A Cas-Embedding Strategy for Minimizing Off-Target Effects of DNA Base Editors

Yajing Liu
ShanghaiTech University
Changyang Zhou
Institute of Neuroscience, State Key Laboratory of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Research Cen
Shisheng Huang
University of Shanghai for Science and Technology
Lu Dang
Affiliated Cancer Hospital & Institute of Guangzhou Medical University, 78 Hengzhigang Road, Guangzhou 510095, China.
Yu Wei
Institute of Neuroscience, Chinese Academy of Sciences
Jun He
ShanghaiTech University
Yin's Zhou
Institute of Neuroscience, State Key Laboratory of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Research Cen
shaoshuai Mao
ShanghaiTech University
Wanyu Tao
School of Life Science and Technology, ShanghaiTech University
Yu Zhang
ShanghaiTech University
Hui Yang
Institute of Neuroscience, Chinese Academy of Sciences https://orcid.org/0000-0003-3590-722X
Xingxu Huang
ShanghaiTech University
tian chi (chitian@shanghaitech.edu.cn)
ShanghaiTech University

Article

Keywords: DNA base editors, Cas-embedding, editing enzyme

Posted Date: July 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-46265/v1

Version of Record: A version of this preprint was published at Nature Communications on November 27th, 2020. See the published version at https://doi.org/10.1038/s41467-020-19690-0.

A Cas-Embedding Strategy for Minimizing Off-Target Effects of DNA Base Editors

Yajing Liu[1,2,6], Changyang Zhou[2,6], Shisheng Huang[1,3,6], Lu Dang[4,6], Yu Wei[2,3,6], Jun He[1,3], Yingsi Zhou[2], Shaoshuai Mao[2], Wanyu Tao[1,3], Yu Zhang[1], Hui Yang[2*], Xingxu Huang[1*], Tian Chi[1*]

[1]School of Life Science and Technology, ShanghaiTech University, Shanghai, China.
[2]Institute of Neuroscience, State Key Laboratory of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China.
[3]University of Chinese Academy of Sciences, Beijing, China.
[4]Cancer Hospital & Institute of Guangzhou Medical University, Guangzhou, China
[6]These authors contributed equally to this work.

Abstract

DNA base editors, typically comprising editing enzymes fused to the N-terminus of nCas9, display off-target effects on DNA and/or RNA, which have remained a obstacle to their clinical applications. Off-target edits are typically countered via rationally designed point mutations, but the approach is tedious and not always effective. Here, we report that the off-target effects of both A>G and C>T editors can be dramatically reduced without compromising the on-target editing simply by inserting the editing enzyme into the middle of nCas9 at tolerant sites identified using a transposon-based genetic screen. Furthermore, employing this Cas-embedding method, we have created a highly specific editor capable of efficient C>T editing at methylated and GC-rich sequences.

Adenine base editor (ABE) and cytosine base editor (CBE) can create substantial off-target edits on RNA[1,2], whereas CBE additionally has off-target effects on DNA[1–7]. Such effects result from the intrinsic properties of the editing enzymes harnessed for base editing, and consequently, can be countered by mutating these enzymes (YE1, SECURE-BE, ABE^{F148A} and SECURE-CBE)[1,2,8,9]. However, this approach requires prior knowledge about the enzyme structure and is labor intensive. Interestingly, the editing enzymes (APOBEC1 and Tad-TadA*) produce more off-target edits on RNA in HEK293T cells when expressed as free proteins than as fusion protein to nCas9[2], suggesting that fusing the enzymes to nCas9 may hinder off-target editing perhaps due to steric hindrance. We hypothesized that inserting the enzymes into the middle of nCas9, rather than fusing it to its N-terminus, might further reduce the off-target effects.

To test this idea, we first used MuA-transposon-based genetic screen[10,11] to identify the nCas9 sites tolerant of deaminase insertions. First, we constructed an "all-in-one" plasmid expressing: 1) nCas9 under the control of IPTG-inducible promoter, 2) the ampicillin-resistant (AmpR) gene bearing a C>T mutation that created a premature stop codon (A118X), and 3) a sgRNA under the control of the J23119 promoter for repairing the premature stop codon. The bacteria transformed with this plasmid would be ampicillin sensitive until A>G editing occurs on the bottom strand of DNA, which corrects the C>T mutation in the top strand to restore the translation (Fig. 1a, Supplementary Fig. 1a). We next used Mu-transposon to randomly insert the DNA encoding TadA-TadA* into the "all-in-one" plasmid. This insertion plasmid library, prepared *in vitro*, was then electroporated into *E. coli* and the cells grown overnight on plates containing kanamycin but no ampicillin. We evaluated the insertion efficiencies at various positions on nCas9 by deep-sequencing the unscreened plasmid library extracted from the recovered cells. At the nCas9 coding sequence, we found 51,393 insertions, with at least one insertion in 99% of amino acid positions (Supplementary Fig. 1b), demonstrating that the Mu-mediated mutagenesis was efficient and unbiased. As expected, insertions were hardly detectable in the Kan resistance gene and the replication-related f1

region (Supplementary Fig. 1b). IPTG was added to the mixture to induce the fusion protein expression (nCas9-TadA-TadA*) before the cells were transferred to plates with ampicillin to select the cells expressing the repaired AmpR gene. Positive clones were picked, and the plasmids were extracted and sequenced to determine the editing efficiencies and TadA-TadA* insertion sites (Fig. 1a and Supplementary Fig. 1c).

In total, 43 insertional sites were found on nCas9 by analyzing the extracted plasmids from the recovered ampicillin resistant colonies. Most of the central fusion ABE variants achieve robust A-to-G editing at the premature stop codon (Supplementary Fig. 1c). The ABE variants with TadA-TadA* inserted into these highly tolerant sites are identical to ABEmax[12] except for the location of TadA-TadA*, which was termed CE-ABE ("CE" stands for Cas Embedded). Among the 43 insertion sites recovered in the screened library, nine sites were clustered together within a short (16-aa) segment, occurring at 1048Thr, 1050Ile, 1051Thr, 1052Leu, 1054Asn, 1056Glu, 1057Ile, 1059Lys, and 1063Ile. The enrichment of these sites in the screened library was specific, because in the unselected library, these sites were inserted only 61, 39, 90, 38, 5, 29, 76, 53, and 25 times, respectively, much less frequently than some other sites (e.g, 1090Pro, inserted 280 times) that were not recovered after screening. Thus, the 16-aa fragment was highly tolerant of insertion and presumably dispensable for nCas9 function. Indeed, this fragment, harbored inside the RuvC III domain in the NUC lobe[13], is not conserved among 28 SpCas9 orthologs (Supplementary Fig. 1d). Therefore, we replaced the 1048Thr-1063Ile region with TadA-TadA* to generate CE-ABE$^{1048\text{-}1063}$.

We next tested the 20 most frequently recovered CE-ABEs in HEK293T cells. At an endogenous site containing multiple As within the editing window (ABE-site1), 12 of the 20 CE-ABEs were as active as ABEmax, with editing efficiencies ranging from 66-89% as compared with 86% for ABEmax (Fig. 1b). The editing efficiencies of the various CE-ABEs in HEK293T cells were largely consistent with their recovery rates in

prokaryotic cells, demonstrating the robustness of the screen. We then examined the off-target RNA editing for the 12 variants at 3 RNA off-target sites known to be highly susceptible to ABEmax. All 12 CE-ABEs showed remarkable reductions in the editing on at least two of the three sites, and four of the 12 variants at all 3 sites (Fig. 1c, the 4 variants marked in red). We futher used RNA-seq to profile the off-target edits by these top 4 CE-ABEs at the entire transcriptome, and found that ABEmax induced massive off-target edits as reported before, which was reduced at least 6× in CE-ABEs, with CE-ABE^{1072} as much as 236× (average from 20,739 to 88 SNVs (single nucleotide variants) (Fig. 1d). In contrast, the on-target editing at ABE-site 1 was either uncompromised (92% and 89%, for ABEmax and CE$^{1048\text{-}1063}$-ABE, respectively) or only mildly reduced (to 75% and 73% for CE^{776} -ABE or CE^{1263}-ABE, respectively), except that for CE-ABE^{1072}, the activity was reduced markedly (to 33%) (Fig. 1e).

