Upgraded adenine base editor (uABE) with minimized RNA off-targeting activity

Jinan Li  
ShanghaiTech University

Wenxia Yu  
ShanghaiTech University

Shisheng Huang  
University of Shanghai for Science and Technology

Susu Wu  
Guangzhou University

Liping Li  
Guangzhou University

Jiankui Zhou  
Guangzhou University  https://orcid.org/0000-0002-7099-4528

Yu Cao  
Shanghai Jiao Tong University School of Medicine  https://orcid.org/0000-0003-1409-2068

Xingxu Huang  (huangxx@shanghaitech.edu.cn)  
ShanghaiTech University

Yunbo Qiao  (ybqiao@gzhu.edu.cn)  
Guangzhou University

Article

Keywords: Adenine base editor; off-target; RNA editing; ABEmax; upgraded ABE.

DOI: https://doi.org/10.21203/rs.3.rs-73017/v1

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

Both adenine base editors (ABEs) and cytosine base editors (CBEs) have been recently revealed to induce transcriptome-wide RNA off-target editing in a guide RNA-independent manner. As the optimized ABE, ABEmax, induces highly efficient A-to-I (inosine) editing within an *E. coli* tRNA-like structure, we construct a reporter system containing *E. coli* Hokb gene with a tRNA-like motif for robust detection of RNA editing activities. Then, we design mutations to disrupt the interaction between TadA and tRNAs in structure-guided principles, and find that Arginine 153 (R153) within TadA is essential for recognizing core tRNA-like structures. Two ABEmax or mini ABEmax variants (TadA* fused with Cas9 (D10A)) with deletion of R153 within TadA and/or TadA* (named as del153/del153* and mini del153) are successfully engineered, showing minimized RNA editing, but comparable DNA on-targeting activities. Moreover, del153 in recently reported ABE8e or ABE8s can also largely reduce their RNA off-targeting activities. Taken together, we develop a strategy to generate upgraded ABEs (uABEs) with minimized RNA off-target activities.

Introduction

Adenine base editors (ABEs), which is originally designed by fusing a wild-type *E. coli* TadA (ecTadA) and/or a laboratory-evolved *E. coli* TadA (TadA*) with a Cas9 (D10A) nickase (Cas9n), can induce efficient A-to-G or T-to-C conversions with very low levels of unwanted mutations or insertions [1, 2]. ABE is designed based on the native structure of homodimerized ecTadA, which can deaminate an adenosine within a transfer RNA (tRNA) [3], with an evolved TadA* being capable of deaminating genomic DNA adenosines [1]. Although ABEs show no detectable guide RNA-independent DNA off-target editing [4, 5], both ABEs and cytosine base editors (CBEs) can induce tens of thousands of A-to-I (inosine) or C-to-U (uracil) RNA edits transcriptome-widely in guide RNA-independent manners in human cells [6, 7]. Engineered CBE and ABE variants bearing rAPOBEC1 mutations [6] or TadA/TadA* mutations[7–9], have been recently reported with reduced RNA off-targeting activities. In these studies [6, 8, 9], GATK HaplotypeCaller, a tool for evaluating germline single nucleotide polymorphisms (SNPs) and indels [10], is employed for analyzing RNA A-to-I edits. It is worth noticing that RNA edits with 0–10% efficiency was not able to be recovered by this tool [6, 8, 9], suggesting a possible underestimation of RNA off-targets, therefore, driving us to further examine the ABE-induced off-target editing of cellular RNAs in depth. According to structure-guided principles, we successfully engineered ABEmax and miniABEmax variants to generate upgraded ABEs (uABEs) that retained DNA on-target editing efficiency while largely decreased RNA-editing activities.

