BRIEF COMMUNICATION
https://doi.org/10.1038/s41592-020-0832-x

nature methods

Check for updates

A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects

Erwei Zuo[1,2,10]✉, Yidi Sun[3,4,10], Tanglong Yuan[1,10], Bingbing He[2,10], Changyang Zhou[2,10], Wenqin Ying[2], Jing Liu[1], Wu Wei[4,5], Rong Zeng[3,6], Yixue Li[4,6,7,8]✉ and Hui Yang[2,9]✉

Cytosine base editors (CBEs) offer a powerful tool for correcting point mutations, yet their DNA and RNA off-target activities have caused concerns in biomedical applications. We describe screens of 23 rationally engineered CBE variants, which reveal mutation residues in the predicted DNA-binding site can dramatically decrease the Cas9-independent off-target effects. Furthermore, we obtained a CBE variant—YE1-BE3-FNLS—that retains high on-target editing efficiency while causing extremely low off-target edits and bystander edits.

Base editors hold great potential for correcting pathogenetic mutations[1]. CBEs use the rat apolipoprotein-B-editing enzyme, catalytic polypeptide-1 APOBEC1 (rAPOBEC1), a cytidine deaminase binding to single-stranded DNA (ssDNA) and is fused to nCas9, to convert cytosine to thymine. However, previous studies have shown unwanted off-target edits of DNA and RNA introduced by CBEs[2–6]. Here, we analyzed the DNA and RNA off-target effects of engineered CBE variants using genome-wide off-target analysis by two-cell embryo injection (GOTI)[2] and RNA sequencing, respectiviely.

Guided by structural and biochemical insights from previous studies[7–11], we designed and introduced various mutations in rAPOBEC1 for tuning DNA[7–10]- or RNA[10,11]-editing activity (Extended Data Fig. 1a). Specifically, the variants included deletions and mutations at the Leu-enriched amino- or carboxy-terminals of rAPOBEC1 (p.X32del, p.R33A, p.K34A, p.X34del, p.X77del, p.X116del, p.X169del, p.X182del, p.P190A and p.P191A) and mutations of putative catalytic active site residues of rAPOBEC1 (p.H61A, p.H61R, p.V62A, p.E63A, p.E63Q, p.C93S, p.C96S). On the basis of the structure of human APOBEC3G[8,9], Arg126 of rAPOBEC1 is predicted to interact with the phosphate backbone of ssDNA (Fig. 1a and Extended Data Fig. 1b), and a variant with an R126E mutation was shown to retain on-target editing activity[7]. Arg128 and Arg132 are near to Arg126. Arg132 has been reported to affect the accessibility of ssDNA[7], so we also introduced p.R128E and p.R132E mutations (Fig. 1a and Extended Data Fig. 1a,b). Further, we examined the effect of the combination of point mutations in the domain responsible for the hydrophobicity of the active site on rAPOBEC1 (p.W90A, p.W90F and p.W90Y), which was reported to narrow the width of the base-editing window[7,8].

We transfected HEK293T cells with plasmids encoding various CBE variants, and firstly evaluated the on-target activity of these 23 variants on 10 genomic loci. Screen results identified five variants which retained their on-target efficiency (Fig. 1b), among which four variants (p.W90F+p.R126E, p.W90Y+p.R126E, p.R126E and p.R132E) had no increase in insertion/deletion polymorphism (indel) rates (Fig. 1b, Supplementary Fig. 1 and Supplementary Table 1). We additionally examined 11 genomic loci to test the on-target activity of these 4 variants and found similar results (Extended Data Fig. 1c). Moreover, these experiments revealed that these three variants (p.W90F+p.R126E, p.W90Y+p.R126E and p.R132E) exhibited narrowed base-editing windows and lower level of indels (Extended Data Fig. 1d–f). These findings are consistent with previous work[7].

We next performed GOTI using CRISPR-Cas9, cytosine base editor 3 (BE3, rAPOBEC1-nCas9-UGI) to evaluate the DNA off-target edits of these variants with high DNA on-target efficiency: $BE3^{R126E}$, $BE3^{R132E}$, $BE3^{W90Y+R126E}$ (hereafter YE1-BE3) and $BE3^{W90F+R126E}$ (hereafter FE1-BE3) (Supplementary Table 2). We note that embryonic development was not deleteriously affected by any of these variants (Extended Data Fig. 2a). The on-target efficiency of these variants was assessed via whole-genome sequencing (WGS) (Extended Data Fig. 3a). Compared with wild-type BE3-treated embryos, the number of DNA off-target single-nucleotide variants (SNVs) in the embryos treated with $BE3^{R126E}$, $BE3^{R132E}$, YE1-BE3 or FE1-BE3 was significantly reduced: from 283 ± 32 in the BE3-treated embryos to 28 ± 6 in $BE3^{R126E}$, 47 ± 8 in $BE3^{R132E}$, 12 ± 2 in YE1-BE3 and 27 ± 19 in FE1-BE3 (Fig. 2a, Extended Data Fig. 4a–e and Supplementary Table 3).

Importantly, there was no significant difference in the number of SNVs between the embryos injected with the three CBE variants

[1]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 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]CAS Key Laboratory of Systems Biology, CAS Center for Excellence in Molecular Cell Science, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China. [4]Bio-Med Big Data Center, Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [5]Center for Biomedical Informatics, Shanghai Children's Hospital, Shanghai Jiao Tong University, Shanghai, China. [6]Department of Life Sciences, Shanghai Tech University, 100 Haike Road, Shanghai, China. [7]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [8]Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China. [9]Shanghai Center for Bioinformation Technology, Shanghai Academy of Science & Technology, Shanghai, China. [10]These authors contributed equally to this work: Erwei Zuo, Yidi Sun, Tanglong Yuan, Bingbing He, Changyang Zhou. ✉e-mail: erweizuo@163.com; yxli@sibs.ac.cn; huiyang@ion.ac.cn

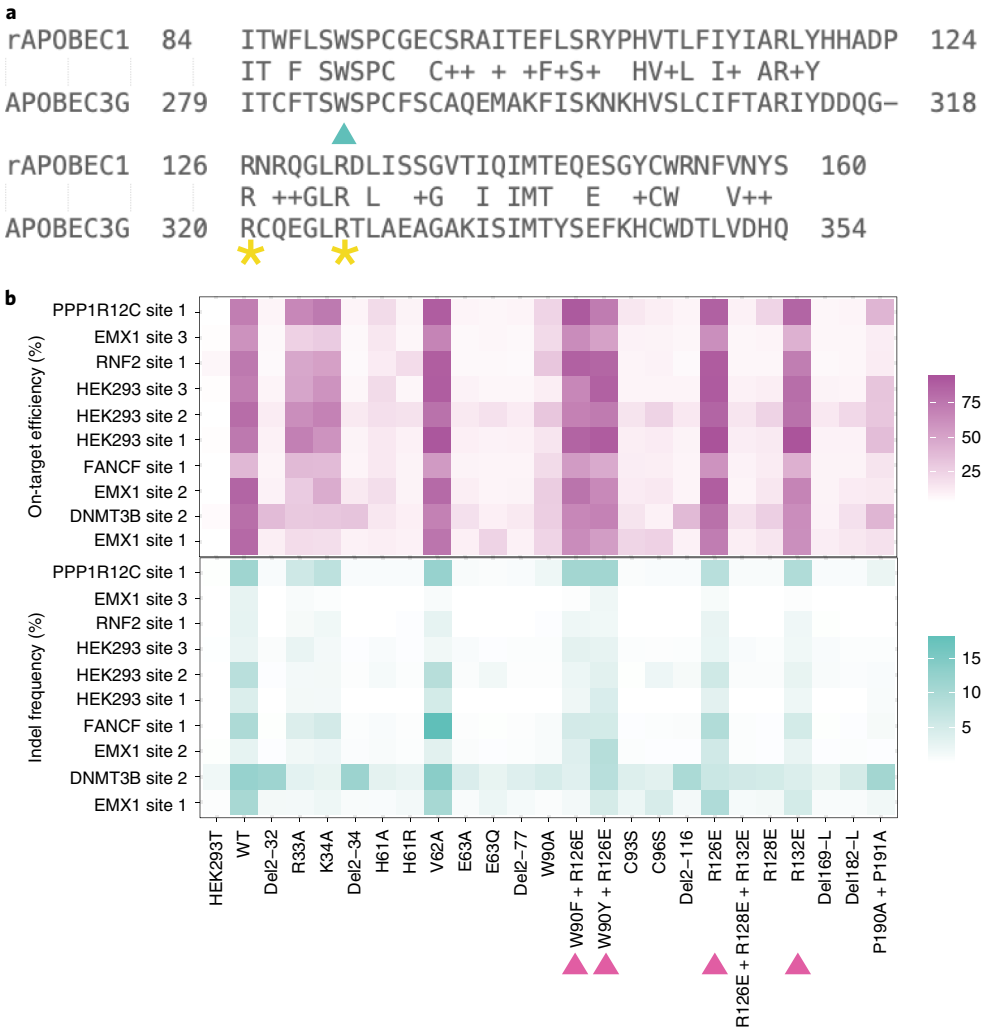


Fig. 1 | On-target efficiency of engineered CBEs. a, A sequence alignment between human APOBEC3G and rAPOBEC1. +, common substitutions. The turquoise triangle indicates residues in the hydrophobic active domain of APOBEC3G, and yellow stars indicate residues in the ssDNA-binding domain. **b**, The on-target efficiency and indel frequencies of engineered CBE variants. $n = 3$ biological replicates for each group. Purple triangles indicate the variants which were subsequently assessed for their off-target effects.