The data above established CE$^{1048\text{-}1063}$-ABE as the optimal CE-ABE with balanced efficiency and specificity. Therefore, we named it CE-ABE and characterized its performance further, at 8 and 12 randomly selected genomic sites in HEK293T cells (Supplementary Fig. 2a) and mouse N2a cells (Supplementary Fig. 2c), respectively. We found that CE-ABE was comparable to ABEmax in terms of editing rates, while the editing window was slightly enlarged (Supplementary Fig. 2b, d and e). We concluded that embedding an adenine deaminase in nCas9 could markedly reduce the off-target effects on RNA with only a minimal impact on on-target editing.

Encouraged by the success in CE-ABEs, we sought to derive the CE versions of CBEs. We first focused on AncBE4max, a highly active CBE consisting of the APOBEC ancestor Anc689 linked to the N-terminus of nCas9[12]. To create the CE version of AncBE4max (named CE-CBE), we relocated Anc689 to position 1048-1063 in nCas9, replacing the native Cas9 sequence in the process. The on-target editing rates of CE-CBE at 9 target sites tested were comparable to those of AncBE4max (Fig. 2a), as were the editing windows (Supplementary Fig. 3a). We next quantified off-target RNA editing using RNA-seq. Unexpectedly, in contrast to BE3, AncBE4max

produced only 74 edits, presumably reflecting the property of Anc689 harnessed in the editor (Supplementary Fig. 4a). Nevertheless, the off-target effect was clearly above the GFP control, and was completely eliminated in CE-CBE, which produced only 15 edits, a level indistinguishable from the background (Supplementary Fig. 4a). Thus, the CE strategy for RNA off-target reduction is apparently applicable to both ABE and CBE.

In contrast to ABEs, which do not affect the genome, CBES are known to display gRNA-independent off-target editing on the genome. We therefore compared the off-target effects of CE-CBE^{1072}, CE-CBE$^{1048\text{-}1063}$ and AncBE4max using a method termed GOTI (genome-wide off-target analysis by two-cell embryo injection), the highly sensitive and physiologically relevant assay for detecting random genomic off-target edits. In this method, the editors were coexpressed with a sgRNA targeting Tyr, and the DNA from E14.5 embryos was sequenced at a depth of ~30 ×. The CE-CBEs proved as efficient as AncBE4max at on-target editing, as expected (Supplementary Fig. 6a). For off-target editing, AncBE4max created an average 773 SNVs per embryo, 43× above the background in the GFP control (18 SNVs) (Fig. 2c). In sharp contrast, an average of only 19 and 31 SNVs per embryo were detected in the CE-CBE$^{1048\text{-}1063}$ and the CE-CBE^{1072} treated cells, respectively, each similar to GFP control (Fig. 2c, Supplementary Fig. 5). About 92% of the SNVs identified in the AncBE4max-edited cells were mutated from G-to-A or C-to-T; this mutation bias was not observed in GFP-treated cells and reduced to ~57% and ~68% in CE$^{1048\text{-}1063}$-CBE- and CE^{1072}-CBE-treated embryos, respectively (Fig. 2d, Supplementary Fig. 5). Those results demonstrate that the embedding strategy could dramatically decrease DNA off-target edits for CBEs.

One of the most active CBEs created so far is A3A-BE comprising APOBEC3A fused to nCas9[14], but unfortunately, A3A-BE also displays the highest gRNA-independent off-target editing on DNA[9]. We have previously found that introducing Y130F into A3A can only partially mitigate the off-target effect, with substantial off-target edits persisting in A3A (Y130F)-BE3, which limits the use of the editor[8,14]. We hypothesized that

111 embedding A3A (Y130F) in nCas9 might help minimize the off-target effect. Therefore, we replaced rA1 of
112 $CE^{1048\text{-}1063}$-CBE with A3A (Y130F), generating $CE^{1048\text{-}1063}$-A3A(Y130), which we compared with A3A (Y130F)-
113 BE4max we generated in this study for the comparison. At 8 different human genes , the two editors displayed
114 similarly high editing rates (Fig. 2b), similar editing windows and similarly neglible off-target effects on RNA
115 (Supplementary Fig. 3b, 4b). In contrast, $CE^{1048\text{-}1063}$-A3A(Y130F) created an average of only 43 off-target edits
116 on DNA, much fewer than BE3-A3A (Y130F) (276 SNVs) and only slightly above the GFP background (18
117 SNVs). We also tested CE^{1072}-A3A , which carried A3A (Y130F) inserted into the 1072 site in nCas9, and
118 observed similarly low numbers of off-target edits (63 SNVs) (Fig. 2c, d and Supplementary Fig. 5). $CE^{1048\text{-}1063}$-
119 A3A(Y130) is thus a highly active CBE that is also highly specific (Supplementary Fig. 6).

120 A unique advantage of BE-A3A is that they can efficiently edit GC-rich and highly methylated regions,
121 contrary to the traditional CBEs utilizing rA1[14]. To determine whether this important advantage of A3A-CBE over
122 traditional CBEs is retained in $CE^{1048\text{-}1063}$-A3A(Y130), we sought to benchmark $CE^{1048\text{-}1063}$-A3A(Y130) against
123 YE1-BE4max, the most active CBE we and others reported[8,9]. We compared the two editors in HEK293T cells at 3
124 highly methylated target genes . Indeed, $CE^{1048\text{-}1063}$-A3A(Y130F) markedly outperformed YE1-BE4max at all the 3
125 sites (Fig. 2e), confirming that $CE^{1048\text{-}1063}$-A3A(Y130F) was preferrable to YE1-BE4max for editing methylated
126 (and by inference, GC-rich) regions. Of note, at these sites, $CE^{1048\text{-}1063}$-A3A(Y130F) was as active as BE4max-
127 A3A(Y130F) but more active than BE3-A3A(Y130F), as expected (Fig. 2e).

128 This study advances the field in three ways. First, we show that CE is a powerful countermeasure of off-
129 target effects of DNA base editors, which is much simpler than the conventional mutagesis method. Although the
130 off-target effects can also be countered by point mutations, these mutations are identified on a case-by-case basis,
131 typically via extensive structure-function analysis, contrary to CE which requires no prior knowledge about the
132 enzymes. We are aware that CE may be insufficient for fully eliminating off-target effects for some editors, but so
133 may be point mutations, and in this case, the two approaches may be used in conjunction to achieve the optimal

effect, as illustrated in CE-A3A(Y130F) where CE and Y130F act together to minimize the off-targets on DNA. Of note, we have recently found that inserting a deaminase domain into a flexible loop on the surface of dCasRx could slightly reduce the off-target effects (~1.8 x) of the RNA base editor on RNA. The data suggest that the CE method is perhaps generalizable to RNA base editors, assuming transposon-based screen on dCasRx can similarly yield optimal insertion sites analogous to the ones discovered in nCas9. The second advance made in this study lies in the creation of $CE^{1048\text{-}1063}$-A3A(Y130F) , a useful reagent for highly specific and efficient editing at GC and methylated sites. Finally, we have developed a universal MuA transposase-based platform for identifying nCas9 sites tolerant of deaminase insertion, which may be useful for optimizating other base editors such as the primer editor.

References

1. Grünewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
2. Zhou, C. *et al.* Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* (2019). doi:10.1038/s41586-019-1314-0
3. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
4. Zuo, E. *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science (80-.).* **364**, 289–292 (2019).
5. Jin, S. *et al.* Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science (80-.).* **364**, 292–295 (2019).
6. 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).
7. Gapinske, M. *et al.* CRISPR-SKIP: Programmable gene splicing with single base editors. *Genome Biol.* **19**, 1–11 (2018).
8. Zuo, E. *et al.* A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* **17**, 600–604 (2020).
9. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).
10. Haapa, S., Taira, S., Heikkinen, E. & Savilahti, H. An efficient and accurate integration of mini-Mu transposons in vitro: A general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Res.* **27**, 2777–2784 (1999).
11. Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34**, 646–651 (2016).

12. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–848 (2018).

13. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).

14. Wang, X. *et al.* Efficient base editing in methylated regions with a human apobec3a-cas9 fusion. *Nat. Biotechnol.* **36**, 946 (2018).

15. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).

16. Liu, Y *et al*. REPAIRx, a specific yet highly efficient programmable A>I RNA base editor. (*The EMBO Journal* accepted)

Materials and Methods

Plasmid construction. Primers and plasmids are listed in Supplementary Table 1. pCMV-nCas9-KanR-AmpR(A118X)-sgRNA, the all-in-one plasmid for insertion screening, was assembled from pCMV-ABEmax (Addgene 112095), pUC57-Kan (Addgene, 51132) and pGL3-U6-sgRNA (Addgene, 51133). The sgRNA expression vector for mammalian cells was constructed using BsaI-digested pGL3-U6-sgRNA-EGFP with annealed DNA oligos (Supplementary Table 1). sgRNA expression vector for GOTI was constructed by cloning annealed DNA oligos (Supplementary tables 1) into BbsI-digested pUC57-sgRNA (Addgene, 51132). CE-ABE and CE-CBE variants were derived from pCMV-ABEmax (Addgene 112095) and pCMV-AncBE4max (112094), respectively.

Transposon library construction and insertion library screening. TadA-TadA* was PCR-amplified from pCMV-ABEmax (Addgene 112095) and cloned into a MuA transposon vector (Supplementary Fig. 1a). The transposon was excised from the vector using BsaI digestion before random insertion into pCMV-nCas9-KanR-AmpR(A118X)-sgRNA in an *in vitro* reaction containing 250 ng transposon, 500 ng nCas9 plasmid and 1 μl of MuA Transposase (F-701, Thermo Fisher) (Supplementary Fig. 1a). The reaction was incubated at 30ºC for one hour to achieve random insertion, followed by 75ºC for 10 minutes to inactivate the MuA Transposase. The DNA was precipitated, resuspended in 5 μl deionized water and electroporated into 100 μL BL21(DE3) electrocompetent cells (Shanghai Weidi Biotechnology, EE1002). A total of 1 ml SOC media was then added and the bacteria were cultured at 37ºC for one hour. The cells were then plated out on several LB agar plates containing 10 μg/mL Kanamycin and incubated at 37ºC for 16 hours. The colonies were then collected by scraping and resuspended in 100 mL LB containing 500 μM IPTG. The culture was incubated for 10-12 hours to induce nCas9 expression and repair the AmpR (A118X) mutation. Decreasing amounts of cells (from 5 mL, 1 mL, 500 μL, 100 μL of culture) were then plated out on 15-cm LB agar plates supplemented with ampicillin (10 μg/mL)

and kanamycin (10 μg/mL). After incubation overnight, the colonies were picked and subjected to Sanger sequencing to evaluate base editing at AmpR(A118X) and to determine the TadA-TadA* insertion sites.

Library sequencing and analysis. The constructed transposon library was sequenced using an Illumina HiSeq X Ten (2 × 150 PE) at the Novogene Bioinformatics Institute, Beijing, China. All cleaned reads were first mapped to the backbone sequence using BWA v0.7.16 with the default parameters. The soft clipped reads were extracted and then mapped to the insertion sequence. All mapped soft clipped reads were checked and the breakpoints were recorded as insertion sites.

Cell culture and transfection. HEK293T (ATCC CRL-3216) and Neuro-2a (N2a) (ATCC HTB-96) cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Hyclone, SH30243.01) supplemented with 10% fetal bovine serum (FBS)(v/v) (Gemini, 900-108) and penicillin and streptomycin (Gibco, 15140122). Cells were passaged once every three days and incubated at 37ºC with 5% CO_2. All cells used in experiment have been tested to exclude mycoplasma contamination.

To evaluate the CE-ABE variants, HEK293T cells were seeded on poly-D-lysine (Sigma, P4707) coated 12-well plates (JET-BIOFIL, TCP010012) and transfected about 14 hours thereafter at approximately 80% density per the manufacturer's protocols (ThermoFisher Scientific, 11668019). To facilitate the detection of off-target RNA editing effects by CE-ABEs, we used "all-in-one" plasmids that expressed CE-ABEs-P2A-GFP together with a sgRNA for the specified target site. In brief, the plasmid (1 μg) was transfected and 48 hours later, the cells with the highest 5% of GFP signal were isolated by FACS for further analysis (for Fig. 2, Fig. 3a-c). To evaluate CE-ABE and CE-CBE (Fig. 2e, Fig3a-b), the editor-expressing vectors (700 ng) were co-transfected with corresponding sgRNA-GFP plasmids (300 ng), and all GFP positive cells were isolated by FACS for further analysis.

On-target genome editing. Cells were isolated as described above before genomic DNA was extracted using QuickExtract™ DNA Extraction Solution (Lucigen). The fragments encompassing the target sites (~200-bp) were PCR-amplified using Phanta Max Super-Fidelity DNA polymerase (Vazyme, P505-03); the primers used are listed in Supplementary Table 1. The amplicons were analyzed by deep sequencing on the Illumina Hiseq X Ten (2×150 PE) platform. The adapter pair of the paired-end reads were removed using AdapterRemoval version 2.2.2, and paired-end read alignments of 11 bp or more bases were combined into a single consensus read. All the processed reads were then mapped to the target sequences using the BWA-MEM algorithm (BWA v0.7.16). For each site, the mutation rate was calculated using bam-read count with parameters -q 20 -b 30. Indels were calculated based on the reads containing at least one inserted or deleted nucleotide in the protospacer. Indel frequency was expressed as the number of indel-containing reads/total mapped reads.

Editing at the predicted off-targets at RNA. Cells (at least 30,000) were isolated as described above, lysed in TRIzol Reagent (Invitrogen, 15596026) and total RNA was extracted. mRNA was reverse transcribed using HiScript II Q RT SuperMix kit (Vazyme, R223-01) and the predicted off-targets were amplified using Phanta Max Super-Fidelity DNA polymerase (Vazyme, P505-03) with the primers listed in Supplementary Table 1. The amplicons were analyzed by Sanger sequencing and the editing rates were calculated using the following tool: https://moriaritylab.shinyapps.io/editr_v10/.

RNA-seq analysis of global off-target edits at the transcriptome. All RNA samples were sequenced using an Illumina HiSeq X Ten (2 × 150 PE) at the Novogene Bioinformatics Institute, Beijing, China, 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 using GATK HaplotypeCaller (version 4.1.2) and filtered with QD (Quality by Depth) < 2. All variants were verified and quantified by bam-readcount with parameters -q 20 -b 30. The depth for a given edit had to be at

least 10× 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 (for ABEs) or C-to-T (for CBEs) edits in the transcribed strand were considered for downstream analysis.

***In vitro* transcription and GOTI.** Linear templates bearing 20-bp sgRNA genes downstream of the T7 promoter were transcribed *in vitro* with the TranscriptAid T7 High Yield Transcription Kit (ThermoFisher Scientific) using the primers listed in the Supplementary Table 1. sgRNAs were purified using the MEGAclear Kit (ThermoFisher Scientific). Linear templates bearing AncBE4max or CE-CBE variants downstream of the T7 promoter were transcribed using mMESSAGE mMACHINE T7 Ultra Kit (Ambion, Life Technologies, AM1345) and the mRNA was purified using RNeasy Protect Mini Kit (QIAGEN, 74124). The RNAs were quantified by UV absorbance before microinjection. In a previous study, we developed the GOTI (genome-wide off-target analysis by two-cell embryo injection) system to evaluate the off-target effects induced by editors.

