nature portfolio

Peer Review File

Engineering TadA ortholog-derived cytosine base editor without motif preference and adenosine activity limitation

Editorial Note: Parts of this Peer Review File have been redacted as indicated to remove third-party material where no permission to publish could be obtained.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this manuscript, the authors report a new cytosine base editor (aTdCBE) that uses an engineered TadA deaminase ortholog to perform cytidine deamination. They screened TadA orthologs for their ability to mediate cytosine base editing in a fluorescence reporter assay, and used MSA-guided engineering to improve the editing efficiency of one particular ortholog (AjTadA). The resulting aTdCBE exhibited efficient cytosine base editing in mammalian cells that was comparable to the efficiency of the previously reported TadCBEd base editor. Compared to TadCBEd, aTdCBE offers lower unwanted A-to-G base editing. The authors also demonstrated that aTdCBE can be delivered using a dual-AAV system via intramuscular injection into a humanized DMD mouse model to yield 40% base editing and substantial phenotypic rescue.

Overall, the manuscript is a solid contribution to the base editing field. However, I have several major concerns about the work that should be addressed prior to publication:

1. The lack of conceptual novelty limits my enthusiasm for the work. There have been extensive previous efforts to develop TadA-based CBEs (Lam et al. Nat. Biotechnol. 2023; Neugebauer et al. Nat. Biotechnol. 2023; Chen et al. Nat. Biotechnol. 2023; Zhang et al. Nat. Commun. 2023; Zhang et al. Nat. Commun. 2024). While the aTdCBE developed in this study has slightly different properties than previously reported TadA-based CBEs, it is unclear how useful these properties will be (as described in more detail below). Therefore, while the study is technically sound, the absence of conceptual novelty limits the impact of the study.

2. The major difference between aTdCBE and the best previously reported TadCBEd is that aTdCBE exhibits lower unwanted A-to-G base editing. While this claim is supported by the data, the absolute level of residual A-to-G base editing present in TadCBEd is very low (<10%), and so it is unclear how useful it is to minimize this from <10% in TadCBEd to <1% in aTdCBE. If the authors could provide evidence of the utility of aTdCBE's reduced A-to-G editing (for example by editing a therapeutically relevant locus in which the residual A-to-G editing leads to a missense mutation and must therefore be minimized as much as possible), then that would considerably increase the impact of the study.

3. A recent paper published in Nature Communications (Zhang et al. 2024) reports an almost identical finding to this manuscript: the development of CBE6 variants that offer substantially minimized residual A-to-G editing compared to TadCBEd. Because the results are so similar and were published in the same journal where the present manuscript is being reviewed, the authors should include additional data comparing aTdCBE to the CBE6 variants reported in the above paper. This additional data will be essential to benchmark the utility of aTdCBE relative to the most recently developed CBE6 variants.

4. While almost all the data in the manuscript is of sufficiently high quality, the RNA off-target editing analysis presented in Extended Data Fig. 5c is not an appropriate analysis. Because RNA off-target editing can occur at a low frequency across all positions of different transcript molecules, individual SNVs cannot be called, since this would underrepresent the total amount of off-target RNA editing. Standard RNA off-target base editing analyses either perform targeted amplicon sequencing of a few representative transcripts (Neugebauer et al. Nat. Biotechnol. 2024), calculate percent A-to-I or C-to-U editing for every A or C in the transcriptome (Grunewald et al. Nature 2019), or calculate average percent A-to-I or C-to-U editing across all As or Cs in the transcriptome (Richter et al. Nat. Biotechnol. 2020; Reichart et al. Nat. Med. 2023; Arbab et al. Science 2023), rather than calling individual SNVs that must surpass some frequency threshold in the overall RNA population. The authors should reanalyze their RNAseq data to calculate the average percent A-to-I or C-to-U editing across all As or Cs in the transcriptome, as is the standard in previously published base editing manuscripts.

Other comments:

5. Some sentences in the introduction are missing appropriate citations. For example, the authors should cite Gaudelli et al. Nature 2017 when discussing how ABEs were developed. Citing a general review on base editing (e.g. Porto et al. Nat. Rev. Drug Discov. 2020) would also be helpful.

6. The claim in Line 75, "To facilitate the clinical translation of therapeutic cytosine base editor into Duchenne muscular dystrophy (DMD) patients", should be rephrased, since this current manuscript is not a pre-clinical study and is instead a demonstration of therapeutic potential.

7. When discussing the fluorescence reporter assay in the first section of the Results, it would be helpful to state explicitly in the text that this reporter is measures C-to-G editing. This is stated in Fig. 1a, but mentioning it more clearly in the text would improve the clarity of the manuscript. In general, statements regarding "cytosine base editing" historically refer to C-to-T editing, so clarifying which cases of cytosine base editing are C-to-G editing versus C-to-T base editing would be helpful.

8. When base editor constructs are mentioned in the text, the authors should clarify whether these constructs contain UGIs. This is related to whether C-to-G or C-to-T edits are desired in each experiment.

9. The authors should reference the papers that first report the base editors against which aTdCBE is compared. Although these papers might be cited elsewhere in the manuscript, it would improve clarity if each paper associated with each specific CBE were referenced when they are mentioned for the first time (line 158).

10. The authors should reference the papers that report the orthogonal R-loop assay (Doman et al. Nat. Biotechnol. 2020; Yu et al. Nat. Commun. 2020) when this is mentioned in the text (line 187).

11. In the main text, the authors should briefly describe the dual-AAV delivery constructs that they used, and reference appropriate literature precedent. From Fig. 4a, it appears that they used split-intein vectors, but inteins are currently not mentioned at all in the text.

Reviewer #2 (Remarks to the Author):

The study titled "Engineering TadA ortholog-derived cytosine base editor without motif preference and adenosine activity limitation" by Li et al. engineered the Acinetobacter junii TadA into a cytidine deaminase and developed a new CBE system, aTdCBE. The editing activity of aTdCBE was evaluated in HEK293T cells and a DMD mouse model. Although the overall editing efficiency of the aTdCBE is satisfactory, there are many missing details and scientifically inaccurate descriptions in this manuscript. Because of this, I do not believe the current manuscript has the novelty or completeness for publication in its current state. Below are my specific comments:

1. The authors analyzed more than 10,000 TadA orthologs and constructed a phylogenetic tree based on sequence similarities. 9 TadA orthologs were picked for experimental evaluation so the authors need to better explain why and how these 9 orthologs were finally selected.
2. The authors designed a reporter system to screen the TadA orthologs for cytosine base editing. According to Line 88-95, cytosine base editing of the stop codon TAG into TAC would rescue the translation of the downstream EGFP, however, cytosine base editing on TAG should lead to another stop codon, TAA, rather than TAC. Therefore, I believe that this reporter system is inherently biased in itself and artificially selects for deaminases with C-to-G activity and would fail to detect the false negative samples with purely C-to-T editing activity. Because of this, I noticed that the authors used CGBE as a control, but this does not mimic the activity of a cytosine base editor.
3. In figure 3a and 3b, the authors are biased in what they chose to present and not present. For

example, according to Figure 3a, editing at TC11 and CC12 is low at the EMX1 site, while in Figure 3b, these two positions did not appear in the plotted analysis.
4. The authors evaluated too few sites to make any concluding statements regarding these deaminases. In the field of base editing, it is very difficult to make any conclusions based off of a few endogenous sites (which could be cherry picked). Furthermore, I do not understand why the authors only chose the particular cytosine base editors they used in Figure 3a as a control for comparing their own. For instance, YE1-BE4max is known to exhibit low off-target editing; however, this is not an aspect of their particular editor that they wanted to evaluate. If it is purely based on activity, then the authors should compare their editor with hA3A or A3B.
5. The authors should evaluate a high-throughput library to evaluate the editing motif preference for their aTdCBE.
6. In line 151, the authors choose theY8R, Y43R, and Y146R (maybe D146R?) for further combination. Can the authors explain why I147R and E46R, which performed better in Fig. 2b was not selected?
7. There are many mistakes in the texts and Figures, for example, "TadCBEd" in some places, while changed into "Tad-CBEd" in other places, and they and many other terminologies were not appropriately cited when appeared in the first time. Line 137, "E25R and D46R variants…", however, in Fig. 2b, they are N25R and E46R. In Line 150, it is "Y146R", while in Fig. 2b, it is D146R. Fig. 3a, SSH2 site, TC5 in the figure legend should be TC2.

