadA, WP_107541930.1, *Staphylococcus xylosus*; cbTadA, WP_111988333.1, *Cellvibrionaceae bacterium AOL6*; dfTadA, WP_117494436.1, *Dorea formicigenerans*; ocTadA, WP_047980683.1, *Ornithinibacillus contaminans*; osTadA, WP_077602817.1, *Oceanobacillus sojae*; pgTadA, WP_026908502.1, *Paucisalibacillus globulus*.

Targeted deep sequencing. HEK293T cells were pelleted by centrifugation. The cell pellet was resuspended in 100 µl of Proteinase K extraction buffer [40 mM Tris-HCl (pH 8.0), 1% Tween-20, 0.2 mM EDTA, 10 mg of Proteinase K, 0.2% Nonidet P-40] and incubated at 60 °C for 15 minutes followed by an incubation at 98 °C for 5 minutes. Genomic DNA from the HEK293T cells was amplified using KOD-Multi & Epi (TOYOBO, KME-101) and the resulting PCR products were analyzed using an Illumina Mini-Seq instrument. The Mini-seq results were analyzed using BE-Analyzer (<http://www.rgenome.net/be-analyzer/>)²³.

Targeted RNA sequencing. In a 24-well plate, HEK293T cells were transfected with 500 ng of ABE expression plasmid and 170 ng of sgRNA through electroporation. Cells were washed with DPBS 24 hours after electroporation. RNA was isolated by using a NucleoSpin[®] RNA Plus kit (MACHEREY-NAGEL, 740984. 250) according to the manufacturer's instructions. Then, for synthesizing cDNAs, reverse transcription was performed by using ReverTra Ace-α[™] (TOYOBO, FSK-101) according to the manufacturer's instructions. For amplifying cDNAs, PCR was performed with KOD-Multi & Epi (TOYOBO, KME-101). The PCR products were analyzed using an Illumina Mini-Seq instrument. To obtain the percentage of adenosines edited to inosines, the number of adenosines converted to guanosines was divided by the total number of adenosines in the products.

Data availability. High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database (SRA; <https://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA661167.

Declarations

ACKNOWLEDGEMENTS

This research was supported by grants from the National Research Foundation of Korea (NRF) (no. 2018M3A9H3022412 to S.B., no. 2020M3A9I4036072 to S.B., and no. 2018R1C1B6004447 to J.-S.W.) and the Technology Innovation Program (no. 20000158 to S.B).

AUTHOR CONTRIBUTIONS

S.B. and J.-S.W. conceived this project; Y.K.J., S.L., S.-A.H., S.-e.P. performed the experiments; Y.K.J. and G.-H.H. performed bioinformatics analyses; J.-S.K. gave critical comments; Y.K.J., J.-S.W., and S.B. wrote the manuscript with the approval of all other authors.

Additional information

Supplementary Information accompanying this paper is available at <http://>

Competing interests: Y.K.J., S.H.L., J.-S.W., and S.B. have filed a patent application based on this work.

References

1. 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).
2. Jeong, Y.K., Song, B. & Bae, S. Current Status and Challenges of DNA Base Editing Tools. *Mol Ther* **28**, 1938–1952 (2020).
3. 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).
4. Liang, P. et al. Genome-wide profiling of adenine base editor specificity by EndoV-sEq. *Nat Commun* **10**, 67 (2019).
5. 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).
6. Grunewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol* **37**, 1041–1048 (2019).
7. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
8. Kim, H.S., Jeong, Y.K., Hur, J.K., Kim, J.S. & Bae, S. Adenine base editors catalyze cytosine conversions in human cells. *Nat Biotechnol* **37**, 1145–1148 (2019).
9. Richter, M.F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* **38**, 883–891 (2020).
10. Gaudelli, N.M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* **38**, 892–900 (2020).
11. Wolf, J., Gerber, A.P. & Keller, W. tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J* **21**, 3841–3851 (2002).
12. Koblan, L.W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* **36**, 843–846 (2018).

13. Fazeli, W. et al. A TUBB6 mutation is associated with autosomal dominant non-progressive congenital facial palsy, bilateral ptosis and velopharyngeal dysfunction. *Hum Mol Genet* **26**, 4055–4066 (2017).
14. Al-Faisal, A.H., Al-Ramahi, I.J., Abudl-Hassan, I.A., Hamdan, A.T. & Barusrux, S. Detection of heterozygous c.1708C > T and c.1978C > G thyroid peroxidase (TPO) mutations in Iraqi patients with toxic and nontoxic goiter. *Comp Clin Path* **23**, 69–75 (2014).
15. Kurt, I.C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* (2020).
16. Zhao, D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* (2020).
17. Arbab, M. et al. Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning. *Cell* **182**, 463–480 e430 (2020).
18. Li, C. et al. Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat Biotechnol* **38**, 875–882 (2020).
19. Zhang, X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat Biotechnol* **38**, 856–860 (2020).
20. Sakata, R.C. et al. Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat Biotechnol* **38**, 865–869 (2020).
21. Grunewald, J. et al. A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat Biotechnol* **38**, 861–864 (2020).
22. Jeong, Y.K., Yu, J. & Bae, S. Construction of non-canonical PAM-targeting adenosine base editors by restriction enzyme-free DNA cloning using CRISPR-Cas9. *Sci Rep* **9**, 4939 (2019).
23. Hwang, G.H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* **19**, 542 (2018).