In this study, we used an improved GOTI system, which employs wild type B6D2F1(C57BL/6 ×DBA2J) mouse embryo. In brief, we separated one blastomere of two-cell embryos and transferred it to another recipient zonae pellucidae, then injected CBE, together with sgRNA, into one of the embryos. Then we transplanted the embryo into recipient mouse with another ICR (albino) embryo to facilitate growth of edited embryos. After 13.5 days of transplantation, we collected all embryo and picked out B6D2F1 mouse embryo based on the eye color. Genomic DNA was extracted from embryos using the DNeasy blood and tissue kit (catalog number 69504, Qiagen) according to the manufacturer's instructions. Part of genomic DNA was using for on-target editing analysis, and the remaining genomic DNA was for WGS as detailed below.

WGS was performed at mean coverage of 30× using an Illumina HiSeq X Ten. Bioinformatic analysis was performed as previously described[4]. In brief, raw sequencing reads were trimmed with Trimmomatic (v0.36) and

duplicates were removed using Sambamba (v0.6.7) before mapping qualified reads to the mouse reference genome (mm10) using BWA (v0.7.16). Three algorithms, Mutect2 (v3.5), Lofreq (v2.1.2) and Strelka (v2.7.1), were used to identify *de novo* variants, with the paired non-injected sample in the same embryo serving as a control. The subset of the SNVs reported by all three algorithms were considered the true variants.

Statistical analysis

No statistical methods were used to predetermine sample size for in vitro or in vivo experiments. All values are shown as mean ± s.e.m. Unpaired Student's t-test (two-tailed) was used for comparisons and $p<0.05$ was considered to be statistically significant.

Acknowledgements

We thank Haibo Zhou for discussion, and Yidi Sun for bioinformatics supporting. This work is supported by National Key R&D Program (2016YFA0500903), National Science Foundation of China (81830004).

Author Contributions

Y.L., C.Z., S.H., H.Y., X.H., and T.C., designed and conceived the project. Y.L., C.Z., and S.H. designed the random insertion screening system. Y.L., L.D. and W. T. carried out the cell experiments with support from S.M and Y.Z.. Y.W., C.Z. and Y.L. performed the GOAT experiments. S.H. performed whole genome sequencing and RNA sequencing data analysis. J.H., performed the Cas9 homology analysis. Y.S.Z. assisted with data analysis. Y.L. and T.C. wrote the manuscript with input from S.H. and C.Z..

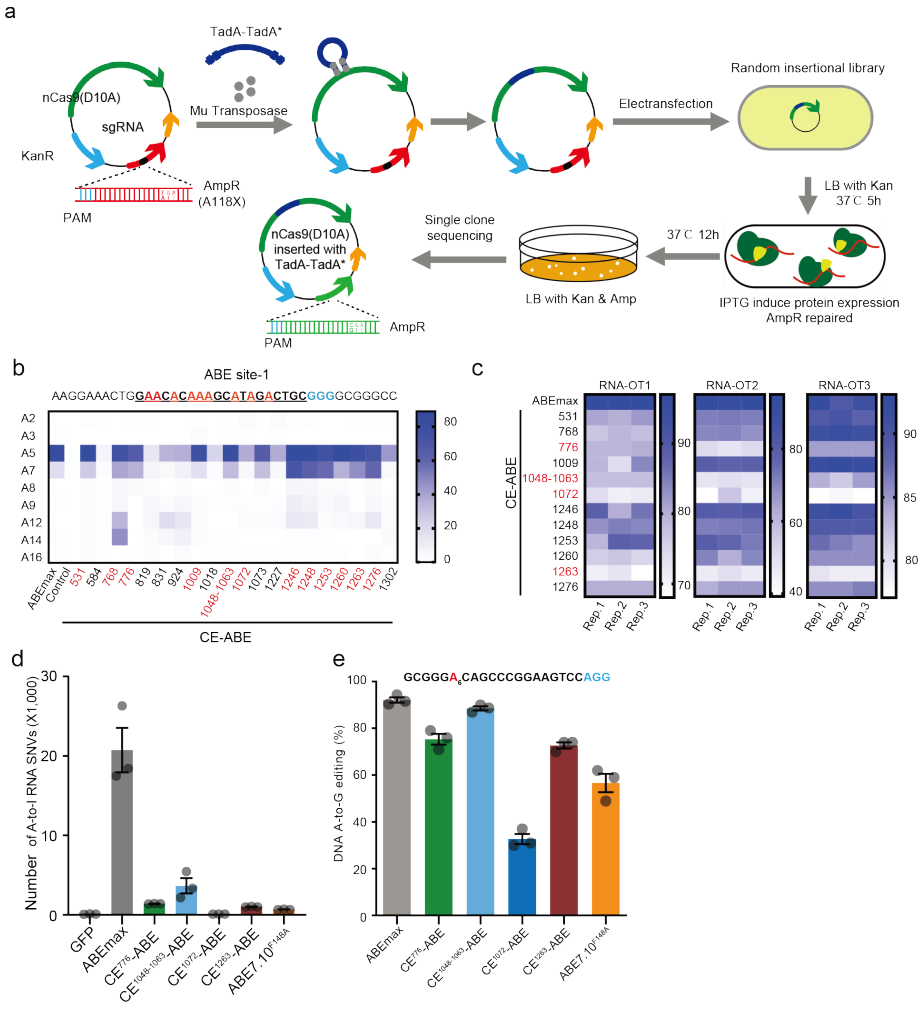


Figure legends

Figure 1. Genetic screen for tolerant sites on nCas9 using transposon-mediated random mutagenesis. a, Workflow of the genetic screen. The sequence of A118X at ampicillin-resistance gene (AmpR) is shown in Fig. S1c. **b**, On-target editing by 20 CE-ABE variants at the ABE-site 1 in HEK293T cells. The heatmap shows the editing rates averaged from three biological replicates. The variants as active as ABEmax were highlighted in red. CE-ABE variants were identical to ABEmax except for the locations of the editing moiety. **c**, Off-target editing by 12 CE-ABE variants at three known RNA sites of ABEmax[15]. The 12 CE-ABE variants were selected from Fig. 1a. Editing rates of three biological replicates are displayed. Four variants with the least off-target effects (red) were selected for further analysis. **d**, Off-target editing by four CE-ABE variants and ABE^{F148A} at the

transcriptome. **e**, and their on-target editing in the same samples, in three biological replicates. The editors were co-expressed with an ABE-site[15] gRNA and GFP, the cells with top 5% GFP fluorescence were sorted for analysis. n, total number of modified adenines.