(BE3^{R126E}, YE1-BE3 and FE1-BE3) and the non-edited control embryos (ten SNVs on average, which is close to the spontaneous mutation rate[2]). Moreover, these CBE variants exhibited significantly reduced mutation bias compared with wild-type BE3, and none of the detected SNVs overlapped with the off-target sites predicted by Cas-OFFinder and CRISPOR (Fig. 2b and Extended Data Figs. 3b and 5). These findings suggest that the CBE variants induce fewer Cas9-independent DNA off-target SNVs.

We also assess the off-target activity on the transcriptomes of transfected HEK293T cells. Compared with wild-type BE3, three variants showed significantly reduced RNA off-target edits at 36 h post-transfection: BE3^{R132E}, YE1-BE3 and FE1-BE3 (Fig. 2c,d and Extended Data Fig. 6). In contrast, the BE3^{R126E} variant showed no decrease of RNA off-target effects 36 h post-transfection (Fig. 2c), but exhibited a significant decrease at the 72 h post-transfection time point (Extended Data Fig. 7a). In comparison with green fluorescent protein (GFP)-transfected control cells, the wild-type BE3 and the BE3^{R126E} variant cells had significantly higher numbers of SNVs, whereas no difference was detected for the BE3-R132E, YE1-BE3 or FE1-BE3 variants with GFP-transfected control cells (Fig. 2c,d and Extended Data Figs. 6 and 7a–c). Together, these results support the notion that the BE3^{R132E}, YE1-BE3 and FE1-BE3 variants are high-fidelity base editors that offer remarkable reductions in the extent of both DNA and RNA off-target effects compared with BE3.

Although these three BE3 variants could significantly reduce the off-target effects, their on-target efficiencies are not as high as for the previously reported BE3-human APOBEC3A (hA3A)[12] (Fig. 3a). We attempted to analyze the off-target effects of BE3-hA3A using GOTI; however, we found that BE3-hA3A was toxic to the embryos (Extended Data Fig. 2a,b), which was possibly caused by the accumulation of Cas9-independent off-target SNVs. We thus introduced BE3-hA3A with a p.Y130F mutation, conferring high DNA on-target efficiency[13] and high fidelity[6]. This BE3-hA3A^{Y130F} editor had high on-target editing efficiency (Fig. 3a and Extended Data Fig. 8a), but still generated a substantial number of DNA off-target SNVs (409 ± 86) (Fig. 3b, Extended Data Fig. 8b–d and Supplementary Tables 2 and 3).

We next generated a variant of the high-fidelity variant YE1-BE3 that had two additional attributes on the basis of the previously reported BE3-flag-tagged nuclear localizaiton signal (FNLS) editor: a DNA sequence codon optimized for expression in human cells, and an N-terminal nuclear localization signal peptide[14]. We found this new YE1-BE3-FNLS variant had comparable on-target efficiency

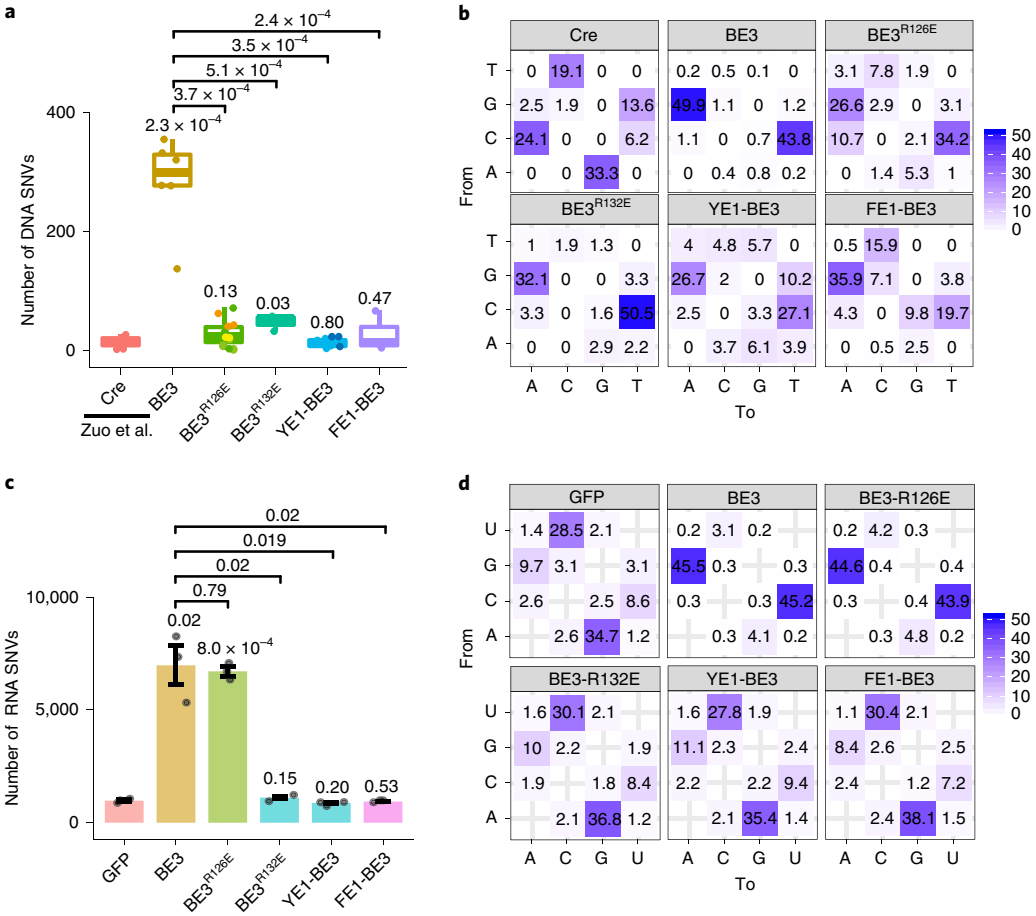


Fig. 2 | DNA and RNA off-target evaluation of engineered CBEs. a, Comparison of the total number of detected DNA off-target SNVs. biologically independent samples for Cre ($n=3$), BE3 ($n=6$), BE3^{R126E} ($n=12$), BE3^{R132E} ($n=3$), YE1-BE3 ($n=8$) and FE1-BE3 ($n=3$). Two Cre samples and six BE3 samples were derived from Zuo et al.[2] and one Cre sample was newly generated in this study. The center line indicates the median, and the bottom and top lines of the box represent the first quartile and third quartile of the values, respectively. Tails extend to the minimum and maximum values. **b**, Distribution of mutation types for the groups transfected with Cre, BE3 or one of the four CBE variants. **c**, Comparison of the total number of detected RNA off-target SNVs at 36 h post-transfection. $n=3$ biologically independent samples for each group. All values are presented as mean ± s.e.m. **d**, Distribution of mutation types for groups transfected with Cre, BE3 or one of the four CBE variants. All *P* values were calculated by two-sided Student's *t*-tests.