Reviewer #3 (Remarks to the Author):

The authors have used an elegant humanized dystrophic mouse model to test the efficacy of a novel cytosine base-derived editor in restoring dystrophin expression. The work is original and it supports the conclusions that the authors claim.
The methods sections should be further clarified: Please include all relevant data in the generation of the humanized model section: sex, crosses with the mdx mouse, further information on sgRNA generation (L, R, PCR primers, template…), further background information on the mouse lines and mutations for both hDMD and mdx mice (i.e. which chromosomes are affected, ... etc.). Please note that the transgenic insertion on chromosome 5 is a tail-to-tail duplication [Yavas et al. 2020 PLoS One. 15:e0244215 (PMID:33362201)]. Also, please detail the WT line that was used for comparison in the extended Figure 7. Some of this information is scattered throughout the manuscript, but should be included in this section.
In figure 4, which areas of the injected muscles where analyzed for dystrophin expression? A representative image showing the entire section of the injected muscle should be included, highlighting the areas analyzed.
Lastly, statistical analyses should be revised. Normal distribution of datasets should be verified before using parametric tests. Also, ANOVA should be used to compare more than two groups. Please, include the statistical analyses used in the figure legends.

Point-to-point response to Reviewers' comments

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

In this manuscript, the authors report a new cytosine base editor (aTdCBE) that uses an engineered TadA deaminase ortholog to perform cytidine deamination. They screened TadA orthologs for their ability to mediate cytosine base editing in a fluorescence reporter assay, and used MSA-guided engineering to improve the editing efficiency of one particular ortholog (AjTadA). The resulting aTdCBE exhibited efficient cytosine base editing in mammalian cells that was comparable to the efficiency of the previously reported TadCBEd base editor. Compared to TadCBEd, aTdCBE offers lower unwanted A-to-G base editing. The authors also demonstrated that aTdCBE can be delivered using a dual-AAV system via intramuscular injection into a humanized DMD mouse model to yield 40% base editing and substantial phenotypic rescue.

Overall, the manuscript is a solid contribution to the base editing field. However, I have several major concerns about the work that should be addressed prior to publication:
Response: We appreciate your highly professional comments and suggestions.

1. The lack of conceptual novelty limits my enthusiasm for the work. There have been extensive previous efforts to develop TadA-based CBEs (Lam et al. Nat. Biotechnol. 2023; Neugebauer et al. Nat. Biotechnol. 2023; Chen et al. Nat. Biotechnol. 2023; Zhang et al. Nat. Commun. 2023; Zhang et al. Nat. Commun. 2024). While the aTdCBE developed in this study has slightly different properties than previously reported TadA-based CBEs, it is unclear how useful these properties will be (as described in more detail below). Therefore, while the study is technically sound, the absence of conceptual novelty limits the impact of the study.
Response: Thank you for your comments and suggestions. We have added application related experiments to highlight the advance of this system. As shown in the next response
2. The major difference between aTdCBE and the best previously reported TadCBEd is that aTdCBE exhibits lower unwanted A-to-G base editing. While this claim is supported by the data, the absolute level of residual A-to-G base editing present in TadCBEd is very low (<10%), and so it is unclear how useful it is to minimize this from <10% in TadCBEd to <1% in aTdCBE. If the authors could provide evidence of the utility of aTdCBE's reduced A-to-G editing (for example by editing a therapeutically relevant locus in which the residual A-to-G editing leads to a missense mutation and must therefore be minimized as much as possible), then that would considerably increase the impact of the study.
Response: We have added application related experiments as below.

“By introducing stop codons(CAA/CAG/CGA to TAA/TAG/TGA), CBEs have shown great potential in the treatment of common diseases such as T-cell acute lymphoblastic leukemia[8], Hepatitis B[9,10], and Acquired immunodeficiency syndrome[11].”

“Moreover, unwanted A-to-G base editing can cause the termination codon to become TGG, limiting the disruption effect of the target gene”

*“We then tested the effects of aTdCBE, TadCBEd, and recently reported CBE6b[34] on introducing premature termination codon in PCSK9 coding region, using four sites containing the CAG codon (**Fig.3f**). We found that TadCBEd produces A-to-G editing at all four sites, while aTdCBE and CBE6b hardly cause A-to-G editing. The aTdCBE performs better at the AC motif site than aTdCBE and CBE6b. The aTadCBE, TadCBEd and CBE6b can generate C-to-T editing efficiencies of up to approximately 72%, 57% and 60% for these CAG codon sites, respectively.”*

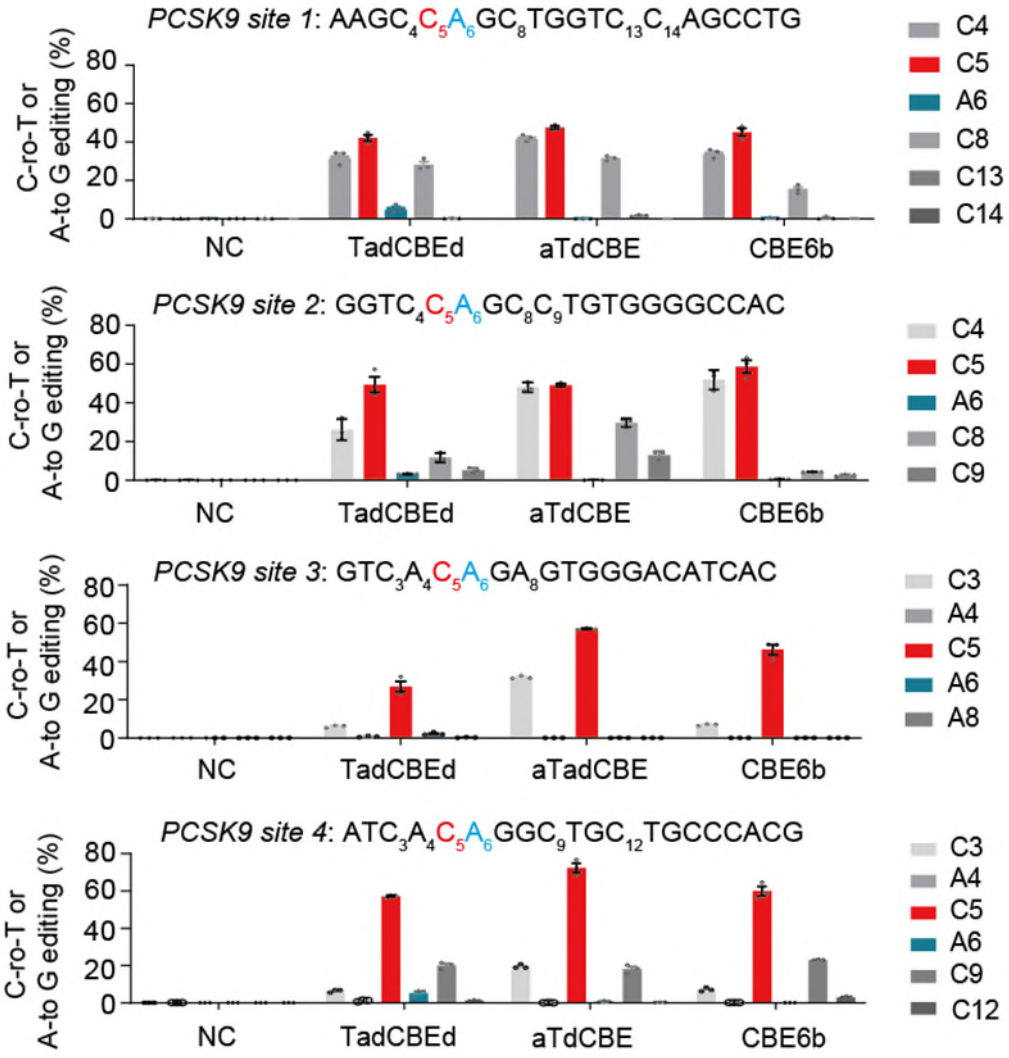


Fig. 3f| Introducing premature termination codons into the PCSK9 coding region using CBEs.

3. A recent paper published in Nature Communications (Zhang et al. 2024) reports an

almost identical finding to this manuscript: the development of CBE6 variants that offer substantially minimized residual A-to-G editing compared to TadCBEd. Because the results are so similar and were published in the same journal where the present manuscript is being reviewed, the authors should include additional data comparing aTdCBE to the CBE6 variants reported in the above paper. This additional data will be essential to benchmark the utility of aTdCBE relative to the most recently developed CBE6 variants.