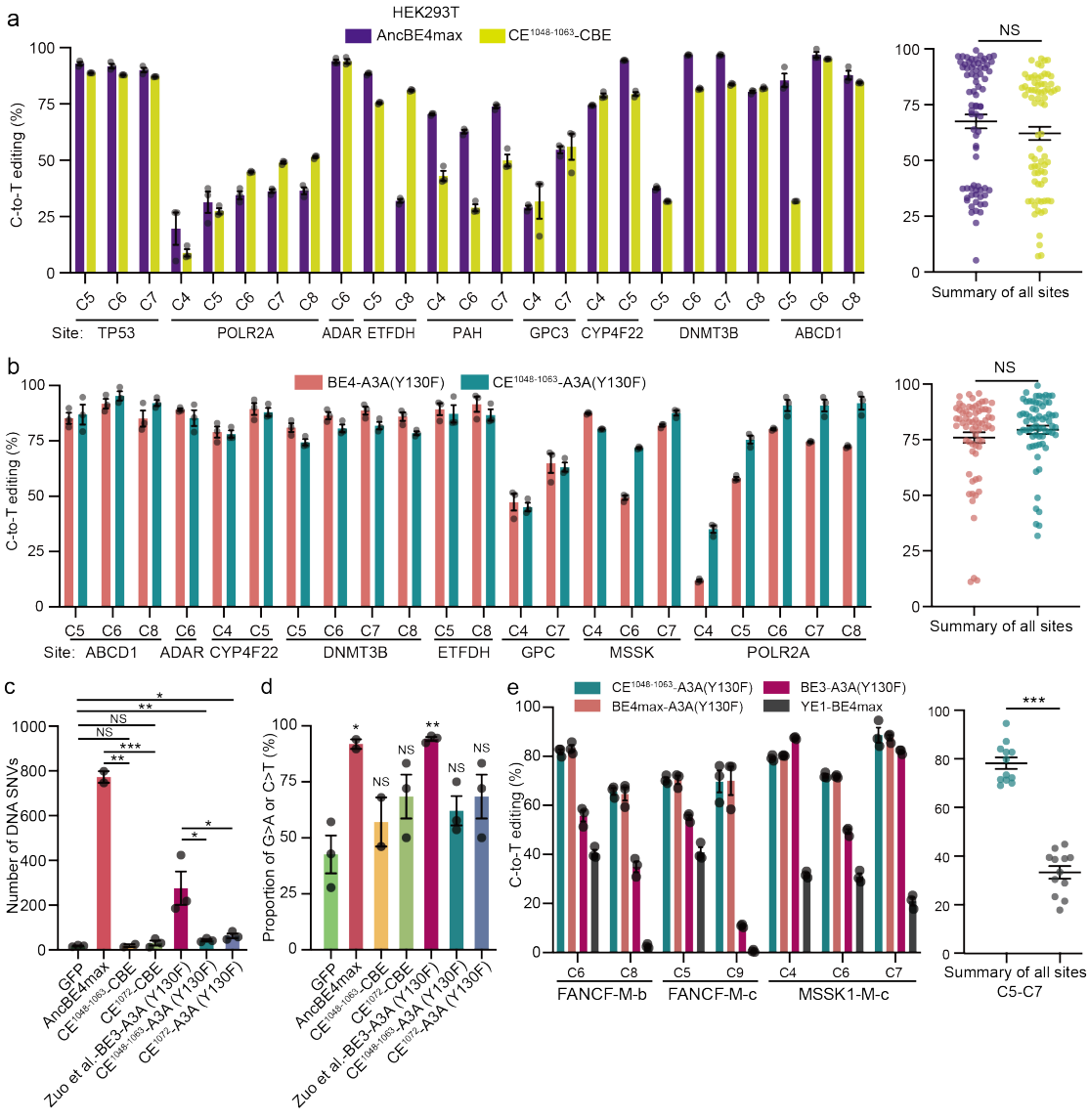
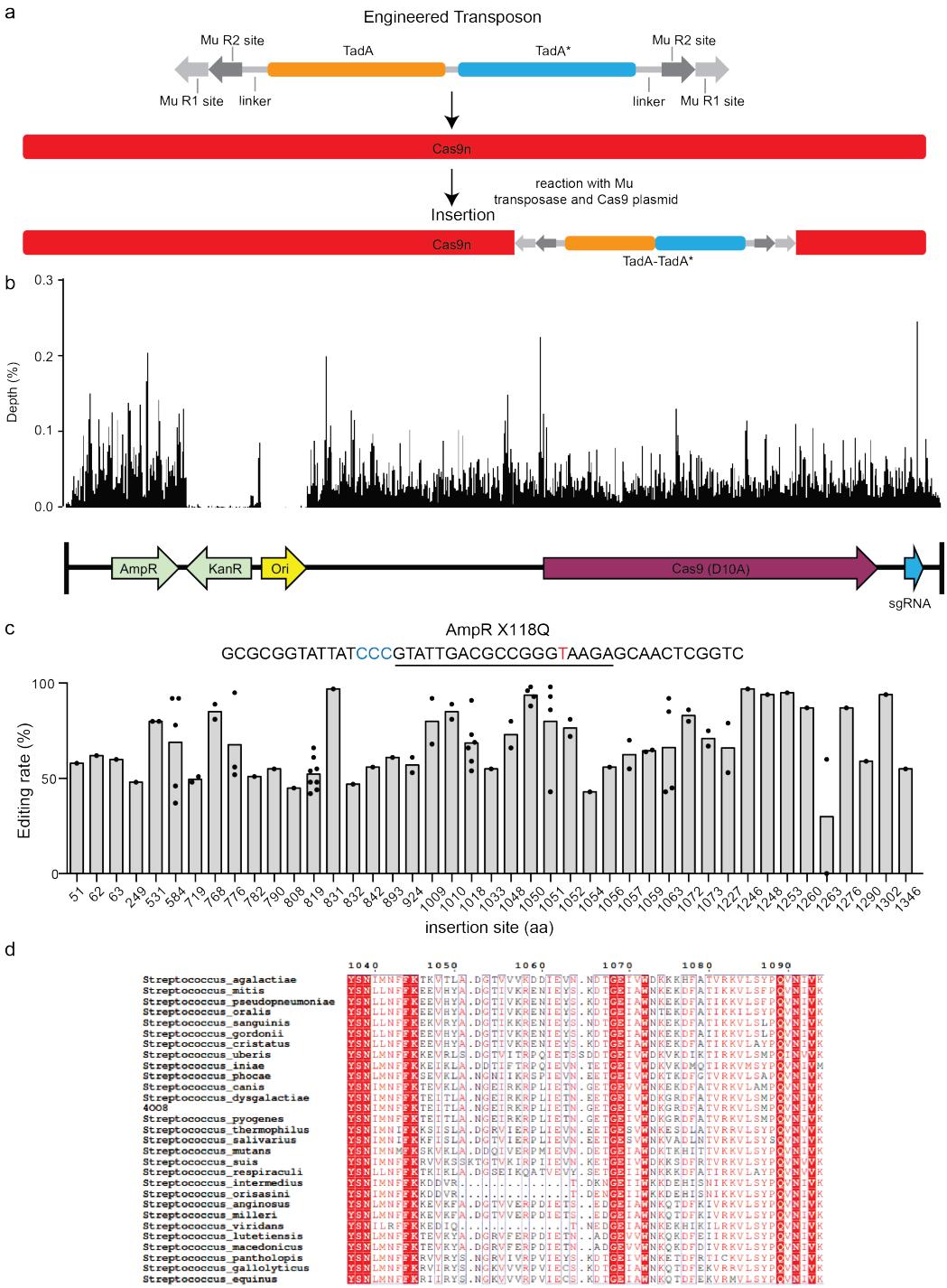