(at the site C5–C7, 78.3 ± 2.8%) with BE3-FNLS (80.8 ± 2.1%), BE3-hA3A (78.7 ± 2.4%) and BE3-hA3A^{Y130F} (79.9 ± 2.4%), but much higher than BE3 (59.2 ± 4.1%) or YE1-BE3 (47.6 ± 3.9%) (Fig. 3a,d–g and Supplementary Table 4 includes the original sequence information from which we calculated the on-target editing efficiency). Moreover, YE1-BE3-FNLS exhibited the lower level of bystander edits (out of C5-C7, 7.3 ± 1.5%) (Fig. 3e) and indels (3.5 ± 1.3%) (Fig. 3f) than BE3-FNLS (19.7 ± 2.8% for bystander edits and 7.2 ± 1.6% for indels). We also compared YE1-BE3-FNLS with BE4max[15] and found that YE1-BE3-FNLS displayed a comparable or higher on-target editing efficiency, but lower level of bystander edits and indels (Extended Data Fig. 9a,b). Besides, YE1-BE3-FNLS also had significantly reduced DNA and RNA off-target effects compared with BE3, reaching levels indistinguishable from unedited controls (Fig. 3b,c, Extended Data Fig. 8b–g and Supplementary Tables 2 and 3). We further examined Cas9-dependent off-target effects of YE1-BE3-FNLS by a previously reported method[16] and found no significant difference in Cas9-dependent off-target effects in cells transfected with YE1-BE3-FNLS compared with cells treated with other BE3 variants (Extended Data Fig. 10). This new CBE variant thus meets dual criteria for an advanced base editor: high on-target efficiency and high fidelity.

We observed that the variants with reduced DNA and RNA off-target effects (BE3^{R132E}, YE1-BE3, FE1-BE3 and YE1-BE3-FNLS) all had narrowed base-editing windows. To explain these results, Rees et al.[1] previously reported that bases located outside the activity window yet within the ssDNA R-loop region may still be edited—albeit at a lower efficiency—particularly if they are located in a favorable editing motif for rAPOBEC1. It is notable that our best-performing variant YE1-BE3-FNLS simultaneously exhibited the highest on-target efficiency and the lowest levels of indels and of bystander edits. A previous study showed that multiple cytosines in the editing window may increase the probability of indels occurring during editing[17]. Compared with BE3-FNLS, YE1-BE3-FNLS substantially narrows the base-editing window, which inherently decreases the distribution of multiple cytosines, potentially explaining our observation of significantly reduced indel frequency.

Very recently, Liu and colleagues[18] also found that YE1 variants display background levels of Cas9-independent DNA and RNA off-target editing, a narrowed editing window and a lower indel frequency. Notably, in our study we found that some of the CBE variants retained high off-target DNA SNV effects while dramatically decreasing off-target RNA effects (and vice versa). It is possible that rAPOBEC1 or hAPOBEC3A may adopt distinct binding modes to accommodate ssDNA versus RNA. This highlights the need for base editor researchers to evaluate the off-target effects of base editors on both DNA and RNA. Specifically, we suspect that p.R132E affects rAPOBEC1's interactions with both DNA and RNA, whereas

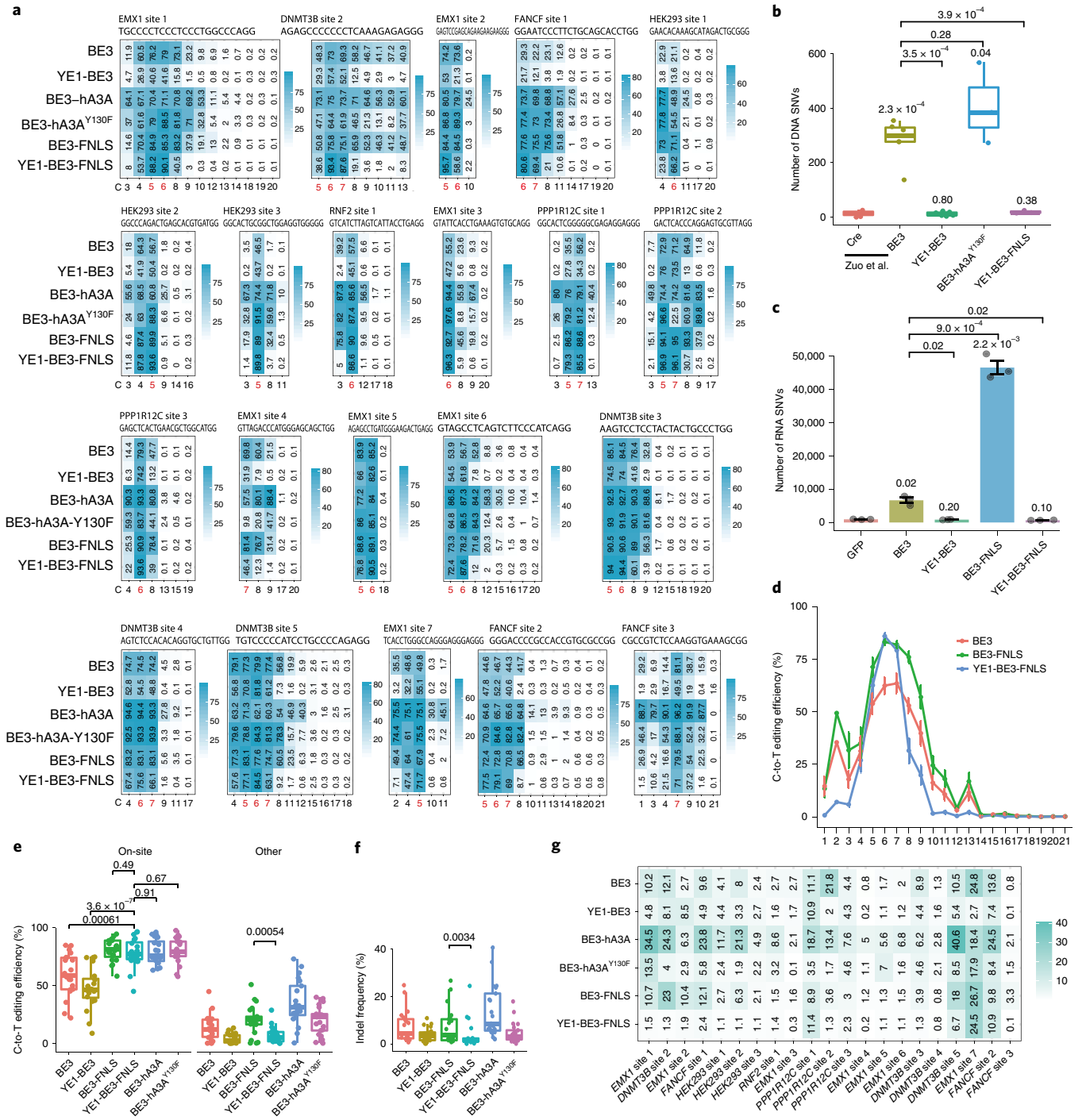


Fig. 3 | Activities of engineered BE3-FNLS and BE3-hA3A. a, The on-target C-to-T editing efficiency of engineered BE3-FNLS and BE3-hA3A at 21 target sites. Sequences of the on-target protospacers and primers are shown in Supplementary Table 4. The red numbers indicate on-site positions, which are those from 5 to 7. **b**, Comparison of the total number of detected DNA off-target SNVs. Biologically independent samples for Cre ($n=3$), BE3 ($n=6$), YE1-BE3 ($n=5$), BE3-hA3A^{Y130F} ($n=3$) and YE1-BE3-FNLS ($n=3$). **c**, The comparison of the total number of detected RNA off-target SNVs at 36 h post-transfection. $n=3$ biologically independent samples for each group. **d**, The C-to-T editing efficiency of engineered variants on BE3-FNLS at each of the 21 target sites. **e**, Comparison of on-site C-to-T editing efficiency and bystanders for the engineered variants on 21 target sites. On-site, positions 5–7. Other, positions outside 5–7. The center line indicates the median, and the bottom and top lines of the box represent the first quartile and third quartile of the values, respectively. Tails extend to the minimum and maximum values. **f**, The indel frequency of the engineered variants for the 21 target sites. $n=21$ independent experiments for each group in **d**, **e** and **f**. **g**, Indel distributions for the engineered variants. All *P* values were calculated by two-sided Student's *t*-tests. All values are presented as mean ± s.e.m in **c** and **d**.

p.R126E mainly affects its DNA-binding and p.Y130F mainly affects its RNA-binding ability. The YE1-BE3-FNLS variant harbors both the p.R126E mutation and a tyrosine substitution at Trp90—positioned at the hydrophobic region of rAPOBEC1 that has been proposed to participate in rAPOBEC1's binding with ssDNA and RNA. We tentatively speculate that the functional contribution of p.W90Y

to the observed increase in YE1-BE3-FNLS's fidelity likely results from alteration of rAPOBEC1–RNA interactions. Our work illustrates examples of how the off-target effects of base editors can be minimized via biological-insight-driven engineering to extend the utility of these powerful gene-editing tools for both research and therapeutic applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-020-0832-x.