Results

Engineering ABEmax with reduced RNA deaminase activities
Considering that cellular RNAs with wide range of aneuploidy copies have been discovered as RNA off-target substrates of ABEs [6–9], we reasoned that Mutect2, a GATK tool for sensitive detection of somatic point mutations in heterogeneous cancer samples [11], might be more appropriate than HaplotypeCaller (for detection of euploid germline SNPs) for detection of RNA edits [12] (Supplementary Fig. 1a). Thus, we reanalyzed ABEs- and their optimized variants (miniABEmax-V82G, AB7.10-F148A, and ABEmaxAW)-induced RNA off-targets [6–9], and found that indeed, Mutect2 recovered 2.7-11-fold number of ABEs-induced RNA edits compared with that using HaplotypeCaller with similar editing signatures (endogenous A-to-I edits were deducted from control sequencing data), revealing these optimized ABE variants still retained a relatively large number of RNA edits (Supplementary Fig. 1b-g). Surprisingly, the overlapped RNA edits from HaplotypeCaller and Mutect2 were as few as 22–68% of HaplotypeCaller-calculated RNA off-targets (Supplementary Fig. 1h-m). We further performed Manhattan plotting for ABEmax-induced RNA off-targets [6] to show efficiency distributions of overlapped, HaplotypeCaller-specific, and Mutect2-specific RNA edits, respectively, demonstrating that the number of Mutect2-specific RNA edits was much more than HaplotypeCaller-specific edits, especially for those edits with 0–10% editing efficiency, which were ignored by HaplotypeCaller (Supplementary Fig. 1l). Meanwhile, lower overlapping ratio was discovered for the samples possessing fewer RNA edits (Supplementary Fig. 1m). Nine MuTect2-specific edits (with > 10% efficiency in RNA-seq data) were randomly selected for PCR validation with cDNAs subjected to RNA-seq experiment. Indeed, all of these amplicons were successfully detected with high or low efficiency of A-to-G mutations (Supplementary Fig. 1n), confirming the reliability of MuTect2-specific edits and suggesting that it is necessary to engineer ABEmax variants based on MuTect2 analysis.

Although an engineered TadA* has been evolved to be capable of deaminating DNA adenines [1, 2], both TadA* and wildtype TadA retain the ability to deaminate cellular RNAs [9]. We analyzed ABEmax-induced RNA edits in ABEmax-overexpressed HEK293T cells from a published RNA-seq dataset (Supplementary Fig. 2a-b) [6]. Meanwhile, we generated our data by co-transfecting HEK293T cells with a sgRNA to efficiently induce DNA A-to-G conversion within ABE site 8 (Supplementary Fig. 2c-e). The cells with highest 15% GFP signal were collected for on-target and off-target analysis. Higher overlapping ratios between two independent duplicates were observed for those A positions of RNA edits with higher editing efficiency, demonstrating the preferential affinity of ABEmax for highly edited RNAs (Supplementary Fig. 2a and 2d). Thus, we calculated the sequence logos for ABEmax-induced RNA edits with differential scope of editing efficiency, showing that higher-edited adenines preferentially located within a conserved motif being more close to UACGA (Supplementary Fig. 2b and 2e), which highly resembles the conserved loop region of tRNA substrate for ecTadA [3]. These data demonstrate that, consistent with a recent report [9], ABEmax induces efficient transcriptome-wide off-target RNA editing harboring core E.coli tRNA-like sequences.

Therefore, we hypothesized that disruption of the interaction between TadA/TadA* heterodimer and tRNA-loop structure may interfere the catalytic activities of ABEmax on RNA. Since there is no crystal structure information for the complexing between ecTadA and tRNA, we referred the co-crystal structure of Staphylococcus aureus TadA (saTadA) and tRNA as well as the alignment of the conserved amino acid sequences between ecTadA and saTadA with high similarity [3, 13], showing that the amino acids...
possibly responsible for interaction with tRNA are conserved between the two types of TadA (Fig. 1a and Supplementary Fig. 3a). Thus, we introduced a series of point mutations into either the TadA or TadA* monomer of ABEmax according to the interacting interface between homodimerized TadA and tRNA [3] to disrupt TadA/TadA* and tRNA interactions [3], and measured their RNA and DNA editing activities (Fig. 1b). To facilitate this test, we generated a robust reporter by cloning the *E.coli Hokb* (ecHokb) gene containing tRNA-like CTACGAA sequence, which has been reported to be highly edited by ecTadA at RNA levels [14], into a CMV promoter-driven vector. Then, this reporter was co-transfected with a sgRNA targeting HEK site 3 and ABEmax or its mutated variants, and the A-to-G editing efficiencies in ecHokb cDNA (reversely transcribed from mRNA) or genomic DNA (gDNA) were determined by deep sequencing on ecHokb cDNA or gDNA amplicons. It showed that both ABEmax and 2xTadA induced highly efficient RNA but not DNA editing within ecHokb locus. Notably, we identified three variants (N46A, H57A, and R153P) with substantially decreased RNA editing activities, especially R153P with most reduced RNA edits comparable to the negative Cas9n control (Fig. 1c and Supplementary Fig. 3b; the endogenous RNA A-to-I edits detected in native HEK293T cells were deducted). In addition, their DNA on-target editing activities were retained (Fig. 1d). Moreover, similar to ABEmax [2], all variants induced very few by-products and indels (Supplementary Fig. 3c-d). Three amino acids, including N46, R153, and the reported site E59 [6], were likely in close contact with tRNA near the enzymatic pocket in structural prediction (Supplementary Fig. 3e). Additionally, ABEmax-R153P variant exhibited comparable DNA on-target A-to-G editing activities for multiple target sites in human cells (HEK293T and U2OS cells) (Supplementary Fig. 4a-b). Thus, we identified three variants, especially ABEmax-R153P, with minimized RNA editing activities in the reporter assay.