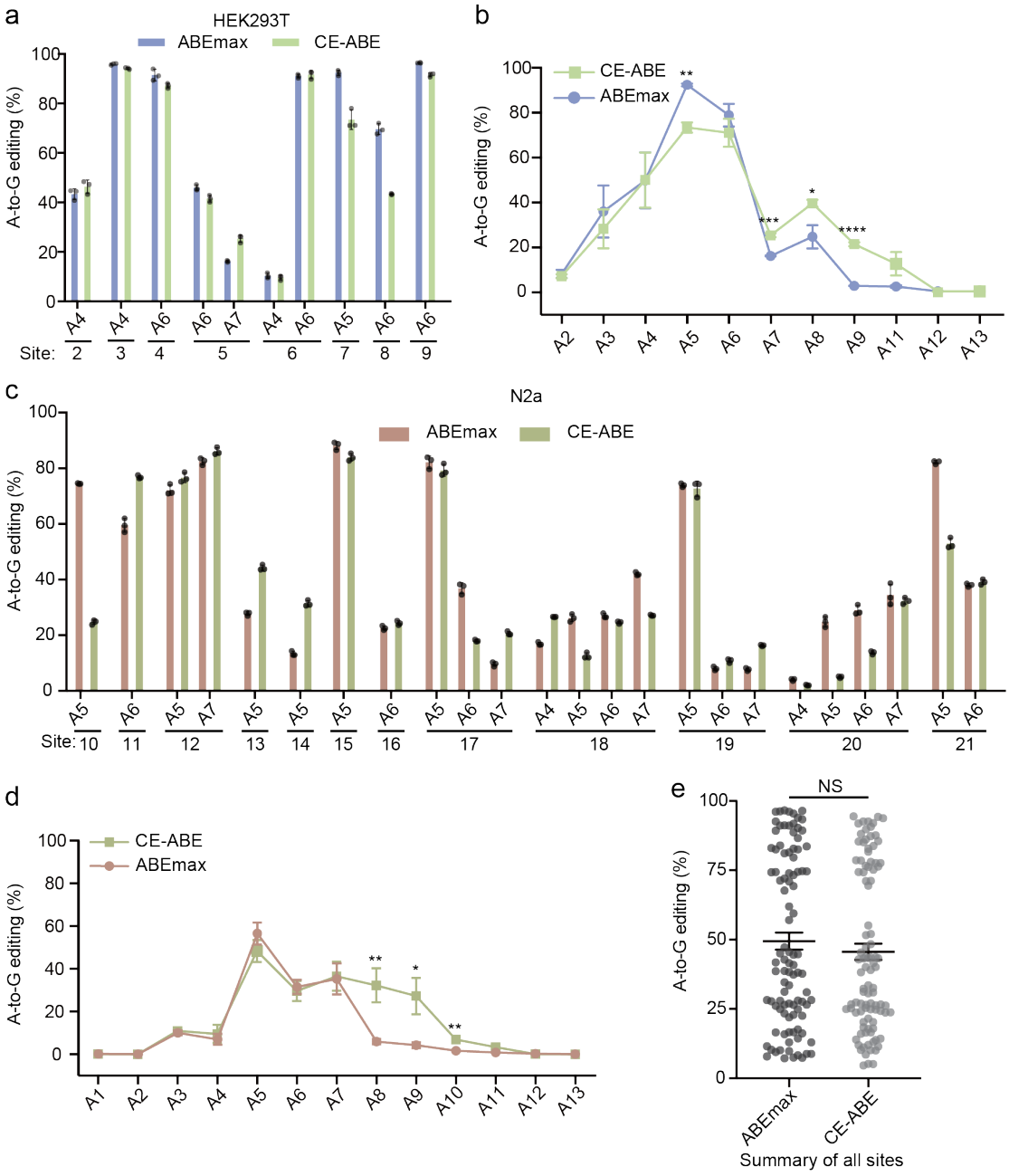
Figure 2. The CE strategy is applicable to CBE: generation and characterization of CE-CBE in HEK293T cells. a, On-target editing of CE-CBE and AncBE4max at nine genomic sites in HEK293T cells. Left: editing efficiencies at various sites; right: summary of all tested sites. GFP-positive cells were analyzed. Values are mean +/-s.e.m. from three biological replicates. **b**, On-target editing of CE-A3A and A3A-BE4max at seven genomic sites in HEK293T cells. Left: editing efficiencies at various sites; right: summary of all tested sites. GFP-positive cells

were analyzed. Values are mean +/-s.e.m. from three biological replicates. **c-d. Total numbers (c) and the** proportions of G>A or C>T mutations (d) for the off-target SNVs detected by GOTI in embryos treated with the indicated editors. The data were all generated in this study except for BE3-A3A(Y130F), which was from reference #8. Values are mean +/-s.e.m. (n=2 to 3). **e**, Editing efficiencies of the indicated editors at methylated genomic sites[14]. Values are mean +/-s.e.m. from three biological replicates.



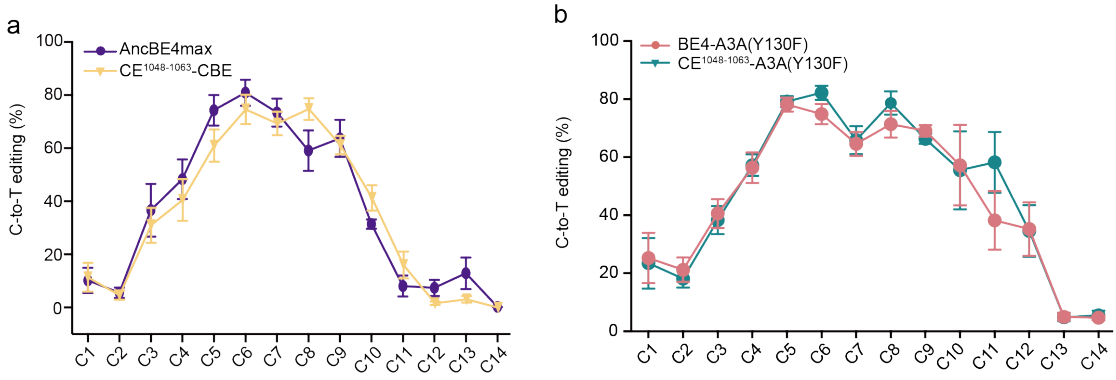
Supplementary Figure 1. Genetic screen for tolerant nCas9 sites. a, Mu transposase catalyzed insertion of TadA-TadA* into nCas9 in vitro. **b,** Insertion site distribution pattern across the entire plasmid in unscreened library, revealed by deep-seq. **c,** nCas9 insertion sites in the screened library and the corresponding A>G conversion rates

at the Amp resistance gene in the same plasmid, as revealed by Sanger sequencing. Each dot represents a single Amp-resistant colony. **d,** Sequence alignment of SpCas9 (1036 to 1096) among 28 orthologs.

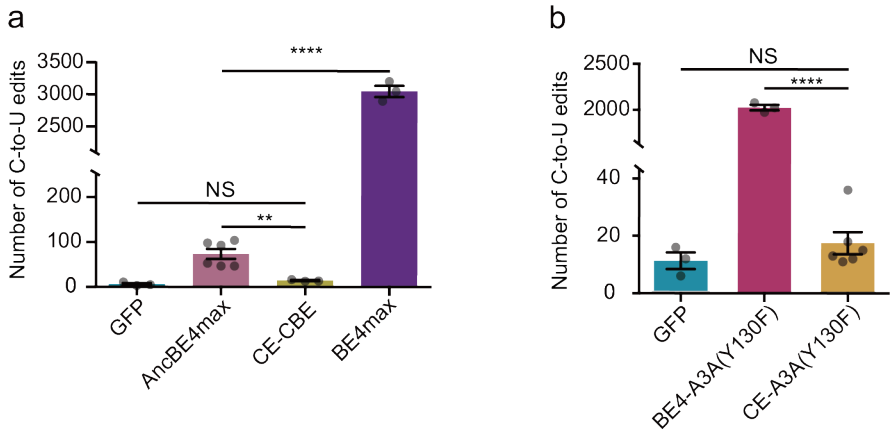


Supplementary Figure 2. On-targeting editing properties of ABEmax and CE-ABE. a, Editing rates at most susceptible As at eight genomic sites in HEK293T cells. **b**, Editing rates at all the As within the editing window, as averaged from the eight sites in HEK293T cells. **c**, Editing rates at the most susceptible As at 12 genomic sites in the mouse N2a cells. **d**, Editing rates at all the As within the editing window, as averaged from the 12 sites in mouse