Received: 18 January 2020; Accepted: 13 April 2020;
Published online: 18 May 2020

References

1. 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).
2. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
3. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
4. Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
5. Grünewald, J. et al. CRISPR adenine and cytosine base editors with reduced RNA off-target activities. **37**, 1041–1048 (2019).
6. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **37**, 1041–1048 (2019).
7. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
8. Holden, L. G. et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121–124 (2008).
9. Chen, K. M. et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. *Nature* **452**, 116–119 (2008).
10. Teng, B. B. et al. Mutational analysis of apolipoprotein B mRNA editing enzyme (APOBEC1). structure-function relationships of RNA editing and dimerization. *J. Lipid Res.* **40**, 623–635 (1999).
11. Teng, B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816–1819 (1993).
12. Gehrke, J. M. et al. An APOBEC3A–Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol.* **36**, 977–982 (2018).
13. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol.* **36**, 946–949 (2018).
14. Zafra, M. P. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol.* **36**, 888–893 (2018).
15. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
16. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
17. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
18. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* (in the press).

Methods

Animal care. Heterozygous Ai9 (full name B6.Cg-Gt (ROSA) 26Sortm9 (CAG-td-Tomato) Hze/J; JAX strain 007909) male mice and female C57BL/6 mice (4 weeks old) were mated for embryo collection. ICR female mice were used for recipients. The animals usage and care complied with the guideline of the Biomedical Research Ethics Committee of Shanghai Institutes for Biological Science, Chinese Academy of Sciences.

Generation of mutant base editor messenger RNA and single guide RNA. T7 promoter was added to base-editor coding region by PCR amplification of plasmid, using primer base editor F and R. T7-base editor PCR product was purified and used as the template for in vitro transcription (IVT) using mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). T7 promoter was added to sgRNA template by PCR amplification of px330. The T7-sgRNA PCR product was purified and used as the template for IVT using MEGA shortscript T7 kit (Life Technologies). T7 promoter was added to Cre template by PCR amplification. T7-Cre PCR product was purified and used as the template for IVT using mMESSAGE mMACHINE T7 ULTRA kit (Life Technologies). Cas9 mRNA, Cre mRNA and sgRNAs were purified using MEGA clear kit (Life Technologies) and eluted in RNase-free water.

sgRNA sequences were as follows:
Locus sequence (5′–3′)
Tyr-C[2] GACCTCAGTTCCCCTTCAAAGGG
Tyr-D[2] CTGTGCCAAGGCAGAAACCCTGG
Tyr-F TGCGGCCAGCTTTCAGGCAGAGG
Primers were as follows:
Name sequence (5′–3′)
Base editor IVT: FTCCGCGGCCGCTAATACGACT
Base editor IVT: RTGGTTCTTTCCGCCTCAGAAGCC
Cre IVT:
FTAATACGACTCACTATAGGGAGACAGATCACCTTTCCTATCAACC
Cre IVT: RTCGGTATTTCCAGCACACTGGA
Tyr-C IVT: FTAATACGACTCACTATAGGGGACCTCAGTTCCCCTTCA-AAGTTTTAGAGCTAGAAATAG
Tyr-D IVT: FTAATACGACTCACTATAGGGCTGTGCCAAGGCAGAAAC-CCGTTTTAGAGCTAGAAATAG
Tyr-F IVT: FTAATACGACTCACTATAGGGTGCGGCCAGCTTTCAGGCAG GTTTTAGAGCTAGAAATAG
sgRNA IVT: RAAAAGCACCGACTCGGTGCC

Blastocyst rate evaluation. Superovulated female C57BL/6 mice (4 weeks old) were mated to C57BL/6 males, and fertilized embryos were collected from oviducts 24 h after human chorionic gonadotropin (hCG) injection. Cas9 mRNA and sgRNA were mixed and injected into the cytoplasm of fertilized eggs in M2 medium containing 5 μg mL^{-1} cytochalasin B (CB) using a FemtoJet microinjector (Eppendorf) with constant flow settings. The injected embryos were cultured in KSOM at 37 °C under 5% CO_2 in air until blastocyst stage.

Two-cell embryo injection, embryo culturing and embryo transplantation. Super ovulated C57BL/6 females (4 weeks old) were mated to heterozygous Ai9 (full name B6.Cg-Gt(ROSA)26Sortm9(CAG-td-Tomato)Hze/J; JAX strain 007909) males, and fertilized embryos were collected from oviducts 23 h after hCG injection. For two-cell editing, the mixture of BE3 mRNA (50 ng $μL^{-1}$) or BE3 variants mRNA (50 ng $μL^{-1}$), sgRNA (50 ng $μL^{-1}$) and Cre mRNA (2 ng $μL^{-1}$) was injected into the cytoplasm of one blastomere of the two-cell embryo 48 h after hCG injection in a droplet of M2 medium containing 5 μg mL^{-1} cytochalasin B (CB) using a FemtoJet microinjector (Eppendorf) with constant flow settings. The injected embryos were cultured in KSOM medium with amino acids at 37 °C under 5% CO_2 in air for 2 h and then transferred into oviducts of pseudopregnant ICR females at 0.5 d post coitum.

Cloning. Site-directed mutagenesis of BE3 (Addgene plasmid no. 73021) was done using NEBuilder HiFi DNA Assembly Master Mix (New England BioLabs). Plasmids encoding BE3-FNLS and BE4max were obtained from Addgene (plasmid no. 110841, no. 112093). Briefly, a primer with an overhang containing the desired point mutation was used to amplify the appropriate vector plasmid by PCR. pCMV-BE3 variants-polyA-pCMV-mCherry-polyA was generated through NEBuilder HiFi DNA Assembly, by combining a PCR-amplified pCMV-mCherry-poly A with a digested pCMV-BE3 variants backbone (Supplementary Sequence 1). pCMV-EGFP-polyA-U6-sgRNA were generated through NEBuilder HiFi DNA Assembly, by combining a PCR-amplified U6-sgRNA with a digested pCMV-EGFP-poly A backbone.

Cell culture, transfections and FACS. HEK293T cells were maintained in DMEM supplemented with 10% FBS in a 37 °C humidified incubator with 5% CO_2. pCMV-BE3 (WT/BE3 variants)-polyA-pCMV-mCherry-polyA and pCMV-EGFP-polyA-U6-sgRNA expression plasmids were co-transfected using Lipofectamine 3000 (ThermoFisher Scientific) according to the manufacturer's protocol. At 36 or 72 h post-transfection, cells were washed with PBS and trypsinized using 0.05% trypsin–EDTA. The cell suspension was filtered through a 40-μm cell strainer, and EGFP/mCherry positive cells were isolated by fluorescence-activated cell sorting (FACS).

RNA sequencing. Around 500,000 cells (top 5% EGFP/mCheery signal) were collected, and RNA was extracted according to the standard protocol. For library construction, mRNAs were fragmented and converted to cDNA using random hexamers or oligo(dT) primers. The 5′ and 3′ ends of cDNA were ligated with adapters, and correctly ligated cDNA fragments were enriched and amplified by PCR. The concentration of the library was assessed using Bioanalyzer. Sequencing was performed on the Illumina HiSeq platforms.

Next-generation sequencing of DNA amplicons. Transfected cells were collected after 72 h, and $EGFP^+mCherry^+$ cells were sorted by FACS. The genomic DNA was isolated using the TIANamp Genomic DNA Kit (TIANGEN) according to the manufacturer's instructions. Genomic sites of interest were amplified by PCR using gene-specific primers (Supplementary Tables 5 and 6) flanking the target sequence. ExTaq (TAKARA) was activated at 95 °C for 3 min, and PCR was performed for 34 cycles at 95 °C for 30 s, 62 °C for 30 s and 72 °C for 1 min, with a final extension at 72 °C for 5 min. DNA amplicons were purified using universal DNA purification kit (TIANGEN) according to the manufacturer's instructions. The amplicons were ligated to adapters and sequencing was performed on the Illumina HiSeq Xten platforms.