Next, we performed RNA-seq analysis to detect transcriptome-wide RNA off-targets induced by ABEmax or R153 substituted variants (R153P, R153P/R153P*, and R153A/R153A*) (Supplementary Fig. 4c). Surprisingly, the total number of RNA off-targets induced by R153P/R153P*, and R153A/R153A* were comparable with ABEmax (Fig. 1e by MuTect2; a bit lower by HaplotypeCaller in Supplementary Fig. 4d), and R153P variant induced even more RNA edits, which was excluded in further test (Supplementary Fig. 4e). Actually, R153P/R153P* and R153A/R153A* was test after the observation of high RNA A-to-I edits induced by R153P variant. Sequence logos derived from highly edited adenines (> 20%) showed that the preference of ABEmax for a consensus TACG motif was diminished upon R153 substitution (Fig. 1f) and the number of RNA edits with > 40% efficiency for R153P/R153P* and R153A/R153A* was much fewer than ABEmax (Fig. 1e and Supplementary Fig. 4f), suggesting that R153P/R153P* and R153A/R153A* mainly affect highly edited RNA edits harboring conserved UACG motif. These data demonstrate that interfering R153 can minimize the RNA-editing catalytic activities of TadA on tRNA loop-like structures (Fig. 1e).

**Deletion of Arginine 153 (del153) reduces RNA off-targeting activities in upgraded ABEs**

Considering the importance of R153 for deamination (Fig. 1), we tried to generate upgraded ABEs with reduced RNA off-targets by deleting R153 from both TadA and/or TadA* within ABEmax (del153/del153*) or mini ABEmax (fused by TadA* and Cas9n [9] (mini del153). As expected, we demonstrated that
compared with ABEmax or mini ABEmax, the RNA off-targets induced by del153/del153* and mini del153 were largely decreased, and there were as few as 291 (MuTect2) or 98 (HaplotypeCaller) RNA A-to-I edits for mini del153 group (Fig. 2a-c; Supplementary Fig. 5a), while both variants retained a relatively high DNA on-targeting activities (Fig. 2a). We also overlapped or merged the ABEmax-, del153/del153*-, or mini del153-induced RNA A-to-I edits using HaplotypeCaller and MuTect2, respectively. Compared with ABEmax, both del153/del153* and mini del153 induced remarkably decreased RNA edits of the overlapped, HaplotypeCaller-specific, MuTect2-specific, and merged edits (Supplementary Fig. 5b). Manhattan plots and histograms further confirmed that both the number and efficiency for del153/del153*- and mini del153-induced RNA A-to-I edits were strikingly decreased (Fig. 2b-c), accompanying with much lower mean frequencies throughout the transcriptome (Supplementary Fig. 5c).

Next, we compared our upgraded ABE (uABE) variants with reported variants possessing reduced RNA off-targeting activities, including ABEmaxAW (E59A/V106W*) [7], ABE-F148A/F148A* [8], and SECURE-ABEs (miniABEmax-K20A/R21A and miniABEmax-V82G (mini V82G)) [9]. The results showed that our upgraded variants, del153/del153* and mini del153 induced much fewer RNA off-targets by both MuTect2 and HaplotypeCaller tools under same experimental conditions (Fig. 2a; Supplementary Fig. 5a) (with much fewer RNA edits by using HaplotypeCaller).