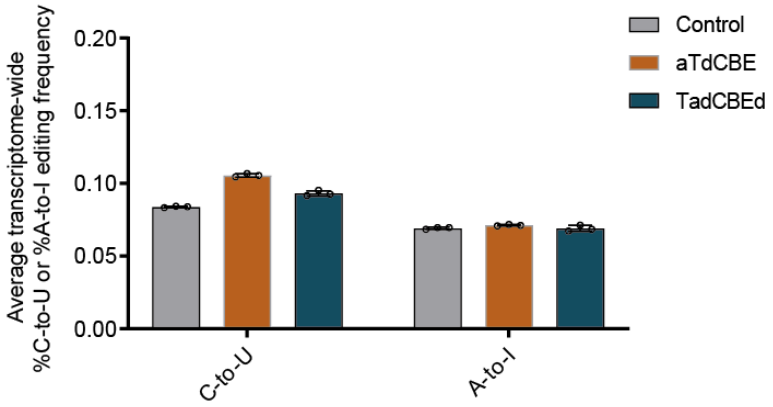
Response: Thank you for your suggestions. We have compared aTdCBE and CBE6b using PCSK9 sites and found that aTdCBE has higher editing efficiency at the AC motif site than CBE6b, while CBE6b has higher editing efficiency at certain motif sites (Fig 3f). Therefore, these two editors can complement each other in terms of functionality.

4. While almost all the data in the manuscript is of sufficiently high quality, the RNA off-target editing analysis presented in Extended Data Fig. 5c is not an appropriate analysis. Because RNA off-target editing can occur at a low frequency across all positions of different transcript molecules, individual SNVs cannot be called, since this would underrepresent the total amount of off-target RNA editing. Standard RNA off-target base editing analyses either perform targeted amplicon sequencing of a few representative transcripts (Neugebauer et al. Nat. Biotechnol. 2024), calculate percent A-to-I or C-to-U editing for every A or C in the transcriptome (Grunewald et al. Nature 2019), or calculate average percent A-to-I or C-to-U editing across all As or Cs in the transcriptome (Richter et al. Nat. Biotechnol. 2020; Reichart et al. Nat. Med. 2023; Arbab et al. Science 2023), rather than calling individual SNVs that must surpass some frequency threshold in the overall RNA population. The authors should reanalyze their RNAseq data to calculate the average percent A-to-I or C-to-U editing across all As or Cs in the transcriptome, as is the standard in previously published base editing manuscripts.

Response: Thank you for your valuable suggestions. We have calculated average percent A-to-I or C-to-U editing across all As or Cs in the transcriptome based on the previously reported method (Richter et al. Nat. Biotechnol. 2020). In addition, we have revised the descriptions and methods, and added a discussion

"*To investigate whether TadA variants with cytosine deaminase activity can cause transcriptome-wide off-target effects, we conducted RNA-seq to evaluate RNA off-target effect for aTdCBE and TadCBEd. Compared to the HEK293T control without the base editor transfection, both aTdCBE and TadCBEd showed a slight off-target effect on cytosine in RNA, while TadCBEd showed lower off-target effects* (**Extended Data Fig. 5c**)**"**

"*Although these TadA deaminases exhibited slight RNA off-target of C-to-U compared to the control, their specificity can be further improved by mutant modification or insertion of deaminases into the Cas9 protein in the future*[37, 38]*."*



Extended Data Fig. 5c| Transcriptome-wide off-target analysis of aTdCBE and TadCBEd.

"RNA-seq for off-target analysis

To quantify the transcriptome deaminases off-target edits, HEK293T cells were cultured in 10-cm dishes with 80% confluence and transfected with 35 μg plasmids containing base editors and gRNA. After 48 hours, about 600,000 transfected cells were sorted by FACS, and RNA was extracted using Trizol (Ambion) for RNA-seq library preparation. An RNA-seq library was generated with a TruSeq Stranded Total RNA library preparation kit according to the standard protocol. The transcriptome libraries were sequenced using a 150-bp paired-end Illumina NovaSeq 6000 platform (Genewiz Co. Ltd.).

The calculation analysis referred to previously published methods[38]. Trimmomatic (v.0.39-2) were using to filter the RNAseq raw data[48]. The clean reads were aligned to the hg38 reference genome with Hisat2 (v.2.2.1)[49]. RNA editing sites were calculated using REDItools (v1.2.1) with "-e -d -p -u -m 60 -T 5-5 -W -n 0.0" parameters[50]. The edited adenosines divided by total adenosines and the edited cytosines divided by total cytosines were calculated separately."

Other comments:

5. Some sentences in the introduction are missing appropriate citations. For example, the authors should cite Gaudelli et al. Nature 2017 when discussing how ABEs were developed. Citing a general review on base editing (e.g. Porto et al. Nat. Rev. Drug Discov. 2020) would also be helpful.

Response: Thanks for your suggestion. We have added the new citations in the revised manuscript accordingly.

6. The claim in Line 75, "To facilitate the clinical translation of therapeutic cytosine base editor into Duchenne muscular dystrophy (DMD) patients", should be rephrased, since this current manuscript is not a pre-clinical study and is instead a

demonstration of therapeutic potential.

Response: Thanks for your kindly reminding. We have rephrased this words accordingly. "*We then established a genetically humanized DMD mouse model and demonstrated the potential of exon skipping and dystrophin restoration using AAV delivery, which may inform future pre-clinical research.***"**

7. When discussing the fluorescence reporter assay in the first section of the Results, it would be helpful to state explicitly in the text that this reporter is measures C-to-G editing. This is stated in Fig. 1a, but mentioning it more clearly in the text would improve the clarity of the manuscript. In general, statements regarding "cytosine base editing" historically refer to C-to-T editing, so clarifying which cases of cytosine base editing are C-to-G editing versus C-to-T base editing would be helpful.

Response: Thank you for your suggestion. We have revised the description in the txt, accordingly.

"*To sensitively detect DNA base editing activity in mammalian cells, we designed a fluorescent reporter system, termed as BFP-*EGFP which contains a BFP and an inactive EGFP carrying stop mutation that can be corrected through C-to-G base editing (**Fig. 1a**)***"**

8. When base editor constructs are mentioned in the text, the authors should clarify whether these constructs contain UGIs. This is related to whether C-to-G or C-to-T edits are desired in each experiment.

Response: Thank you for your reminding. We have added the descriptions in the figure legends accordingly.

9. The authors should reference the papers that first report the base editors against which aTdCBE is compared. Although these papers might be cited elsewhere in the manuscript, it would improve clarity if each paper associated with each specific CBE were referenced when they are mentioned for the first time (line 158).

Response: Thank you for your suggestion. We have added the reference accordingly.

10. The authors should reference the papers that report the orthogonal R-loop assay (Doman et al. Nat. Biotechnol. 2020; Yu et al. Nat. Commun. 2020) when this is mentioned in the text (line 187).

Response: Thank you for your reminding. We have added the reference accordingly.

11. In the main text, the authors should briefly describe the dual-AAV delivery constructs that they used, and reference appropriate literature precedent. From Fig. 4a, it appears that they used split-intein vectors, but inteins are currently not mentioned at all in the text.

Response: Sorry for the undefined description. We have added the descriptions accordingly.

“*The dual-AAV delivery system was designed to express two separate fragments of a base editor, which are subsequently spliced into the full-length protein via the Rhodothermus marinus (Rma) intein, as reported in a previous study*[44, 45]*. The intein sequences, specifically the N- and C-terminal segments, were synthesized by Genewiz (Suzhou, China). These segments were then integrated into the 573 and 574 amino acid residues of the aTdCBE backbones using Gibson cloning of PCR-amplified inserts.***”**

Reviewer #2 (Remarks to the Author):

The study titled "Engineering TadA ortholog-derived cytosine base editor without motif preference and adenosine activity limitation" by Li et al. engineered the Acinetobacter junii TadA into a cytidine deaminase and developed a new CBE system, aTdCBE. The editing activity of aTdCBE was evaluated in HEK293T cells and a DMD mouse model. Although the overall editing efficiency of the aTdCBE is satisfactory, there are many missing details and scientifically inaccurate descriptions in this manuscript. Because of this, I do not believe the current manuscript has the novelty or completeness for publication in its current state. Below are my specific comments:

Response: We appreciate your valuable suggestions for improving the quality and completeness of our manuscript.

1. The authors analyzed more than 10,000 TadA orthologs and constructed a phylogenetic tree based on sequence similarities. 9 TadA orthologs were picked for experimental evaluation so the authors need to better explain why and how these 9 orthologs were finally selected.

Response: Sorry for the undefined description.

"*MCL (v14-137)[43] was used to cluster the redundant proteins with identity over than 70%, and nine TadA orthologs were randomly selected from different clusters for experimental screening.***"**

2. The authors designed a reporter system to screen the TadA orthologs for cytosine base editing. According to Line 88-95, cytosine base editing of the stop codon TAG into TAC would rescue the translation of the downstream EGFP, however, cytosine base editing on TAG should lead to another stop codon, TAA, rather than TAC. Therefore, I believe that this reporter system is inherently biased in itself and artificially selects for deaminases with C-to-G activity and would fail to detect the false negative samples with purely C-to-T editing activity. Because of this, I noticed that the authors used CGBE as a control, but this does not mimic the activity of a cytosine base editor.