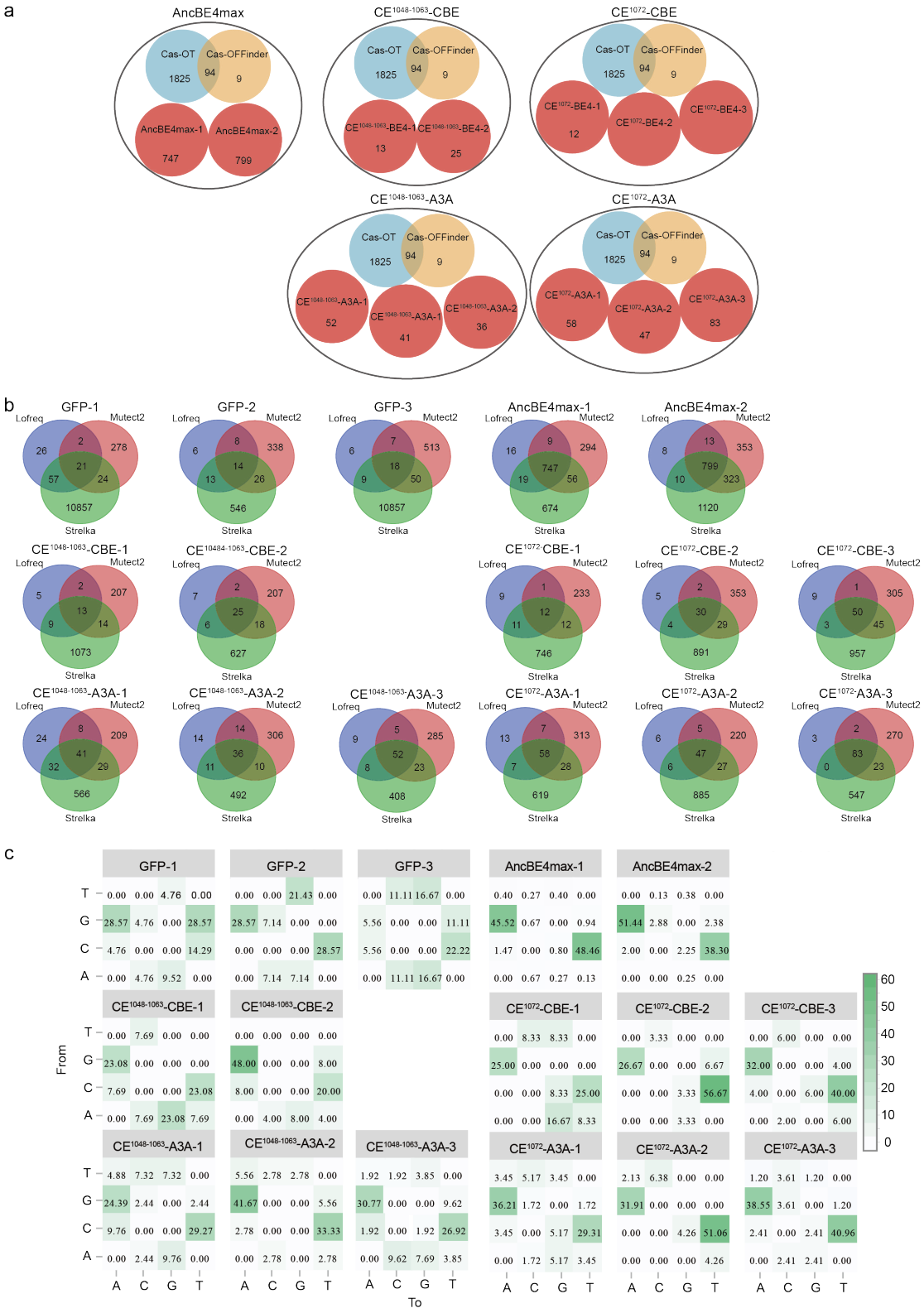
N2a cells. In **a** and **c**, cells with top 5% GFP fluorescence were analyzed, and values are mean +/-s.e.m. from three biological replicates. **e**, Editing frequencies at all the sites shown in **a** and **c**. n = 20 samples from three independent experiments.



Supplementary Figure 3. CE did not alter the editing window. a and b are the same as Fig. 2a and Fig. 2b, respectively, except that the average editing rates for all the Cs within the editing windows (C1-C14) are shown. Values are mean +/-s.e.m. (n=3 biological replicates).

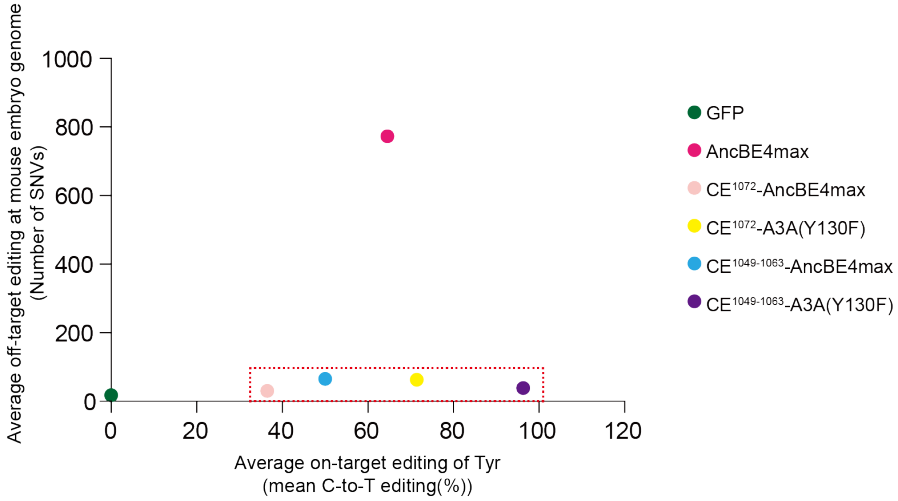


Supplementary Figure 4. CE suppressed RNA off-target effects of CBE. a, RNA off-target SNVs created by AncBE4max and its CE version (CE-CBE). BE4max was included as a control. **b,** RNA off-target SNVs created by BE4-A3A(Y130F) and its CE version (CE-A3A (Y130F)). Cells with top 5% GPF fluorescence was analyzed 48 h post-transfection. Values are mean +/-s.e.m. (n=3 biological replicates).



Supplementary Figure 5. Characteristics of CBE variants induced off-target SNVs. **a,** The SNVs detected by GOTI in **Fig. 2c** overlaps with the predicted off-target by Cas-OFFinder and CRISPOR. **b,** Venn diagram of the

SNVs discovered in each embryo through WGS data. **c,** Distribution of SNV types for samples transfected with GFP, AncBE4max or one of the four CBE variants.



Supplementary Figure 6. On-target editing rates and off-target SNVs for all the variants in mouse embryos, as determined by the GOTI assays. The result shows that $CE^{1049\text{-}1063}$-A3A(Y130F) displays the best balance of editing efficiency and specificity.