We further characterize the DNA on-target editing activities of del153/del153* and mini del153 variants for another eight sites. It demonstrated that the DNA on-targeting activities of del153/del153* and mini del153 were highly similar to ABEmax for nearly all of detected sites, except for ABE site 12 with a bit lower but satisfactory editing efficiency (Fig. 2d; Supplementary Fig. 6a). Similar to SECURE-BE3-induced RNA C-to-U edits with perfect reducing effect [6], del153/del153* and mini del153 variants-induced RNA A-to-I edits were decreased to only dozens or hundreds of off-targets when using HaplotypeCaller (Supplementary Fig. 5a). Comprehensively considering the on-targeting and off-targeting activities of engineered variants, we designate del153/del153* and mini del153 as our best optimized ABE variants with minimized RNA editing activity. Recently reported ABE8e and ABE8s containing evolved mutations within TadA/TadA* possess increased DNA on-targeting activities as well as elevated RNA off-targeting activities [15, 16]. We tried to generate upgraded ABes with higher DNA on-targeting and lower RNA off-targeting activity by deletion of R153 from ABE8e or ABE8s, demonstrating that RNA off-targets were remarkably decreased from HaplotypeCaller or MuTect2 calculations, and the number of ABE8s-del153 exhibited comparable number of RNA edits with mini ABEmax (Fig. 2e and Supplementary Fig. 6b). Notably, ABE8s-del153 and ABE8e-del153 showed comparable or slight lower levels of DNA A-to-G editing activities; however, the on-targeting activity of f ABE8e/8 s or ABE8e/8 s-del153 was much higher than ABEmax or mini ABEmax, and the editing window of ABE8e/8 s or ABE8e/8 s-del153 was also much wider than ABEmax or mini ABEmax (Fig. 2f; Supplementary Fig. 6c). In collection, we propose that deletion of R153 is a good strategy for reducing RNA off-targeting activities in upgraded ABes.

**Discussion**
Our description of ABEmax-induced transcriptome-wide RNA off-targeting with high frequency and efficiency confirmed recent studies [7–9], although the number of RNA A-to-I edits was variable, possibly because of differential expression of ABEs [8] (Supplementary Fig. 6) and detection methods. When we noticed the rare distribution of RNA edits with 0–10% efficiency using HaplotypeCaller [6, 8, 9], we started using MuTect2, a widely used tool for calling somatic mutations in cancers [11, 17], which might be more suitable for analyzing SNPs within aneuploid mRNAs. Surprisingly, we identified 2.7-11-fold number of ABEs-induced RNA edits, only 22–68% of which was overlapped with HaplotypeCaller-generated edits (Supplementary Fig. 1). We conclude that the number of BEs-induced RNA off-targets is underscored, especially for those with < 10% editing efficiency, which may result in some poisoned or oncogenic proteins in therapeutic cases [18]. The sequence logos analysis suggests that TadA/TadA* preferentially edit cellular RNAs with UACGA motif, not regarding the secondary structure of RNAs. In addition, there are also a large number of HaplotypeCaller-recovered RNA A-to-I edits that cannot be captured by MuTect2, some of which can be validated by Sanger sequencing (data not shown). Detection performances of different tools to call RNA mutations highly depend on the sequencing depths, detected regions, and variant allele frequencies [12], which may lead to differential results from different tool. Therefore, barely using MuTect2 is not the best way to examine RNA A-to-I edits, and developing a new tool, such combining HaplotypeCaller and MuTect2 [11, 19], for more accurate evaluation of RNA off-targeting effect will be quite helpful [20].