Response: Thank you for your comments. Due to the inability of C-to-T editing to disrupt stop codon, it is difficult to design a sensitive EGFP activated fluorescence reporting system to detect C-to-T editing activity. We believe that for the same target, the C-to-G and C-to-T activities of the cytosine base editor without UGI are theoretically positively correlated. Of course, the C-to-G and C-to-T activities of certain TadA base editors with adenosine deaminase and cytosine deaminase activity may also be affected by A-to-G activity, as A-to-G activity alters the sequence context and may affect the formation and repair of AP sites. Previously reported CBEs based on cytosine deaminase and UGI can both produce pure C-to-T editing. Therefore, I believe that cytosine deaminases with efficient C-to-G editing activity can be developed into highly active CBEs by

fusing UGI. In addition, in this study, we compared aTdCBE, TadCBEd and Td-CBEmax using high-throughput library experiments and found that their average C-To-G activity in the editing window was very low.

Table R1 The base editing activity of CBEs was evaluated using high-throughput library.

	C-to-G (%)	C-to-T (%)	C-to-G / C-to-T (%)
TadCBEd	0.2096	18.2724	1.1473
aTdCBE	0.1999	20.3789	0.9808
Td-CBEmax	0.1164	10.9952	1.0585

3. In figure 3a and 3b, the authors are biased in what they chose to present and not present. For example, according to Figure 3a, editing at TC11 and CC12 is low at the EMX1 site, while in Figure 3b, these two positions did not appear in the plotted analysis.

Response: Sorry for the missing statement. *"The editing window of CBEs was primarily concentrated between bases 4-7. Consequently, the analysis focused on preference motifs exhibiting higher editing efficiency within this range."*

4. The authors evaluated too few sites to make any concluding statements regarding these deaminases. In the field of base editing, it is very difficult to make any conclusions based off of a few endogenous sites (which could be cherry picked). Furthermore, I do not understand why the authors only chose the particular cytosine base editors they used in Figure 3a as a control for comparing their own. For instance, YE1-BE4max is known to exhibit low off-target editing; however, this is not an aspect of their particular editor that they wanted to evaluate. If it is purely based on activity, then the authors should compare their editor with hA3A or A3B.

Response: Detecting the editing efficiency of 25 endogenous loci with various editors is not as easy in the field of gene editing. Chen, L., et al (PMID: 36357717), Neugebauer, M. E., et al (PMID: 36357719) , and Zhang, S., et al (PMID: 36702837) have tested 23, 9, and 12 endogenous Cas9 loci, respectively.
This work mainly focuses on comparing TadA cytosine deaminases, so we have chosen the best version reported in the above three articles. Considering that YE1-BE4max is currently the most commonly used cytosine base editor, we chose it as the control. However, we agree with using hA3A variant to validate the 25 endogenous loci.

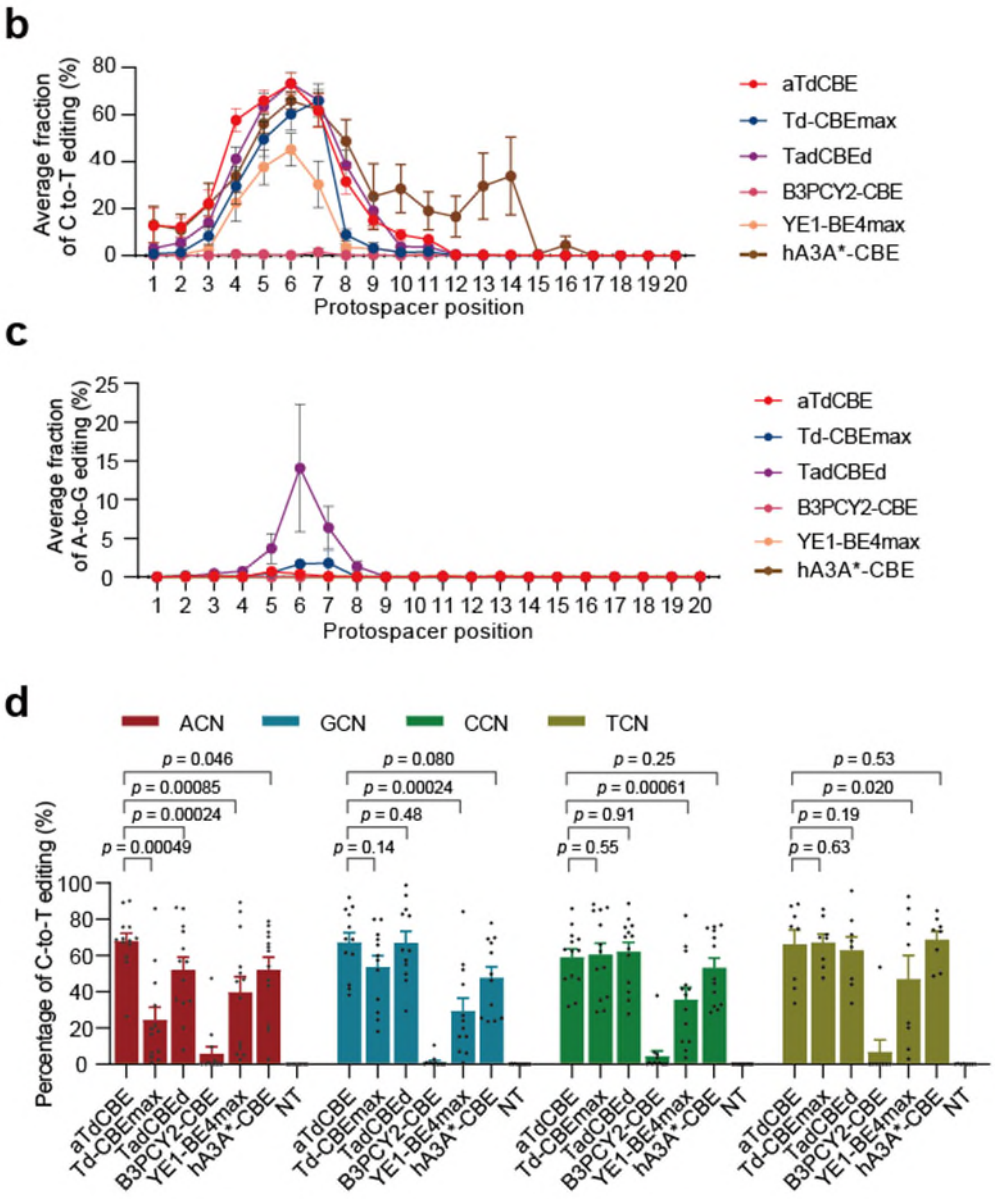


Fig. 3b-d| aTdCBE enables robust genomic base editing in mammalian cells. b. Base editing activity window plots showing mean C-to-T editing at all tested target positions. **c.** Base editing activity window plots showing mean A-to-G editing at all tested target positions. **d**.Comparison of motif preference for 25 endogenous loci by base editors derived from aTdCBE, Td-CBEmax, TadCBEd, B3PCY2-CBE, hA3A*-CBE and YE1-BE4max in HEK293T cells.The editing window of CBEs was primarily concentrated between bases 4-7. Consequently, the analysis focused on preference motifs exhibiting higher editing efficiency within this range.

5. The authors should evaluate a high-throughput library to evaluate the editing motif preference for their aTdCBE.

Response: Thank you for your valuable suggestion. We have added high-throughput library experiments to evaluate the editing motif preference of

aTdCBE, Td-CBEmax, TadCBEd and B3PCY2-CBE.