Figures

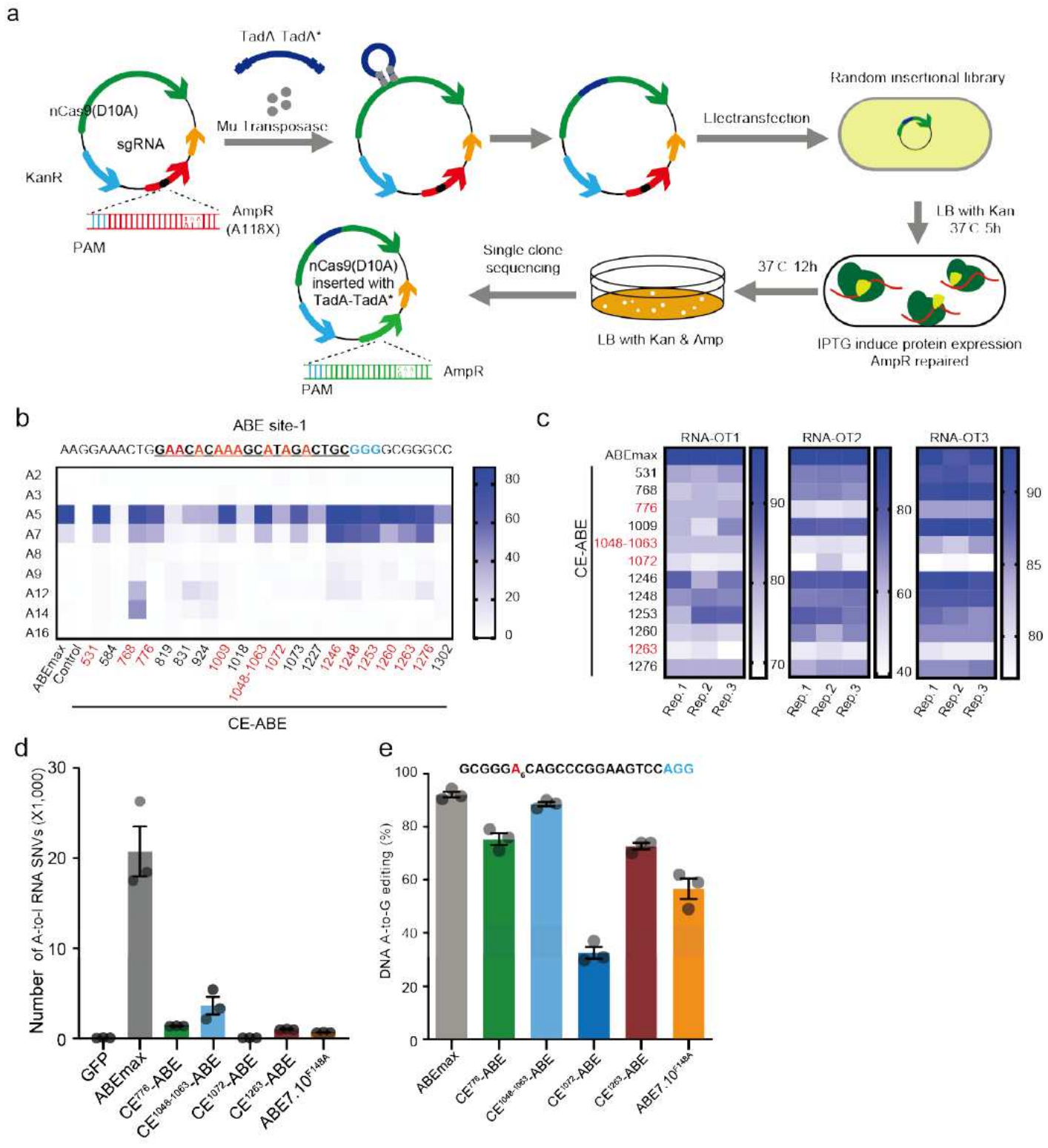


Figure 1

Genetic screen for tolerant sites on nCas9 using transposon-mediated random mutagenesis. a, Workflow of the genetic screen. The sequence of A118X at ampicillin-resistance gene (AmpR) is shown in Fig. S1c. b, On-target editing by 20 CE-ABE variants at the ABE-site 1 in HEK293T cells. The heatmap shows the

editing rates averaged from three biological replicates. The variants as active as ABEmax were highlighted in red. CE-ABE variants were identical to ABEmax except for the locations of the editing moiety. c, Off-target editing by 12 CE-ABE variants at three known RNA sites of ABEmax15. The 12 CE-ABE variants were selected from Fig. 1a. Editing rates of three biological replicates are displayed. Four variants with the least off-target effects (red) were selected for further analysis. d, Off-target editing by four CE-ABE variants and ABEF148A at the transcriptome. e, and their on-target editing in the same samples, in three biological replicates. The editors were co-expressed with an ABE-site15 gRNA and GFP, the cells with top 5% GFP fluorescence were sorted for analysis. n, total number of modified adenines.

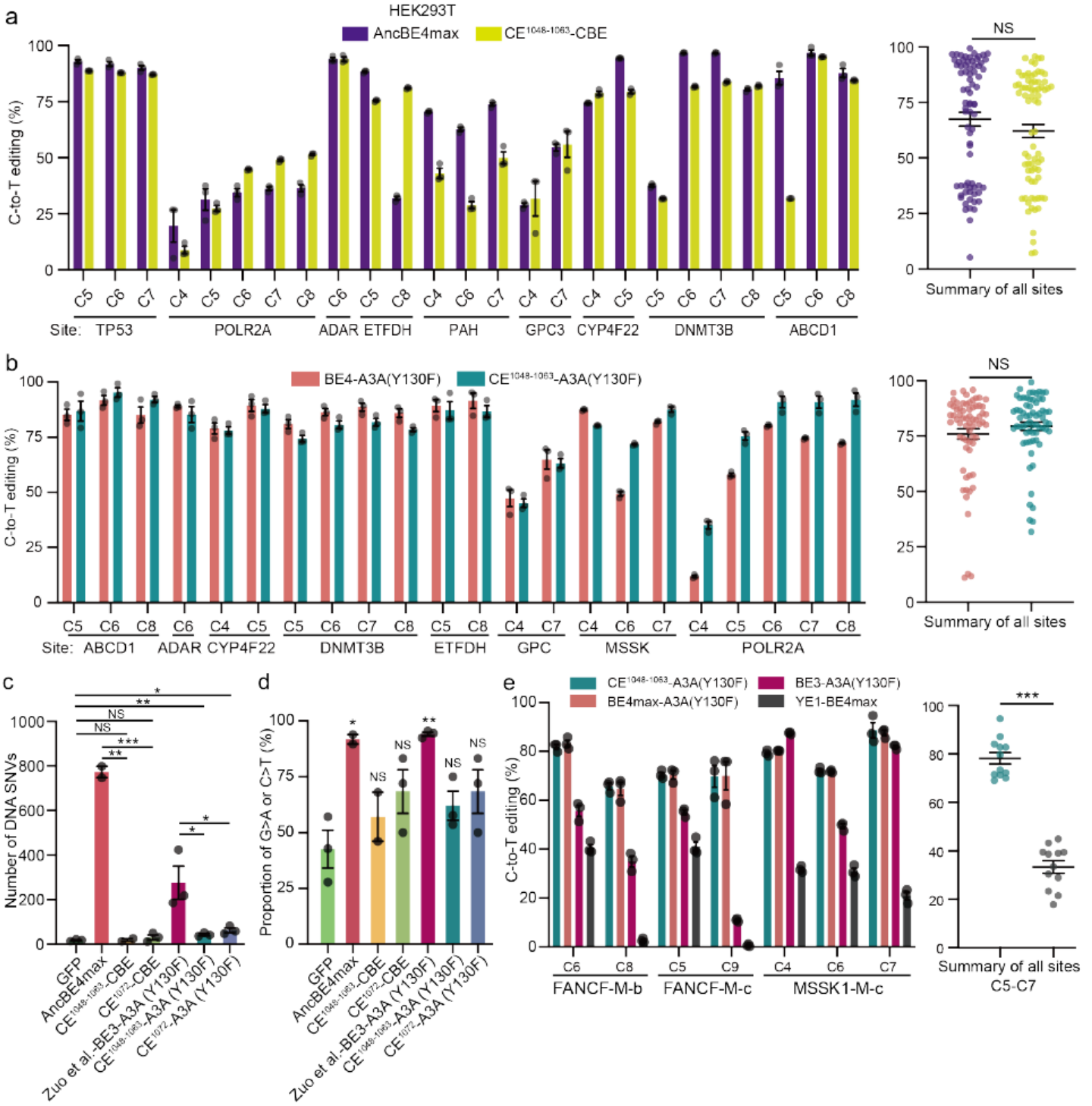


Figure 2

The CE strategy is applicable to CBE: generation and characterization of CE-CBE in HEK293T cells. a, On-target editing of CE-CBE and AncBE4max at nine genomic sites in HEK293T cells. Left: editing efficiencies at various sites; right: summary of all tested sites. GFP-positive cells were analyzed. Values are mean +/-s.e.m. from three biological replicates. b, On-target editing of CE-A3A and A3A-BE4max at seven genomic sites in HEK293T cells. Left: editing efficiencies at various sites; right: summary of all tested sites. GFP-positive cells were analyzed. Values are mean +/-s.e.m. from three biological replicates. c-d. Total numbers (c) and the proportions of G>A or C>T mutations (d) for the off-target SNVs detected by GOTI in embryos treated with the indicated editors. The data were all generated in this study except for BE3-A3A(Y130F), which was from reference #8. Values are mean +/-s.e.m. (n=2 to 3). e, Editing efficiencies of the indicated editors at methylated genomic sites14. Values are mean +/-s.e.m. from three biological replicates.

Supplementary Files

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

- SupplementalTable1.xlsx