Based on a structure-guided design [3] to disrupt the interaction between TadA/TadA* and tRNA-like mRNAs with conserved UACGA motif [9], we successfully identify R153 as an important amino acid for deaminase activity of TadA/TadA*, supported by R153A/R153A* variants-induced lower efficiency of RNA A-to-I edits. Interestingly, the RNA editing efficiency for our reporter ecHokb and those efficiently edited RNAs with tRNA loop-like structures by ABEmax was decreased upon R153 mutation, whereas the total number of RNA edits was not significantly changed (Fig. 1). It indicates that R153 might be required for TadA to specifically bind to t-RNA loop like RNAs, and deletion of R153 within TadA/TadA* in del153/del153* and mini del153 variants strikingly reduces the number of RNA off-targets, with high DNA on-targeting activity retained (Fig. 2), further confirming the reasonability of our strategy. However, mutation of R153 into “P” or “A” may retain its structural interaction with RNAs, while mutation into an acidic amino acid “E” may disrupt this interaction, because R153E mutant possesses very low on-targeting activity (data not shown). Moreover, our del153/del153* and mini del153 variants show better optimizing effects than the reported versions under our experimental conditions. When comparing with the perfect reducing effect of SUCURE-BE3-induced RNA C-to-U edits [6], del153/del153*, and mini del153 variants-induced RNA A-to-I edits are decreased to only dozens or hundreds of off-targets when using HaplotypeCaller. Considering a slight lower DNA on-targeting efficiency for mini del153 occasionally, del153/del153* is priorly recommended for targets with low targeting efficiency. We also combine del153 strategy with evolved ABEs, ABE8e and ABE8s [15, 16], demonstrate that deletion of R153, the residue nearby some of mutated acids in ABE8e/ABE8s [15, 16], can also remarkably reduce the number of RNA edits but retain their on-targeting activities in most cases (Fig. 2). Therefore, ABE8e-del153 and ABE8s-del153 are suitable for desiring higher DNA on-targeting and lower RNA off-targeting activities. Besides, it
has been reported that BEs-induced RNA off-target editing acts in a sgRNA-independent manner [6, 9], and we do not consider the sgRNA-dependent effects in the current study. While these findings remind us to reconsider the off-targeting activities of our and others’ reported dCas9-fused epigenome editing tools [21, 22].

In sum, we reveal R153 of TadA/TadA* as an essential amino acid for its RNA interacting ability, and we successfully optimize ABEs by deletion of R153 from TadA/TadA* to generate upgraded ABE (uABE) variants, which greatly reduce the number of RNA edits while retain high DNA on-targeting activity. The successful engineering of CBEs and ABEs variants in our and other two studies [6–9] expands our understanding of desired and undesired features of DNA and RNA editing activities of base editors, and provides a feasible pathway available to engineer base editors based on structural analysis to minimize the unwanted properties while retaining the desired on-target editing ability for CBEs or ABEs.

**Materials And Methods**

**Plasmid construction**

Briefly, the synthesized DNA oligos for sgRNA-expressing plasmid construction were annealed and cloned into pGL3-U6-sgRNA-PGK-EGFP with U6 promoter (Addgene #107721). Oligos are showed in Supplementary Table 1. Editors were constructed by insertion of amplified DNA product into linearized ABEmax (NdeI/Bg1II digest of pCMV_ABEmax, Addgene #112095). DNA products were amplified by Phanta Max Super-Fidelity DNA Polymerase (Vazyme, P505) using mutant site primers (such as TadA-N46A-F/R) and two fragments primer sets (CMX-NdeI-F and nCas9-Bg1II-R). Primers are showed in Supplementary Table 2.

**Cell culture and transfection**

HEK293T and U2OS cells were purchased from ATCC and cultured in DMEM (10566, Gibco/Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS) (v/v) (Gemini, 900-108) and 1% Penicillin Streptomycin at 37°C with 5% CO2.

For deep sequencing samples, HEK293T cells were seeded on 24-well plates (JETBIOFIL) and transfected at approximately 70% confluence with editors (628 ng) and sgRNAs (373 ng) using Lipofectamine LTX (ThermoFisher Scientific, 15338100) according to the manufacturer’s protocol. GFP positive cells were harvested from fluorescence-activated cell sorting (FACS) 48 h after transfection.

For RNA sequencing samples, HEK293T cells were seeded on 6 cm dish (JETBIOFIL) and transfected at approximately 70% confluence with editors (4 µg) and sgRNA-expressing plasmids (2 µg) using Lipofectamine LTX (ThermoFisher Scientific, 15338100) according to the manufacturer’s protocol. GFP signal positive cells of top 15% were harvested from fluorescence-activated cell sorting (FACS) 48 h after transfection.

**RNA and genomic DNA extraction**
Genomic DNA of HEK293T and U2OS cells was extracted using phenol-chloroform method, and embryos genomic DNA was extracted using QuickExtract™ DNA Extraction Solution (Lucigen, QE09050). For RNA extraction, cells harvested from FACS were immediately treated with TRIzol reagent (Vazyme, R401-01), according to the manufacturer’s instructions.