*“To verify the motif preference of the TadA deaminases, we performed high-throughput library experiments using a library containing 11,868 paired sgRNA[34]. Consistent with previous results, the aTdCBE has better editing efficiency for AC sites **(Fig.3e)**.”*

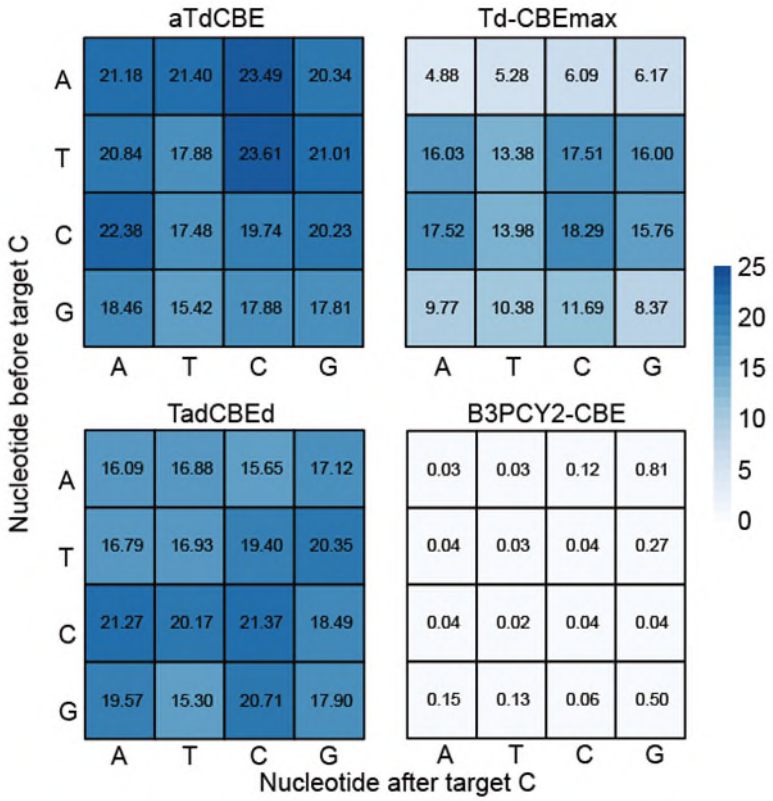


Fig. 3e| High-throughput library experiments to evaluate the motif preference of aTdCBE, Td-CBEmax, TadCBEd and B3PCY2-CBE. **f.** Introducing premature termination codons into the PCSK9 coding region using CBEs. The numbers in the grid represent the average editing efficiency of cytosines with corresponding motifs at positions 4-7 for each sgRNA.

*“**High-throughput library experiments***

The cell line previously constructed using 11,868 pairs of sgRNA lentivirus plasmid library was used to detect motif preference in the base editor[34]. For each 10-cm dish, 35 μg plasmids that encode CBEs and mCherry were transfected using PEI. After 48 hours, transfected cells were harvested using FACS followed by genomic DNA extraction. The PCR products were sequenced using a 150-bp paired-end Illumina NovaSeq 6000 platform (Genewiz Co. Ltd.). High-throughput sequencing datasets were processed using CRISPResso2 to calculate editing efficiency of each target. The target sites were excluded with a coverage depth of less than 100 in each sample. Cytosines in positions 4-7 of the target sequences were used to statistically analyze motif preferences.”

6. In line 151, the authors choose theY8R, Y43R, and Y146R (maybe D146R?) for further combination. Can the authors explain why I147R and E46R, which performed better in Fig. 2b was not selected?

Response: Sorry for the mistakenly marking of P46R as E46R in Fig.2b. We have made correction in the revised version. Due to AjTadA.v2 containing P46G mutations, we consider that introducing adjacent I47R mutation has a high potential to disrupt the effect of the P46G mutation.

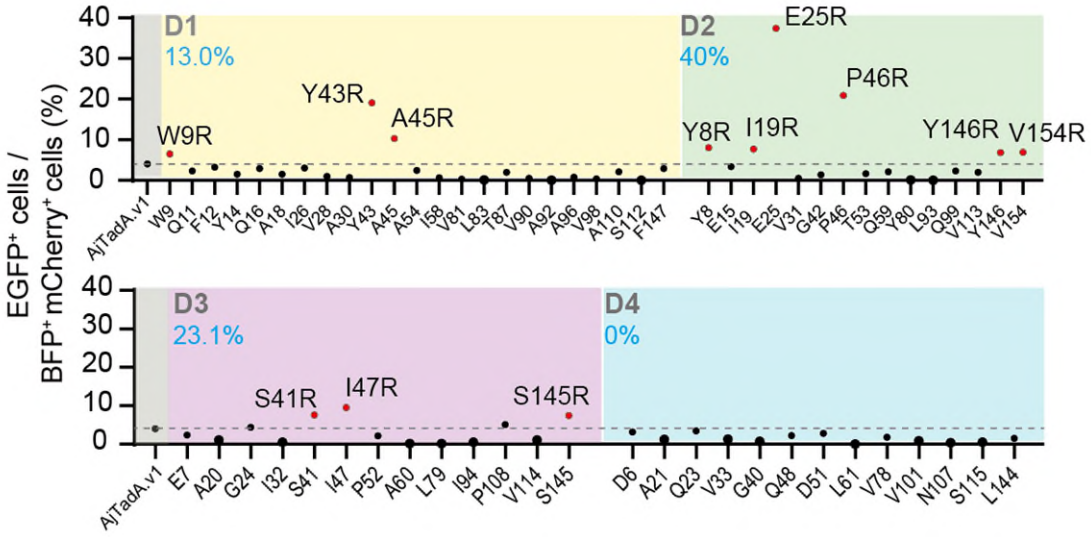


Fig. 2b| Substitutions of non-positively charged amino acids of AjTadA at the AHC sites. Each dot represents activity for a single variant.

7. There are many mistakes in the texts and Figures, for example, "TadCBEd" in some places, while changed into "Tad-CBEd" in other places, and they and many other terminologies were not appropriately cited when appeared in the first time. Line 137, "E25R and D46R variants…", however, in Fig. 2b, they are N25R and E46R. In Line 150, it is "Y146R", while in Fig. 2b, it is D146R. Fig. 3a, SSH2 site, TC5 in the figure legend should be TC2.

Response: Thank you for your valuable comments. We carefully examined these issues and made corresponding corrections.

Reviewer #3 (Remarks to the Author):

The authors have used an elegant humanized dystrophic mouse model to test the efficacy of a novel cytosine base-derived editor in restoring dystrophin expression. The work is original and it supports the conclusions that the authors claim.
Response: We greatly appreciate your interest and recognition in our work on restoring dystrophy proteins.

The methods sections should be further clarified: Please include all relevant data in the generation of the humanized model section: sex, crosses with the mdx mouse, further information on sgRNA generation (L, R, PCR primers, template…), further background information on the mouse lines and mutations for both hDMD and mdx mice (i.e. which chromosomes are affected, ... etc.). Please note that the transgenic insertion on chromosome 5 is a tail-to-tail duplication [Yavas et al. 2020 PLoS One. 15:e0244215 (PMID:33362201)]. Also, please detail the WT line that was used for comparison in the extended Figure 7. Some of this information is scattered throughout the manuscript, but should be included in this section.
Response: Thank you for your detailed feedback, which has highlighted areas in our methods section that required further clarification. We appreciate your guidance in ensuring that our manuscript provides a comprehensive and clear description of the experimental models and methods used.
In response to your comments, we have revised the methods section to include a more detailed description of the generation of our humanized model, the sgRNA design, and the background information on the mouse lines. We have also ensured that all relevant details are appropriately consolidated within this section for clarity and ease of understanding. Revised methods section addition:

" *To generate the humanized DMD$^{\Delta E54}$ mice, we employed the CRISPR/Cas9 system on the embryos obtained from mating STOCK Tg(DMD)72Thoen/J male and female mice (#018900). Specifically, we designed two sgRNAs targeting the flanking introns of human DMD exon 54 on Chr.5. The sequences of these sgRNAs are gRNA1: gTTTCTGCAAGTGCAGAGAGG and gRNA2: GGTGTGTGGAGTGAGATACT. Each sgRNA template was appended with the T7 promoter sequence (TAATACGACTCACTATAg) for efficient transcription. All sequences are listed in Supplementary Table 1."*

"Exclusively male mice were utilized for all experiments, including grip strength tests, creatine kinase (CK) analysis, and AAV injections."

For comparative analyses, wild-type (WT) mice were derived from crosses between STOCK Tg (DMD) 72Thoen/J mice (#018900) and mdx mice, which carry a c.2977C>T, p.Gln993 mutation in exon 23 on Chr.X, known for its impact on dystrophin expression.***"**

Regarding the transgenic insertion on chromosome 5, noted as a tail-to-tail duplication [Yavas et al. 2020, PLoS One. 15:e0244215 (PMID:33362201)], we have revised Figure 4a to accurately reflect this configuration.

[Redacted]

Fig. 4a | Strategy for generating humanized DMD mouse model.

We hope that these revisions adequately address your concerns and enhance the clarity of our methods section. We are grateful for your assistance in improving the quality and accuracy of our manuscript and welcome any further feedback.

In figure 4, which areas of the injected muscles where analyzed for dystrophin expression? A representative image showing the entire section of the injected muscle should be included, highlighting the areas analyzed.