Figures

Figure 1

Screening for decreased ABE-induced cytosine deamination. **a** Schematic diagram of on-target adenine conversion and cytosine conversion mediated by the catalysis activity of an ABE. **b** Different adenosine deaminases from various adenine base editors are indicated in the diagram to the left. On the right, the graphs show the base substitution activity of each ABE at A4 and C6 in the FANCF and RNF2 sites. Residue numbering is based on the amino acid position in *E. coli* TadA. ($n \geq 3$, $\pm$ S.D.) **c** Sequence

alignment of wtTadA, TadA7.10 TadA8e, saTadA, and 10 TadA orthologs from various species. The positions corresponding to P48 and D108 of wtTadA are highlighted by blue and red boxes and labeled. d The apo structure of wtTadA from *E. coli* (green; PDB code 1Z3A) was superposed onto the holo structure of saTadA bound to RNA (pink and gray; PDB code 2B3J). The catalytically important zinc ion in the *E. coli* wtTadA structure is represented as a gray ball and labelled. e Analysis of adenine conversion and cytosine conversion induced by 33 ABE variants at the FANCF and RNF2 sites. The combinatorial mutations F84L/D108Q/F149A and V88T/L117W/M118K/V120T are abbreviated as LQA and TWKT, respectively. Efficiencies of conversions are shown in a heat map format ($\log_2(\text{fold change})$), with each box representing the mean of independent replicates normalized to the editing efficiency of ABEmax. f Specificities of adenine conversion. Each dot indicates the adenine conversion efficiency compared to the cytosine conversion efficiency for an independent replicate. Variants that contain V106W, D108Q, F148A, or F149A mutations exhibit the highest specificity for adenine editing and are highlighted with yellow boxes. A variant containing P48R conversely displayed the highest specificity for cytosine editing and is highlighted with a pale red box.

Figure 2

D108Q is a key mutation for reducing the frequency of cytosine deamination. **a** Analysis of the adenine substitution efficiency of nine ABE variants at four endogenous sites, in *FANCF*, *RNF2*, *ABLIM3*, and *CSRNP3*. All four sites contain adenine in position 4, counted from the 5' end of the Cas9 target spacer. ABE8eQ, ABE8eWQ, and ABE8sQ are abbreviations for ABE8e-D108Q, ABE8e-V106W/D108Q, and ABE8.17-D108Q-m, respectively. ($n \geq 3$, $\pm$ S.D.) **b** Analysis of the cytosine substitution efficiency of nine ABE variants at the same sites tested in **a**. The four endogenous sites have a cytosine target motif within the editing window. ($n \geq 3$, $\pm$ S.D.) **c** The adenine substitution rate relative to the cytosine substitution rate at all four endogenous sites, represented as box-and-whisker plots. The center line indicates the median, the X mark indicates the mean, the box represents the first and third quantiles, and the whiskers indicate

the maximum and minimum values. (n ≥ 12) d Schematic diagrams of the cytosine target motif binding region in eTadA. The key mutation, D108Q, affects the binding affinity of eTadA to the cytosine target motif so that only adenine, but not cytosine, is converted by ABE containing the D108Q mutation. e The average A-to-I conversion frequencies in four mRNA transcripts after treatment with each of the nine ABE variants. The values were determined by high-throughput sequencing.

Figure 3

The P48R mutation promotes specific cytosine conversion with low adenine editing activity. a Schematic diagrams of six CBEs and ABEs. “N” and “C” indicate the amino- and carboxyl-terminal ends, respectively. Each component is connected with appropriate linkers. b The substitution rates at each base in the CSRNP3 target site are indicated as a heat map. The numbers indicate the position counting from the 5' end of the Cas9 target spacer. Cytosine target motifs and adenines within the editing window are highlighted with brown and pink boxes, respectively. (Mean, n = 3) c The proportions of cytosines that were converted to other bases are indicated as stacked columns. To obtain these data, HEK293T cells were treated with six base editors targeted to four endogenous sites, in FANCF, RNF2, ABLIM3, and CSRNP3, which contain cytosine target motifs within the editing windows. (n = 3, ± S.D.) d Pie graphs indicating the proportion of pathogenic human mutations in the ClinVar database that are targetable by ABE compared with those targetable by canonical CBEs. CBE targetable T>C mutations are indicated in the upper graph and G>C mutations are indicated in the lower graph. Among the mutations, ABE targetable sites are counted according to the cytosine target motif and editing window. e and f Correcting pathogenic mutations in TUBB6 and TPO. The target sequence in the TUBB6 gene contains a pathogenic T>C mutation (e). The target sequence in the TPO gene contains a pathogenic G>C mutation (f). In each case, the pathogenic mutation is contained in a cytosine target motif that is highlighted with a purple box. Two types of base editor-induced substitutions are possible in the target sites, which depend on the choice of base editor. The major base editor-induced substitution patterns are shown in the upper schematic diagrams and the base editor-induced substitution outcomes are shown in the lower charts

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)