**Targeted deep sequencing**

Target sites were amplified with primers in Supplementary Table 3 using Phanta® Max Super-Fidelity DNA Polymerase (Vazyme, P505). PCR products with different barcodes were pooled together for deep sequencing on Illumina Nextseq 500 (2×150 PE) platform at the Novogene Bioinformatics Institute, Beijing, China. BWA and Samtools was employed for mapping the pair-end reads to human reference genome (hg38), and VarDict was used to call single nucleotide variants (SNVs) in amplicon aware mode. The aligned reads were visualized by using the Integrated Genome Viewer (IGV) and tabbed using Pysamstats.

**RNA off-target analysis by RNA-seq**

The libraries were sequenced on an Illumina HiseqXten-PE150, at a depth of around 20 million reads per sample. The reads were mapped to the human reference genome (hg38) by STAR software (Version 2.5.1), annotation from GENCODE version v30 was used. After removing duplication, variants were identified by GATK (version 4.1.2) Mutect2 or HaplotypeCaller. For Mutect2 method, variants were filtered with FilterMutectCalls. For HaplotypeCaller method, variants were first filtered with QD (Quality by Depth) < 2, then all variants were verified and quantified by bam-readcount with parameters -q 20 -b 30. The depth for a given edit should be at least 10x and these edits were required to have at least 99% of reads supporting the reference allele in the wild-type samples. Finally, only A-to-G edits in transcribed strand were considered for downstream analysis. Motif or sequence logo was analyzed by WebLogo (v3.6.0) for RNA edits. The downloaded data subjected to RNA off-target analysis from four published papers were listed in Source Data for Sup Figures. Detailed information for called mutations was provided in Source Data for called mutations from RNA-seq data.

**Structural analysis**

A structural model for TadA-RNA complex was generated using coordinates from PDB ID 2B3J by PyMol (The PyMOL Molecular Graphics System, Version 1.9 Schrödinger, LLC.). TadA from Staphylococcus aureus (SaTadA) was shown as cartoon model in gray and the RNA bound was shown as stick model rendered by elements, with the Zn\(^{2+}\) ion as green sphere. The residues critical for the RNA binding of TadA were shown in ball-and-stick model and labelled with single-letter codes in red.

**Statistics**

Results were obtained from three independent or indicated experiments and were presented as the mean ± SD. For analyzing relative expression in FPKM, the FPKM from RNA-seq data in 3 embryos was used
and presented as the mean ± s.d.. All original data presented in main figures were provided in Source Data for Main Figures, and original data presented in supplementary figures were listed in Source Data for Sup Figures. Statistical analyses and graphing were carried out using GraphPad Prism 7.0. Comparisons of mean values were analyzed by Student's t-test.

**Data availability**

All sequencing data was provided in SRA (accession number: PRJNA660634). All other results in the present study are available upon request.

**Declarations**

**Authors’ contributions**

J.L., W.Y., and Y.Q. designed, conceived, and performed the experiments and co-wrote the manuscript; S.H. and Y.C. performed computational analysis; S.W., M.L., and J.Z. helped with the data analysis. Y.Q. and X.H. designed, conceived, and supervised the work and co-wrote the manuscript.

**Acknowledgements**

This work was supported in part by National Key Research and Development Program (2018YFC1004700, Y.Q.; 2016YFA0500903, X.H.), Local Grants (17JC1420103, X.H.; Excellent Youth Foundation of Guangdong Scientific Committee, 2020B1515020018, Y.Q.), National Natural Science Foundation of China (81830004).

**Competing interests**

The authors declare that they have no competing interests.

**References**

1. Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR: Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 2017, 551:464–471.
2. Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A, Liu DR: Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat Biotechnol* 2018, 36:843–846.
3. Losey HC, Ruthenburg AJ, Verdine GL: Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* 2006, 13:153–159.
4. Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu JL, Zhang F, Gao C: Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 2019, 364:292–295.
5. Zuo E, Sun Y, Wei W, Yuan T, Ying W, Sun H, Yuan L, Steinmetz LM, Li Y, Yang H: Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 2019, **364**:289–292.

6. Grunewald J, Zhou R, Garcia SP, Iyer S, Lareau CA, Aryee MJ, Joung JK: Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* 2019, **569**:433–437.

7. Rees HA, Wilson C, Doman JL, Liu DR: Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci Adv* 2019, **5**:eaax5717.