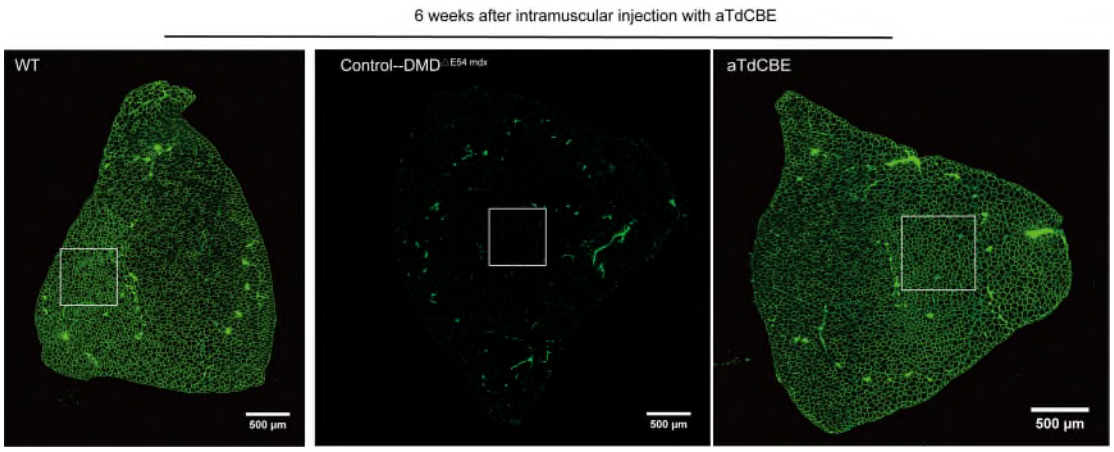
Response: Thank you for your insightful feedback.

"*Tissues were divided into distinct segments for targeted assessment. Specifically, the distal region was allocated for evaluating DNA editing and exon skipping efficiency, the middle portion was dedicated to Western blot analysis of dystrophin expression, and the proximal segment was reserved for immunofluorescent analysis of dystrophin levels at six weeks after treatment.*

"

Additionally, we have added Extended Data Figure 8 to the manuscript. This new figure includes a representative image showing the entire section of the injected muscle. We have clearly highlighted the specific areas that were analyzed for dystrophin expression to ensure clarity and provide a thorough understanding of the scope and focus of our analysis.

"



Extended Data Fig. 8 Rescue of dystrophin expression following intramuscular injection of aTdCBE after 6 weeks. Dystrophin immunohistochemistry of TA muscle. Control mice were injected with saline.

Images shown in both Fig. 4e and Extended Data Fig. 8 were obtained from the same tissue at 20x magnification. Fig. 4e showed the local region staining image rather than the reconstituted whole-tissue scanning image in Extended Data Fig. 8, and highlighted with white boxes. Dystrophin is shown in green. Scale bar, 500 μm.

Lastly, statistical analyses should be revised. Normal distribution of datasets should be verified before using parametric tests. Also, ANOVA should be used to compare more than two groups. Please, include the statistical analyses used in the figure legends.

Response: Thank you for your valuable suggestion. Due to the limitation of data sample size, it is difficult to determine whether the data follows a normal distribution. So we used the non-parametric testing method Mann Whitney U test to calculate the p-value in the revised manuscript.

"*P-values determined by one-sided Mann-Whitney U-test.* **"**

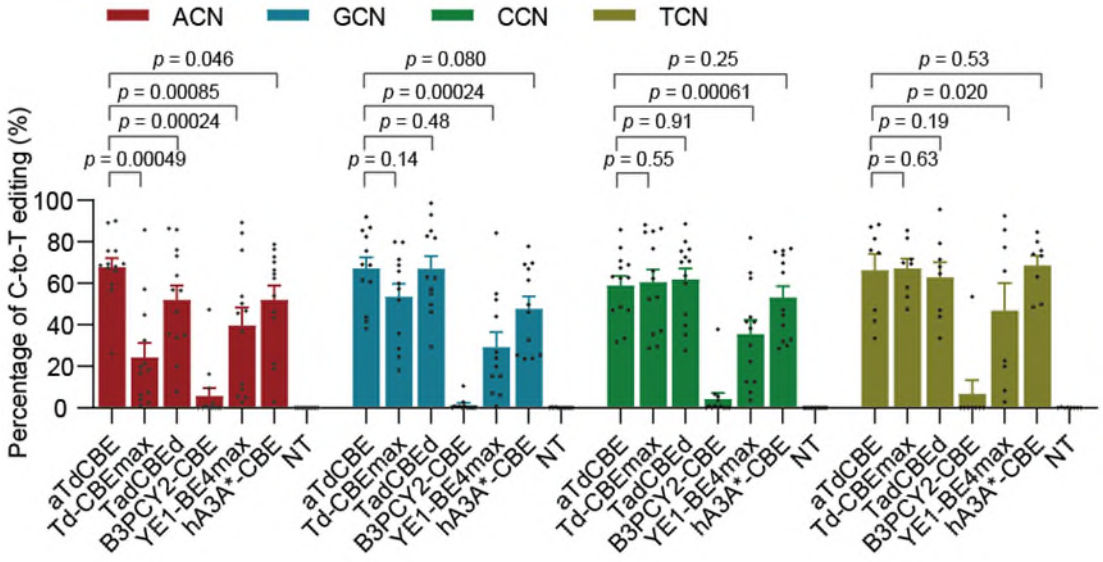


Fig. 3d| Comparison of motif preference for 25 endogenous loci by base editors derived from aTdCBE, Td-CBEmax, TadCBEd, B3PCY2-CBE, hA3A*-CBE and YE1-BE4max in HEK293T cells.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have addressed my major concerns, and the revised manuscript is substantially improved. I have a few remaining minor comments:

1. Some nomenclature errors: in some cases the new CBE is referred to as aTdCBE or aTadCBE (e.g. Fig. 3f)—the authors should pick a consistent nomenclature and correct all instances of this naming.

2. While the authors added a brief description of the dual-AAV delivery strategy to the methods section, it would be clearest to include a sentence in the main text describing this strategy. For example, added to lines 247-248.

Reviewer #2 (Remarks to the Author):

The revised manuscript has addressed some of my concerns, however, there are still lots of mistakes that need careful revision.
1. I still do not believe the authors' use of this reporter is adequate as it is not stringently established for discovering C-to-T editors. If the authors claim that there is a "theoretical" positive correlation between C-to-G and C-to-T, I hope the authors can validate this across at least a few sites. From previous literature, C-to-G is highly site and sequence dependent because it relies on translesion synthesis upon base excision repair (Koblan, 2021, NBT) whereas C-to-T is independent of translesion synthesis.
2. I thank the author for evaluating hA3A as this serves as a great control.
3. I also thank the author for now including a high-throughput library experiment.
4. Lastly, I also worry about the lack of conceptual novelty as reviewer 1 stated as there have been already many many advances for new cytosine base editors and many have demonstrated superior properties–I do not see a need for yet another TadA-derived CBE to be published in a notable journal like Nature Communications.

Minor comments below:
1. Line 146, "12 variants with significantly enhanced base editing activity" – statistical analysis is required when using the word "significantly".
2. Line 169-170, the authors should explain why they choose to compare these CBEs.
3. Line 196, "PCSK9" should be italic.
4. Line 214-216, citations are required.
5. Extended Data Fig. 2, the R72, L73, N126 looks quite conserved, but why were these not included? The V31 seems to be mislabeled.
6. Line 238, "SAS", Line 244, "CK", Line 247, "TA muscle", full names are required as they are the first time they appear.
7. Figure 3 and other places, there are at least three kinds of spellings for "C-to-T conversion" in the y axes.

Reviewer #3 (Remarks to the Author):

The authors have addressed the majority of my main concerns. However, the statistical analysis needs to be revised. The authors should run a normality test (D'Agostini or Shapiro–Wilk should work with their dataset). If the distribution is indeed not normal, they should run a non-parametric test for three or more groups such as a Kruskal-Wallis test, and a suitable posthoc test, such as Dunn's. Non parametric data should be represented as medians using box plots or similar. Mann-Whitney test is used to compare two groups, and therefore introduces error in multiple comparisons.