Targeted amplicon sequencing data analysis. Sequencing data was firstly demultiplexed by Cutadapt (v2.8) on the basis of sample barcode. The demultiplexed reads were then processed by CRISPResso2 (ref. [19]) for the quantification of mutations, insertions and deletions at each target site. Alternatively, we also aligned the demultiplexed reads to the target genome by NCBI BLASTN[20] (v2.2.29+), BWA[21] (v0.7.12) and Bowtie2 (ref. [22]) (v2.2.9), respectively, quantified the mutated alleles by each method and compared the results with those from CRISPResso2. Note that we found that the other three tools[13,20–24] have difficulty aligning consecutive mutated sequences to the target sites, thus resulting in the loss of many consecutively mutated reads when the on-target site has continuous Cs (Supplementary Fig. 2). Here we recommended CRISPResso2, recently developed by Clement et al.[19] solved this problem by extending the Needleman–Wunsch algorithm to increase the detection sensitivity of these fragments.

FACS for GOTI. To isolate mouse embryonic cells, the prepared tissues were dissociated enzymatically in an incubation solution of 5 mL trypsin–EDTA (0.05%) at 37 °C for 30 min. The digestion was stopped by adding 5 mL of DMEM with 10% FBS. Fetal tissues were then homogenized by being passed 30-40 times through 1-mL pipette tips. The cell suspension was centrifuged for 6 min (800 r.p.m.), and the pellet was resuspended in DMEM with 10% FBS. Finally, the cell suspension was filtered through a 40-μm cell strainer, and $tdtomato^+tdtomato^-$ cells were isolated by FACS. Samples were found to be >95% pure when assessed with a second round of flow cytometry and fluorescence microscopy analysis.

WGS and RNA-seq data analysis. DNeasy blood and tissue kit (catalog number 69504, Qiagen) was used to extract genomic DNA from cells following the manufacturer's instructions. WGS was performed at mean coverages of 50× by Illumina HiSeq X Ten. BWA (v0.7.12) was used to map qualified sequencing reads to the reference genome (mm10). The mapped BAM files were then sorted and marked using Picard tools (v2.3.0). To identify the genome-wide de novo SNVs with high confidence, we conducted single-nucleotide variation calling on three algorithms, Mutect2 (v3.5), Lofreq (v2.1.2) and Strelka (v2.7.1) with default parameters, separately[25–27]. In parallel, Mutect2 (v3.5), Scalpel (v0.5.3) and Strelka (v2.7.1) were run individually for the detection of whole-genome de novo indels with default parameters[25,27,28]. The overlap of three algorithms of SNVs or indels were considered as the true variants. All the sequencing data were deposited in NCBI Sequence Read Archive (SRA) under project accession PRJNA527003.

Potential off-targets of targeted sites were predicted using two previous reported algorithms, Cas-OFFinder (http://www.rgenome.net/cas-offinder/) and CRISPOR (http://crispor.tefor.net/) with all possible mismatches[29,30].

The SNVs and indels were annotated with annovar (version 2016-02-01) using RefSeq database[31].

For RNA-seq data analysis, FastQC (v0.11.3) and Trimmomatic (v0.36)[32] were used for quality control. Qualified reads were mapped to the reference genome (Ensemble GRCh38) using STAR (v2.5.2b)[33] in two-pass mode with default parameters. Picard tools (v2.3.0) was then applied to sort and mark duplicates of the mapped BAM files. The refined BAM files were subject to split reads that spanned splice junctions, local realignment, base recalibration and variant calling with SplitNCigarReads, IndelRealigner, BaseRecalibrator and HaplotypeCaller tools from GATK (v3.5)[34], respectively.

Structure prediction. Amino acid sequences of rat APOBEC1 and human APOBEC3G were retrieved from UniProt (https://www.uniprot.org/) and sequence alignment was performed with NCBI blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). The structure of rAPOBEC1 was predicted by Protein Structure Prediction

Server, (PS)² [35,36], according to the consensus sequence and secondary structure information for proteins with known structures. The crystal structure of APOBEC3G was downloaded from PDB (http://www.rcsb.org/3d-view/3IQS) and presented using PyMOL (v2.3.2).

Statistical analysis. R version 3.5.1 (http://www.R-project.org/) was used to conduct all the statistical analyses in this work. All tests conducted were two-sided, and the difference was considered significant at $P < 0.05$. In box-and-whisker plots, the center line indicates the median, the bottom and top lines of the box represents the first quartile and third quartile of the values, respectively. The bottom and top lines represent the minimum and maximum values.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All the sequencing data were deposited in NCBI Sequence Read Archive (SRA) under project accession PRJNA527003 and https://www.biosino.org/node/project/detail/OEP000272. The data that support the findings of this study are available from the corresponding author upon request.

References

19. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
21. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
22. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
23. Park, J., Lim, K., Kim, J. S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286–288 (2017).
24. Wang, X. et al. CRISPR–DAV: CRISPR NGS data analysis and visualization pipeline. *Bioinformatics* **33**, 3811–3812 (2017).
25. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
26. Wilm, A. et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
27. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
28. Fang, H. et al. Indel variant analysis of short-read sequencing data with Scalpel. *Nat. Protoc.* **11**, 2529–2548 (2016).
29. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
30. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).
31. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
33. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
34. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
35. Chen, C. C., Hwang, J. K. & Yang, J. M. (PS)2: protein structure prediction server. *Nucleic Acids Res.* **34**, W152–157 (2006).
36. Huang, T. T. et al. (PS)2: protein structure prediction server version 3.0. *Nucleic Acids Res.* **43**, W338–342 (2015).

Acknowledgements

We thank the FACS facility staff H. Wu, L. Quan and S. Qian at ION, M. Zhang at IPS, and L. Yuan at Big Data Platform (SIBS, CAS). This study was supported by the R&D Program of China (2018YFC2000100 and 2017YFC1001302, 2017YFC0909701), the CAS Strategic Priority Research Program (XDB32060000, XDBS01060100), the National Natural Science Foundation of China (31871502, 31522037, 31822035, 31922048, 31925016, 91957122), the Basic Frontier Scientific Research Program of Chinese Academy of Sciences From 0 to 1 original innovation project (ZDBS-LY-SM001), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX05), the Shanghai City Committee of science and technology project (18411953700, 18JC1410100, 16JC1420202), the National Science and Technology Major Project (2015ZX10004801-005), the National Key Research and Development Program of China (2017YFA0505500, 2016YFC0901704) the Agricultural Science and Technology Innovation Program to E.Z. and the International Partnership Program of Chinese Academy of Sciences (153D31KYSB20170059).

Author contributions

E.Z. designed and performed experiments. Y.S., W.W., R.Z. and Y.L. performed data analysis. T.Y., B.H. and J.L. performed PCR analysis. W.Y. performed mouse embryo transfer. C.Z. performed cell transfection experiments. H.Y. and Y.L. supervised the project and designed experiments. Y.S. and H.Y. wrote the paper.

Competing interests

The authors disclose a patent application relating to aspects of this work (engineered base editors).