8. Zhou C, Sun Y, Yan R, Liu Y, Zuo E, Gu C, Han L, Wei Y, Hu X, Zeng R, et al: Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* 2019, **571**:275–278.

9. Grunewald J, Zhou R, Iyer S, Lareau CA, Garcia SP, Aryee MJ, Joung JK: CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat Biotechnol* 2019, **37**:1041–1048.

10. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011, **43**:491–498.

11. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G: Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013, **31**:213–219.

12. Liu F, Zhang Y, Zhang L, Li Z, Fang Q, Gao R, Zeng Z: Systematic comparative analysis of single-nucleotide variant detection methods from single-cell RNA sequencing data. *Genome Biol* 2019, **20**:242.

13. Wolf J, Gerber AP, Keller W: *tadA*, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J* 2002, **21**:3841–3851.

14. Bar-Yaacov D, Mordret E, Towers R, Biniashvili T, Soyris C, Schwartz S, Dahan O, Pilpel Y: RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res* 2017, **27**:1696–1703.

15. Gaudelli NM, Lam DK, Rees HA, Sola-Esteves NM, Barrera LA, Born DA, Edwards A, Gehrke JM, Lee SJ, Liquori AJ, et al: Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat Biotechnol* 2020, **38**:892–900.

16. Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, et al: Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol* 2020, **38**:883–891.

17. Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, Bare JC, P’ng C, Waggott D, Sabelnykova VY, et al: Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat Methods* 2015, **12**:623–630.

18. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, et al: Patterns of somatic mutation in human cancer genomes. *Nature* 2007, **446**:153–158.
19. Boutros PC, Ewing AD, Ellrott K, Norman TC, Dang KK, Hu Y, Kellen MR, Suver C, Bare JC, Stein LD, et al: Global optimization of somatic variant identification in cancer genomes with a global community challenge. *Nat Genet* 2014, **46**:318–319.

20. Neums L, Suenaga S, Beyerlein P, Anders S, Koestler D, Mariani A, Chien J: VaDiR: an integrated approach to Variant Detection in RNA. *Gigascience* 2018, **7**.

21. Liu J. LL, Zhou C., Meng B., Wei Y., Yang G., Lu Z., Shen Q., Zhang Y., Yang H., Qiao Y.: Allele-specific genome editing of imprinting genes by preferentially targeting non-methylated loci using *Staphylococcus aureus* Cas9 (SaCas9). *Science Bulletin* 2019, **64**:1592–1600.

22. Pulecio J, Verma N, Mejia-Ramirez E, Huangfu D, Raya A: CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* 2017, **21**:431–447.

**Figures**
Figure 1

Arginine 153 within TadA/TadA* is essential for specificity of ABEmax-induced RNA deamination. (a) Summary of predicted E. coli TadA-tRNA interactions. A classical tRNAArg2 possessing CUACGAA and the potential interaction between the residues from TadA (purple) and tRNA loop nucleotides was shown. (b) Design for the engineered ABEmax (for the 1st round of engineering) with indicated amino acid substitutions. (c) Schematic overview of reporter assays for RNA editing. The upper diagram represents a
gene Hokb cloned from E. coli that is driven by a CMV promoter (pCMV). A sgRNA targeting HEK-site 3 (sgHEK site 3) was constructed into a pGL3-U6 vector. The lower bar plot showed the percentage of A-to-G (A-to-G for genomic DNA test; A-to-I for RNA-generated cDNA test) conversion for Hokb gene in the reporter, which was co-transfected with sgHEK-site 3 and ABEmax variants into HEK293T cells. After transfection for 72 h, cells were harvested for genomic DNA and RNA extraction and reverse transcription, followed by PCR amplification and deep-sequencing. A, adenine; I, inosine; ABEmax, codon optimized adenine base editor. (d) The genomic DNA and RNA-generated cDNA were subjected to PCR amplification of fragments containing HEK site 3 targeting sites for ABEmax variants by amplicon deep-sequencing. The A-to-G editing efficiency on adenine 7 (A7) within the sgRNA sequence was shown. (e) The number (upper, by MuTect2) and efficiency distribution (lower in Jitter plot) of RNA A-to-I edits were presented for RNA-seq experiments in HEK293T cells that expressed Cas9n (rep1 and rep2), ABEmax (rep1 and rep2), ABEmax (R153P/R153P*; rep1 and rep2), or ABEmax (R153A/R153A*; rep1 and rep2) and sgRNA targeting HEK site 8. Two independent replicates were shown. Each dot in the lower panel represents an edited adenine position in RNA. Non-transfected HEK293T transcriptome server as control, and the RNA A-to-I edits overlapping with endogenous A-to-I edits were excluded. (f) Sequence logos centered by the edited adenine (A) from a single replicate of RNA-seq data in e (rep1 for each group) for edited adenines with indicated editing efficiencies (>40%, >20%, and all). The nucleotides around the edited A were shown (9 nucleotides in total). The changed sequence logos for ABEmax variants were highlighted in gray. (g) Bar plot showing the on-target DNA A-to-G editing efficiencies of ABEmax, ABEmax (R153P/R153P*), and ABEmax (R153A/R153A*) with 3 gRNAs (ABE site 1-3). The constructs for these three ABEmax variants were presented in Supplementary Fig. 4c. Data was generated from two independent replicates from deep-sequencing analysis. The efficiencies for most highly edited adenines for each sgRNA on-target site within the editing window were reported. Error bars represent s.d.
Figure 2