Point-to-point response to Reviewers' comments

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

The authors have addressed my major concerns, and the revised manuscript is substantially improved. I have a few remaining minor comments:

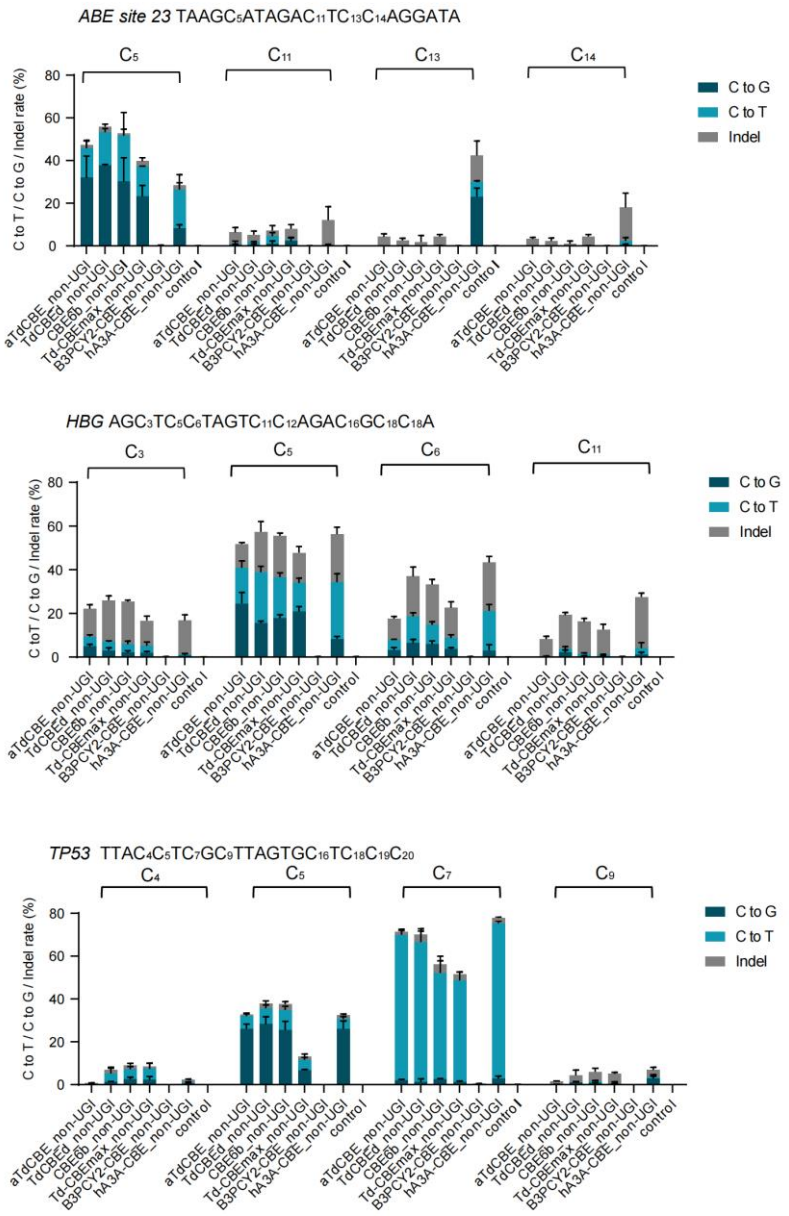
1. Some nomenclature errors: in some cases the new CBE is referred to as aTdCBE or aTadCBE (e.g. Fig. 3f)—the authors should pick a consistent nomenclature and correct all instances of this naming.
Response: Thank you for your reminding. We have carefully checked and corrected it.

2. While the authors added a brief description of the dual-AAV delivery strategy to the methods section, it would be clearest to include a sentence in the main text describing this strategy. For example, added to lines 247-248.
Response: Thank you for your suggestion. We have added the descriptions in the main text. "*Given the vector size beyond the genome packaging capacity of AAV, we have developed a strategy where each fragment of the base editor is expressed individually by two separate AAV vectors. This Rhodothermus marinus (Rma) intein facilitated the precise autocatalytic splicing of the two fragments, thereby reconstructing the full-length, active form of aTdCBE within the target cells.***".**

Reviewer #2 (Remarks to the Author):

The revised manuscript has addressed some of my concerns, however, there are still lots of mistakes that need careful revision.
1. I still do not believe the authors' use of this reporter is adequate as it is not stringently established for discovering C-to-T editors. If the authors claim that there is a "theoretical" positive correlation between C-to-G and C-to-T, I hope the authors can validate this across at least a few sites. From previous literature, C-to-G is highly site and sequence dependent because it relies on translesion synthesis upon base excision repair (Koblan, 2021, NBT) whereas C-to-T is independent of translesion synthesis.
Response: Thank you for your suggestion. We have conducted experiments to validate it.
"*Due to the engineered AjTadA was obtained through C-to-G fluorescence reporter screening.To investigate whether the edited product of engineered AjTadA exhibits C-to-G preference compared to other deaminase enzymes, we removed the UGIs of aTdCBE and TadCBEd performed better than Td-*

CBEmax, B3PCY2-CBE, hA3A-CBE and CBE6, and generated 6 CGBEs. Compared to the effect of sites and sequences on editing product preference, the effect of deaminase variants on C-to-G preference is relatively low (**Extended Data Fig. 6**).*"



Extended Data Fig. 6 | Compare the editing products of the base editors without UGI at three endogenous loci. Data are presented as means ± s.d. Values represent n = 3 independent biological replicates. All of the above base editors do not contain UGI.

2. I thank the author for evaluating hA3A as this serves as a great control.

3. I also thank the author for now including a high-throughput library experiment.

4. Lastly, I also worry about the lack of conceptual novelty as reviewer 1 stated as there have been already many many advances for new cytosine base editors and many have demonstrated superior properties–I do not see a need for yet another TadA-derived CBE to be published in a notable journal like Nature Communications.

Response: Thank you for candidly raising these concerns. There are several reasons to support the suitability of this manuscript for publication in Nature Communications.

Firstly, base editing technology plays a very important role in the field of genome editing. Due to its advantages such as small size, high DNA level specificity, and rich functional diversity compared to other deaminase enzymes, TadA deaminase is the most popular in the field. Our research fully demonstrates that aTdCBE has a significant improvement compared to previous TadA based deaminase. Although more cytosine deaminase variants were reported during our manuscript submission and review period (PMID: 38402281; PMID: 36229683； PMID: 38831042), these variants will not fundamentally affect the novelty of this manuscript.

Secondary，in this study, we introduced a feasible protein engineering strategy that does not rely on protein structure, artificial intelligence, or phage-directed evolution. Through screening dozens of mutants, we have improved the catalytic activity of the TadA deaminase by over 20-fold using fluorescence reporting system. Unlike other methods, our approach is more easily replicable for most ordinary laboratories，making it applicable for the engineering of other enzymes.

Thirdly, the successful delivery of aTdCBE to a humanized DMD mouse model represents a significant milestone in the field of DMD therapy. Human *DMD* mutations are clustered in specific “hotspot” areas of the gene, predominantly within exons 45 to 55 and exons 2 to 10 (PMID: 29404407). Notably, exon 55 skipping has the potential to restore dystrophin function in approximately 2% of DMD patients (PMID: 29404407). However, the *in vivo* application of base editing to induce exon 55 skipping has not been reported previously. This study demonstrates the successful achievement of robust exon 55 skipping and restoration of dystrophin expression, emphasizing the therapeutic potential of this technology in treating genetic disorders. This groundbreaking result underscores the crucial need for further exploration and refinement of TadA-derived CBEs for their future applications in gene therapy and precision medicine.

In light of these advancements and achievements, it is evident that the publication of a TadA-derived CBE in a reputable journal like Nature Communications is not only justified but also essential for disseminating important scientific findings to the research community. The continued development and optimization of gene editing technologies, including TadA-derived CBEs, are vital for addressing complex genetic diseases and advancing the field of molecular medicine.

Minor comments below:

1. Line 146, "12 variants with significantly enhanced base editing activity" – statistical analysis is required when using the word "significantly".

Response: Thank you for your reminding.The sentence has been revised.

2. Line 169-170, the authors should explain why they choose to compare these CBEs.

Response: Thank you for your suggestion. We have added a description here "*In order to compare the editing efficiency of our aTdCBE with other CBEs, Td-CBEmax (PMID: 36357717), TadCBEd (PMID: 36624149), B3PCY2-CBE (PMID: 36702837), hA3A^{W104A}(hA3A*)-CBE (PMID: 30125268) and YE1-BE4max (PMID: 32042165) were selected due to their diverse properties, including differences in editing efficiency, off-target effects, substrate preferences, and applicability across various cell types and model organisms. By including these specific CBEs in our study, we aimed to provide a comprehensive analysis of the current state-of-the-art cytidine base editing tools and facilitate a comparative assessment of their performance under standardized experimental conditions.***"** 。

3. Line 196, "PCSK9" should be italic.

Response: Thank you for your reminding. We have corrected it.