Additional information

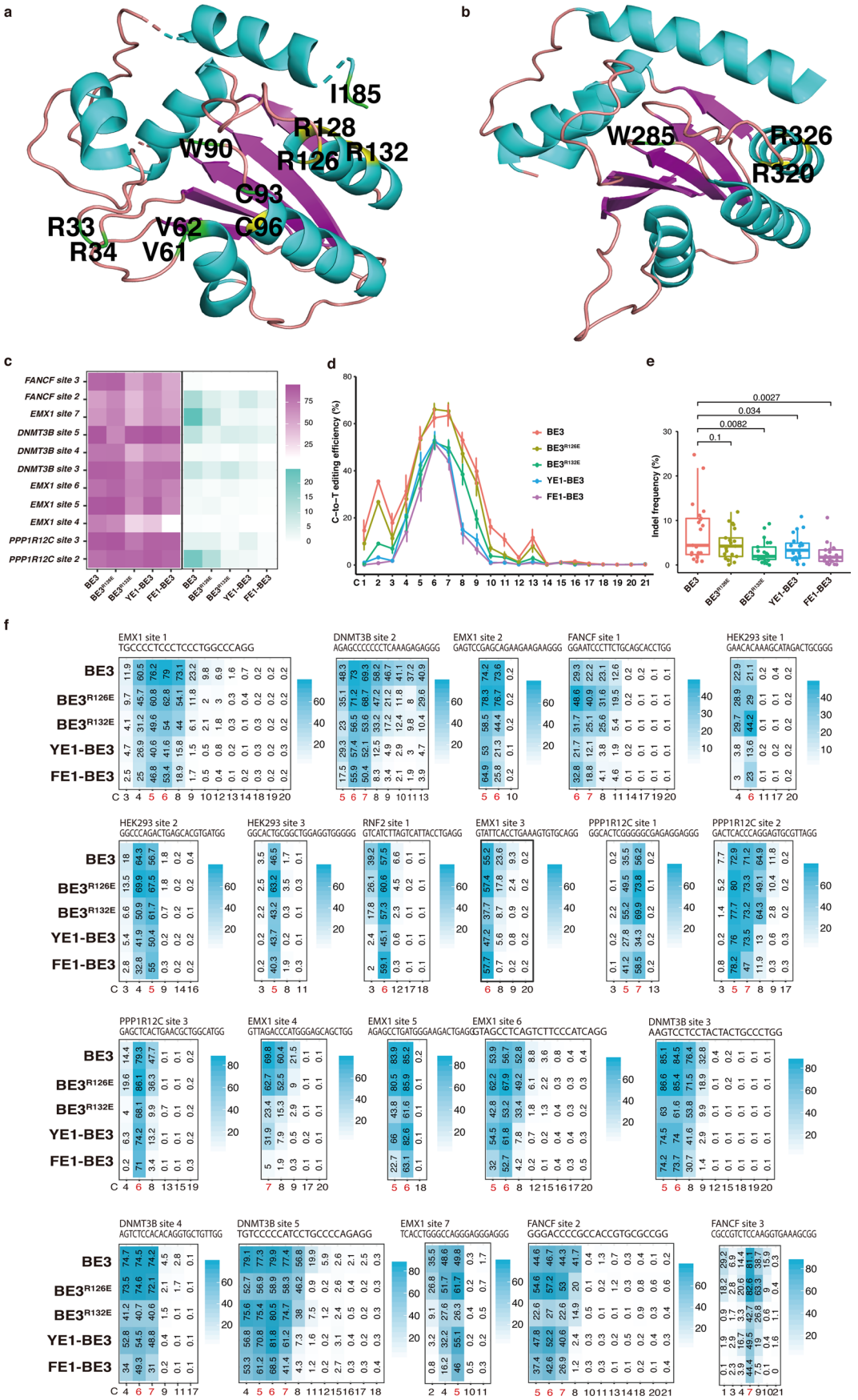
Extended data is available for this paper at https://doi.org/10.1038/s41592-020-0832-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-020-0832-x.

Correspondence and requests for materials should be addressed to E.Z., Y.L. or H.Y.

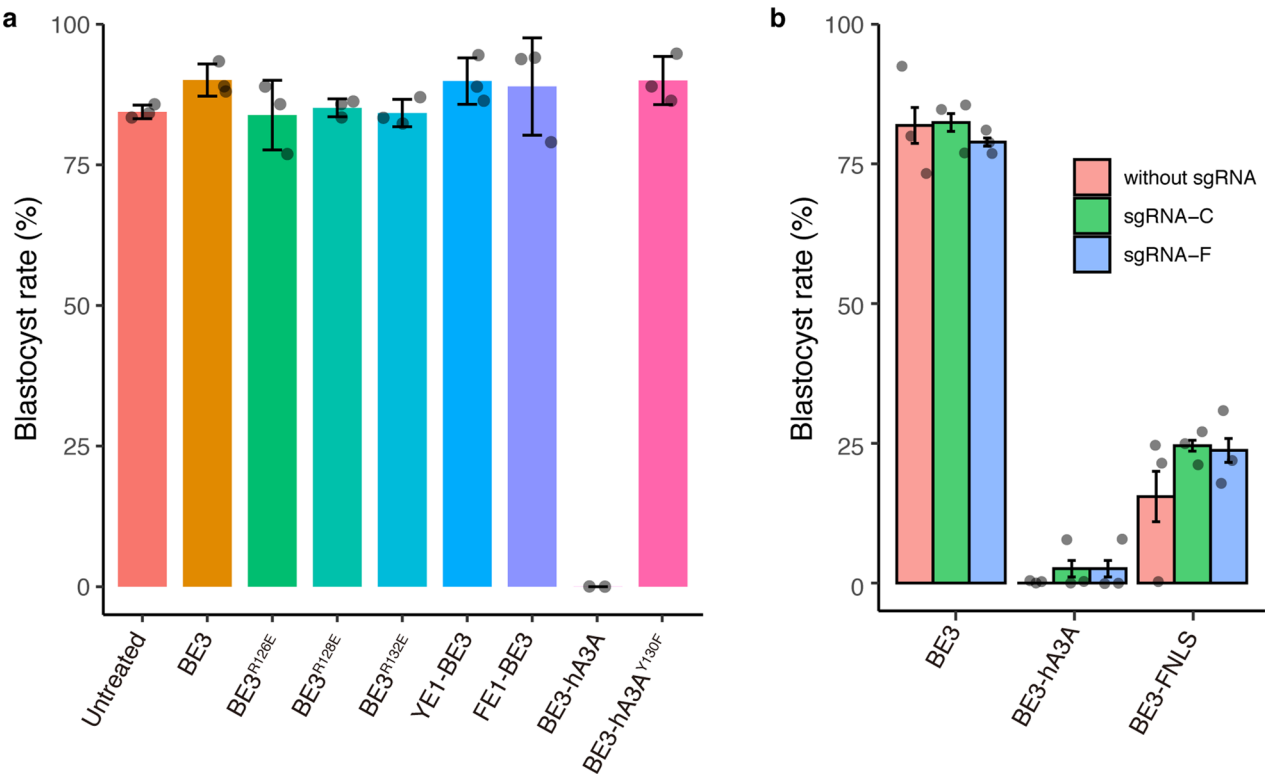
Peer review information Lei, Tang was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

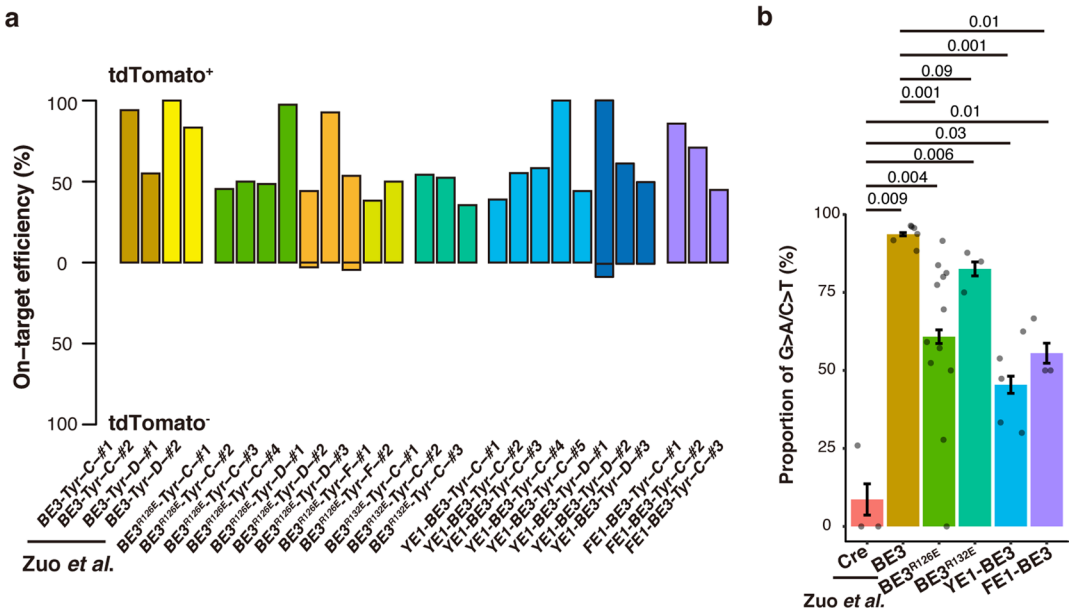


Extended Data Fig. 1 | See next page for caption.