Deletion of R153 from TadA/TadA* results in minimized RNA off-targeting activities. (a) Left panel: design for the second round of engineered ABEmax variants with R153 deletions (del). * represents amino acid substitutions or deletions constructed in TadA*. Mini ABEmax was fused by TadA* and Cas9n. R153 was deleted from TadA* to generate mini del153. Middle panel: heat maps showing the DNA on-target A-to-G editing efficiency in HEK293T cells co-expressing ABEmax or its variants (as shown in left panel).
with sgRNAs targeting ABE site 1, 2 and 3 respectively. Efficiencies were presented in heat map as log2(fold change) for each site; each box representing the DNA A-to-G editing efficiency normalized to the that observed with ABEmax for each on-target DNA sites (ABE site 1, 2, and 3). The efficiencies for edited adenines within the window (4-8 of the protospacers) of the on-target sites for these three sites were shown. Right panel: the number of RNA A-to-I edits (calculated by MuTect2) induced by engineered ABEmax variants was presented; ABEmax (rep3) and Cas9n (rep3) served as controls. Non-transfected HEK293T transcriptome server as control, and the RNA A-to-I edits overlapping with endogenous A-to-I edits were excluded. (b) Manhattan plots showing the representative distributions of RNA A-to-I edits across all chromosomes induced by ABEmax (rep 3), del153/dele153* (rep 1), and mini del153 (rep 1) with a sgRNA targeting ABE site 8 in HEK293T cells from RNA-seq data. (c) The diagram showing the number of RNA A-to-I edits induced by ABEmax (rep 3 and rep4), del153/dele153* (rep 1 and rep 2), and mini del153 (rep 1 and rep2) with a sgRNA targeting ABE site 8 in HEK293T cells from RNA-seq analysis. (d) Bar plots showing the DNA on-target A-to-G editing efficiencies of ABEmax, del153/dele153* and mini del153 with 8 sgRNAs. The efficiencies for most highly edited adenines for each sgRNA on-target site within the editing window are reported, and error bars represent s.d. (e) The number of RNA edits calculated by HaplotypeCaller or MuTect2 or overlapped RNA edits was presented for RNA-seq experiments in HEK293T cells that expressed Cas9n, mini ABEmax, ABE8e, ABE8e-del153, ABE8s, ABE8s-del153 and sgRNA targeting HEK site 8. Two independent replicates were shown. (f) Heat maps showing the on-target A-to-G editing efficiencies of Cas9n, ABEmax, del153/dele153*, mini ABEmax, mini del153, ABE8e, ABE8e-del153, ABE8s, ABE8s-del153 with 6 sgRNAs. Data are produced from two independent replicates. A-to-G editing efficiencies are shown in heat map format for the editing window 2-9. Numbering at the bottom of the heat map represents spacer position and 1 was calculated from the most PAM-distal nucleotides.

**Supplementary Files**

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

- SupplementalFigures.pdf
- SupplementaryFigureLegends.docx
- SupplementaryinformationTable.docx
- SourceDataforMainFigures.xlsx
- SourceDataforSupplementaryFigures.xlsx
- SourceDataforcalledmutationsfromRNAseqdata.xlsx