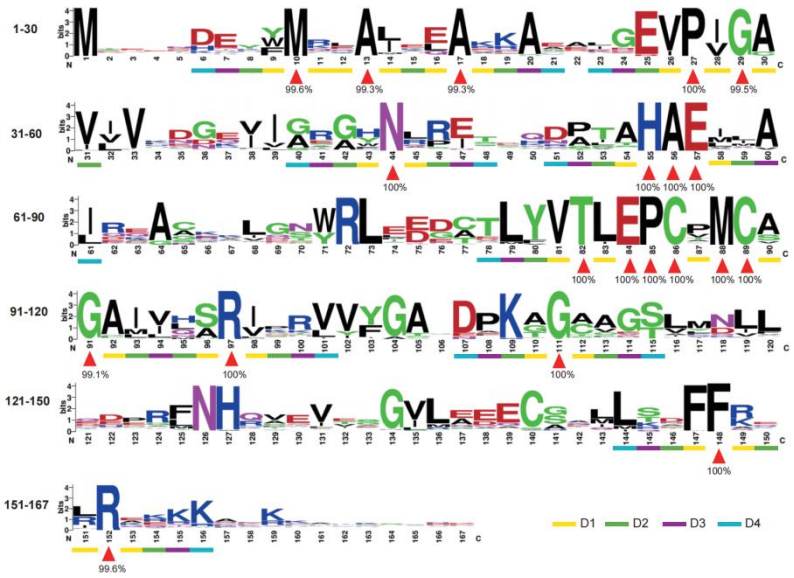
4. Line 214-216, citations are required.

Response: The citations have been added as requested.

5. Extended Data Fig. 2, the R72, L73, N126 looks quite conserved, but why were these not included? The V31 seems to be mislabeled.

Response: Thank you for your reminding. According to multiple sequence alignment, the conservativeness of R72, L73, and N126 is 98.80%, 98.15%, and 97.82%, respectively. Due to the low proportion of other amino acids at these three sites, the weblogo failed to display them.

The color of V31 was indeed labled incorrectly, and we have corrected it.



Extended Data Fig. 2 | Conservative site analysis of TadA protein sequence based on multiple sequence alignment. The red triangle indicates highly conserved amino acids. The yellow, green, purple and blue squares represent adjacent amino acids with different distances from highly conserved sites. The amino acid frequency visualized by webogo(https://weblogo.berkeley.edu/logo.cgi).

6. Line 238, "SAS", Line 244, "CK", Line 247, "TA muscle", full names are required as they are the first time they appear.

Response: Thank you for the valuable suggestions. The full names have been added as requested.

7. Figure 3 and other places, there are at least three kinds of spellings for "C-to-T conversion" in the y axes.

Response: Thanks for carefully reviewing the manuscript. We have revised the figures in the manuscript to ensure consistency.

Reviewer #3 (Remarks to the Author):

The authors have addressed the majority of my main concerns. However, the statistical analysis needs to be revised. The authors should run a normality test (D'Agostini or Shapiro–Wilk should work with their dataset). If the distribution is indeed not normal, they should run a non-parametric test for three or more groups such as a Kruskal-Wallis test, and a suitable posthoc test, such as Dunn's. Non parametric data should be represented as medians using box plots or similar. Mann-Whitney test is used to compare two groups, and therefore introduces error in multiple comparisons.

Response: Thank you for the valuable suggestions.

Following the suggestions, we conducted a normal distribution test on the data using Shapiro Wilk (Table R1) and found that some data samples did not follow a

normal distribution ($p<0.05$).

However, we have concerns about applying the Kruskal Wallis test to this study. A assumption of Kruskal Wallis test is that the observations in each group need to be independent of each other. However, the editing efficiency data of multiple editors at multiple corresponding loci does not meet this assumption. In addition, we also reviewed relevant literature in the field of genome editing and found that Mann-Whitney test, T-test, and Tukey's test are the most commonly used statistical methods in similar situations (Table R2). The Kruskal Wallis test has not been found to be applied in similar studies.

As we all know, T-test and Tukey's test require data to follow a normal distribution, while Mann-Whitney test is suitable for both normal and non-normal distribution data. Therefore, we still believe that the Mann-Whitney test is the more suitable method here. However, in order to control the problem of type I errors during multiple comparisons, we agree to use the commonly used Benjamini-Hochberg procedure to adjust the p values (Fig R2).

Table R1: Shapiro Wilk test of the samples

	P value			
	ACN	**GCN**	**CCN**	**TCN**
aTdCBE	**0.03901**	0.3417	0.612	0.612
Td-CBEmax	**0.02352**	0.2458	0.09948	0.7212
TadCBEd	0.7181	0.7538	0.2483	0.9083
B3PCY2-CBE	**1.453e-05**	**2.051e-05**	**6.066e-06**	**1.074e-06**
YE1-BE4max	0.1262	0.1827	0.5589	0.1335
hA3A*-CBEA	0.09792	0.06712	**0.02263**	0.2301

Table R2: Statistical methods of relevant literature

Mann-Whitney test			
Kleinstiver et al. Nat Biotechnol. 2019	PMID: 30742127	Fig.4	Cas12a
Sakata et al. Nat Biotechnol. 2020	PMID: 32483365	Fig.1	Base editor
Jin et al. Nat Biotechnol. 2021	PMID: 33859403	Fig.4	Prime editor
Lam et al. Nat Biotechnol. 2023	PMID: 36624149	Fig.5	Base editor
Dependent sample t-test			
Zong et al. Nat Biotechnol. 2022	PMID: 35332341	Fig.1-2	Prime editor
Wu et al. Nat Chem Biol. 2023	PMID: 37400536	Fig.2	Cas12f
Tong et al. Nat Biotechnol. 2023	PMID: 35953673	Fig.1	Cas13
Qi et al. Nat Chem Biol. 2024	PMID: 38052959	Fig.3	Cas9
Zhang et al. Nat Commun. 2024	PMID: 38402281	FigS23-S25	CBE6
Yang et al. Nat Chem Biol. 2024	PMID: 38553609	Fig.3	haA3A-CBE
Li et al. Nat Commun. 2024	PMID: 38280857	Fig.3	TnpB
Tukey's test			

Kim et al. Nat Biomed Eng. 2020	PMID: 31937939.	Fig.3	Cas9
Tsuchida et al. Mol Cell. 2022	PMID: 35219382	Fig.5	Cas12d
Hino et al. Cell. 2023	PMID: 37776859	Fig 5	Cas12f

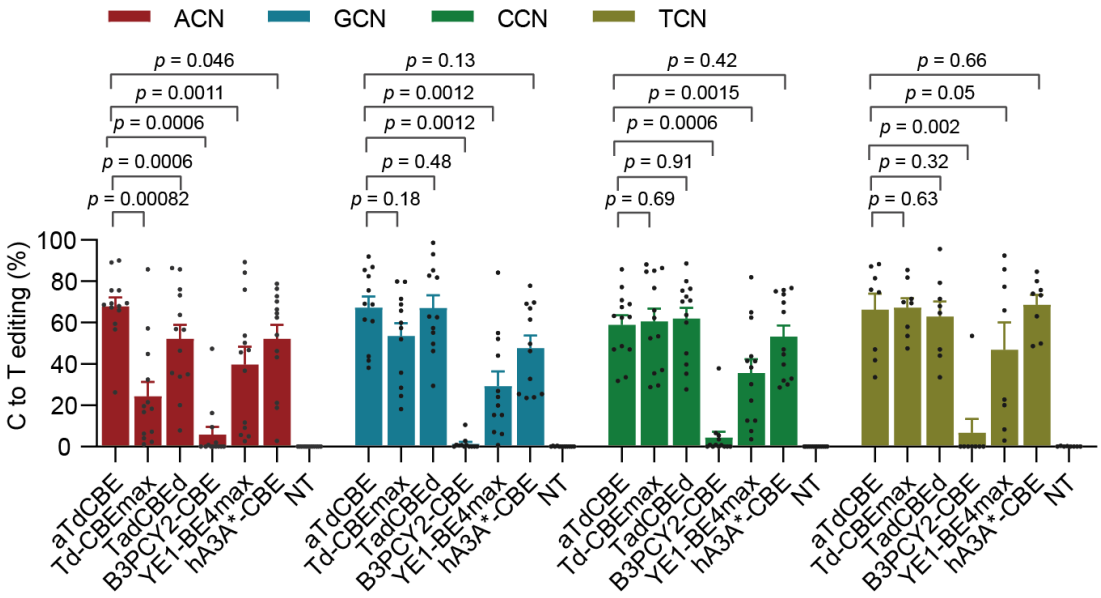


Fig. 3d | Comparison of motif preference for 25 endogenous loci by base editors derived from aTdCBE, Td-CBEmax, TadCBEd, B3PCY2-CBE, hA3A*-CBE and YE1-BE4max in HEK293T cells.The editing window of CBEs was primarily concentrated between bases 4-7. Consequently, the analysis focused on preference motifs exhibiting higher editing efficiency within this range.*P*-values determined by one-sided Mann-Whitney U-test and adjusted by Benjamini-Hochberg procedure.

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

I thank the author for the revised manuscript and for addressing my previous questions. I do not have any additional concerns on top of what I had already said before.

Reviewer #3 (Remarks to the Author):

The authors have addressed my main concerns, and I believe that the manuscript is now suitable for publication.