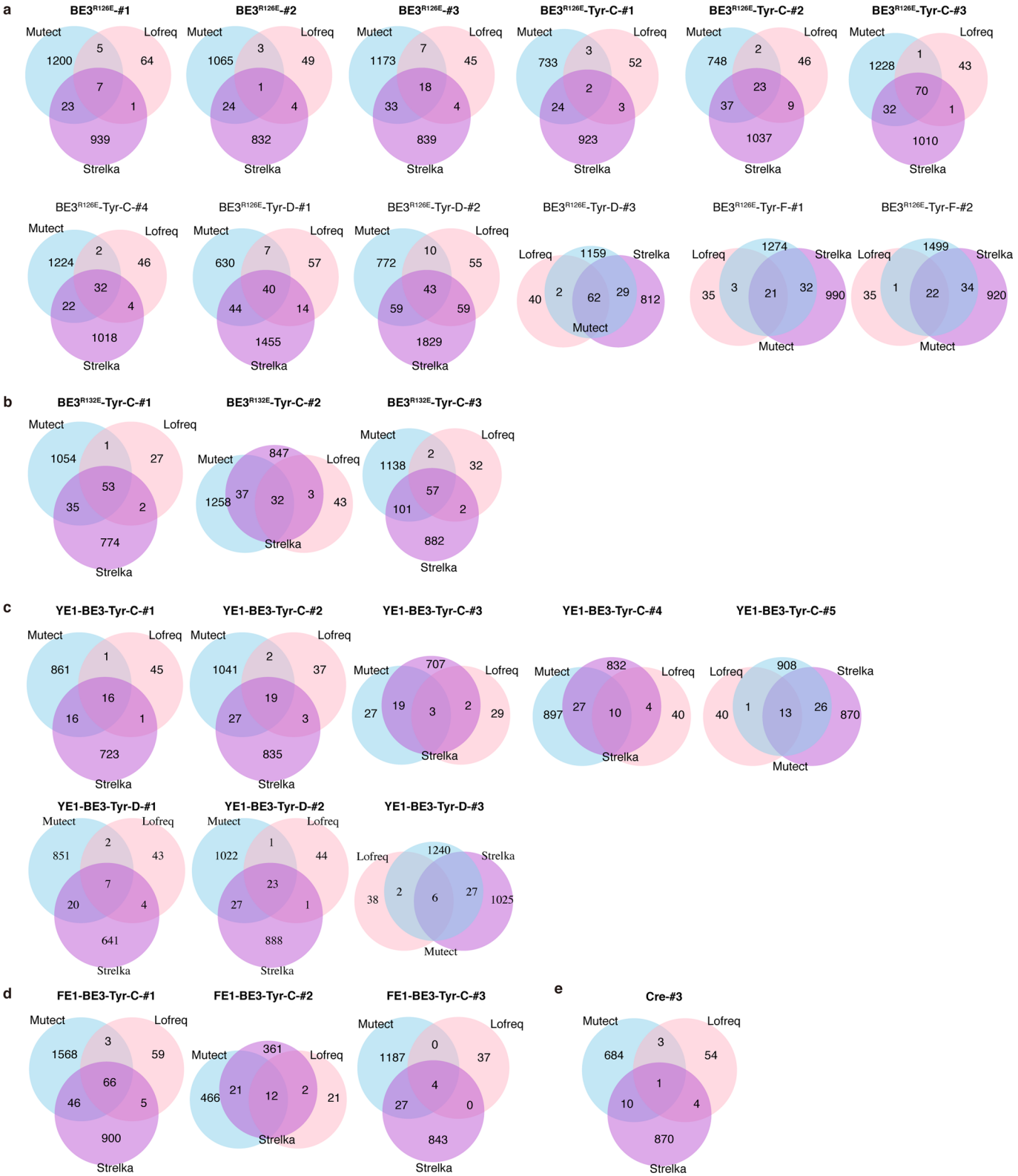
Extended Data Fig. 1 | The on-target editing ofthe BE3 and BE3 variants in different target sites. a, The mutated residues are highlighted in the predicted structure of rAPOBEC1. Green and yellow colors indicate residues in the helix and the loop of the structure, respectively. **b**, The crystal structure of APOBEC3G. **c**, The on-target efficiency and indel frequencies of different versions of engineered CBEs for additional 11 target sites. **d**, The C-to-T editing efficiency for the engineered variants at each C of the 21 target sites. n = 21 independent experiments for each group. All values are presented as mean ± s.e.m. **e**, The indel frequency comparison among the engineered variants for the 21 target sites. n = 21 independent experiments for each group. P value was calculated by two-sided Student's t-tests. Box-and-whisker plots: center line indicates median, the bottom and top lines of the box represents the first quartile and third quartile of the values, respectively. The bottom and top of the vertical line represent the minimum and maximum value. **f**, The on-target C-to-T editing efficiency of engineered BE3 variants at each target site. n = 3 biologically independent samples for each group. P value was calculated by two-sided Student's t-tests. Sequences of the on-target protospacers and primers were shown in Extended Data Table 5. The data for BE3 and YE1-BE3 are also used in Figs. 3a and 3d-g.



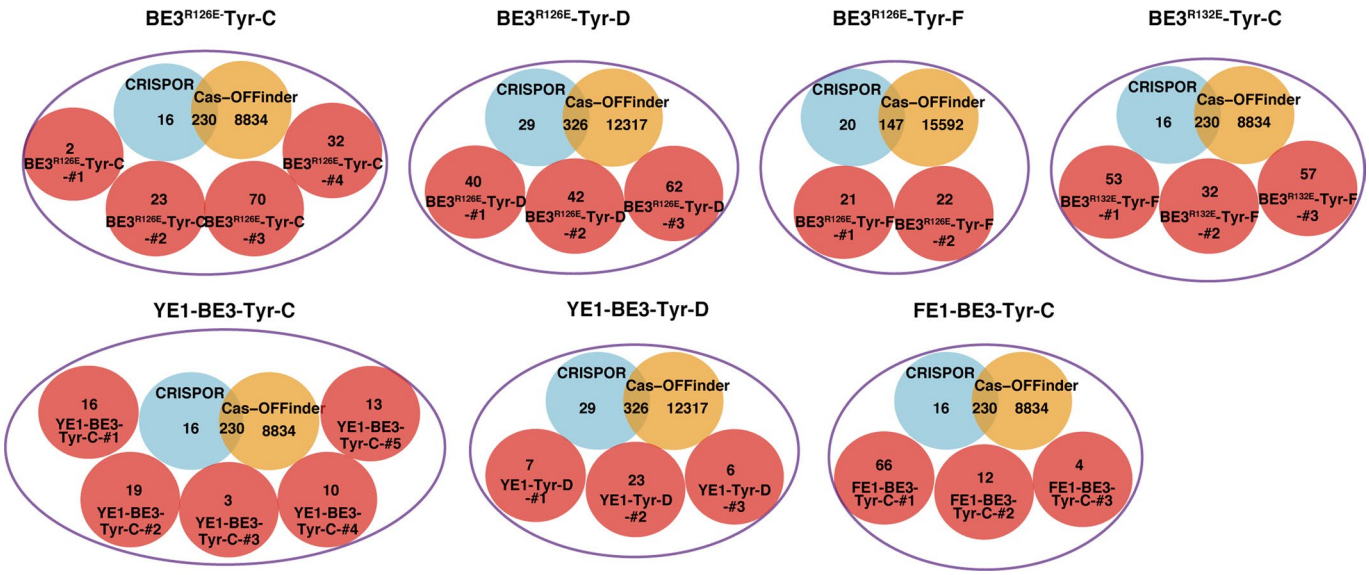
Extended Data Fig. 2 | The embryonicdevelopment rates for BE3 and BE3 variants. a, The blastocyst rate of BE3 and BE3 variants with sgRNA-D. All values are presented as mean ± s.e.m. **b**, The blastocyst rate for BE3-hA3A and BE3-FNLSwith additional sgRNAs. All values are presented as mean ± s.e.m. n = 3 biologically independent samples for each group.



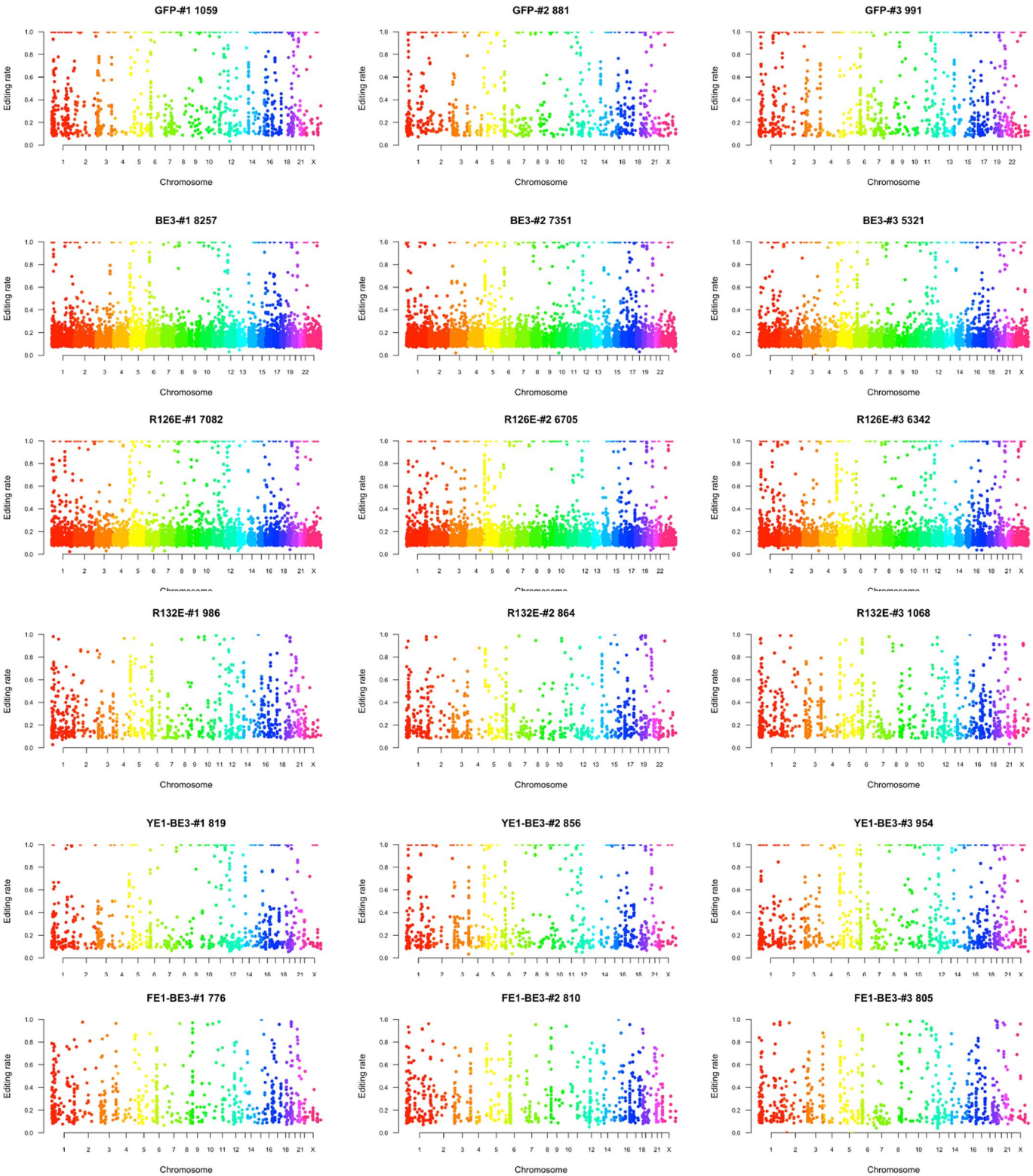
Extended Data Fig. 3 | On-target editing efficiency and characteristics of off-target SNVs of engineered CBEs. a, On-target editing efficiency of BE3 and CBE variants from WGS data. Two BE3 embryos without sgRNAs were not shown as they have no target site. **b**, Comparison of C-to-T and G-to-A conversions between CBE variants-treated and Cre or BE3 groups. n = 3 biologically independent samples for Cre, n = 6 biologically independent samples for BE3, n = 12 biologically independent samples for $BE3^{R126E}$, n = 3 biologically independent samples for $BE3^{R132E}$, n = 8 biologically independent samples for YE1-BE3, and n = 3 biologically independent samples for FE1-BE3. Two Cre samples and six BE3 samples were derived from Zuo et al.22 and one Cre sample was newly generated in this study. All values are presented as mean ± s.e.m. P value was calculated by two-sided Student's t-test.



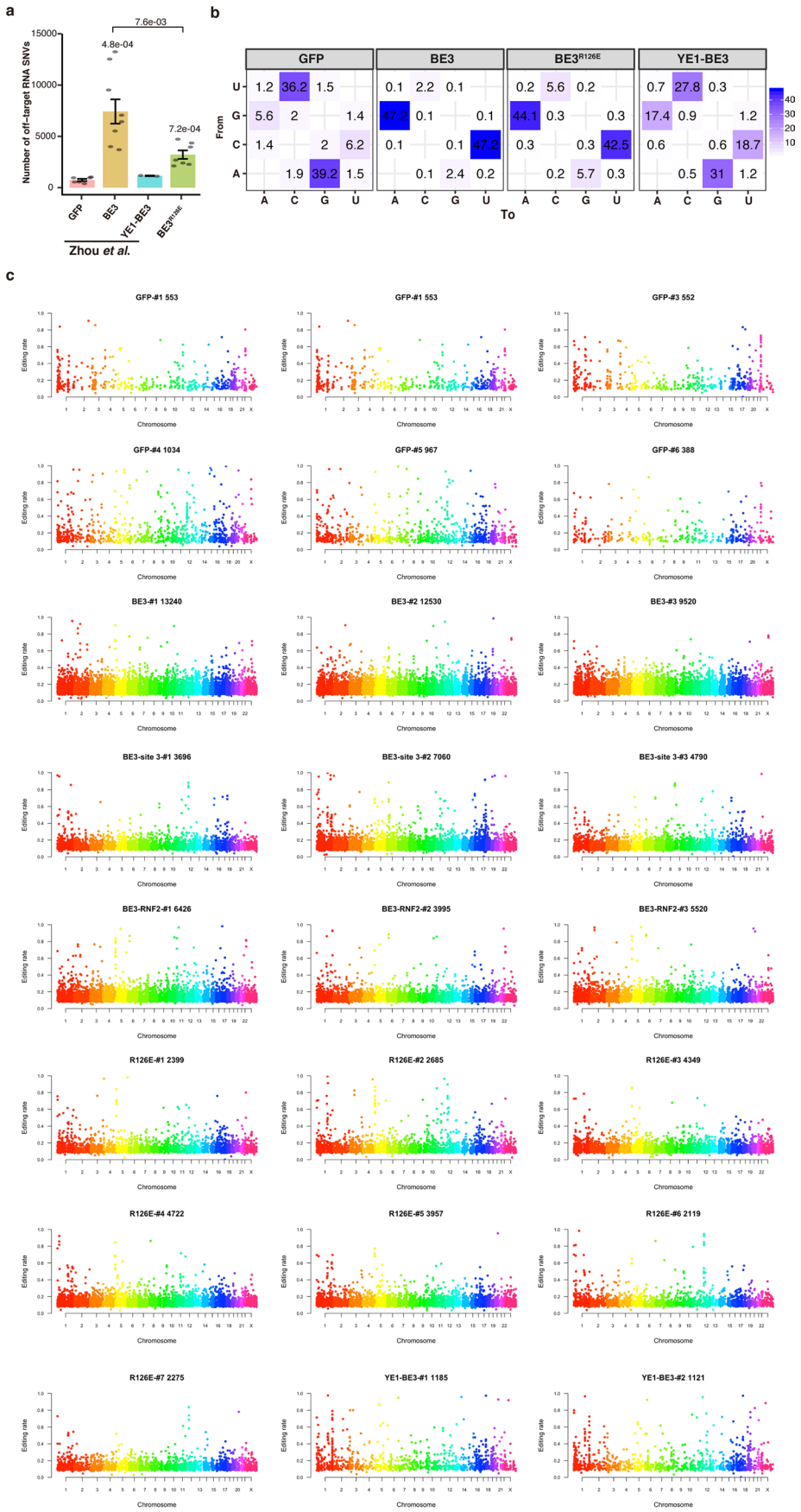
Extended Data Fig. 4 | Venn diagrams of SNVs detected in each embryo by WGS data using the indicated software tools. a, SNVs identified in BE3^{R126E}-treated embryos. **b**, SNVs identified in BE3^{R132E}-treated embryos. **c**, SNVs identified in YE1-BE3-treated embryos. **d**, SNVs identified in FE1-BE3-treated embryos. **e**, SNVs identified in the newly generated Cre-treated embryo.



Extended Data Fig. 5 | Characteristics of off-target SNVs of engineered CBEs. a, The overlap among SNVs detected from our analysis with predicted off-targets sites by Cas-OFFinder and CRISPOR.

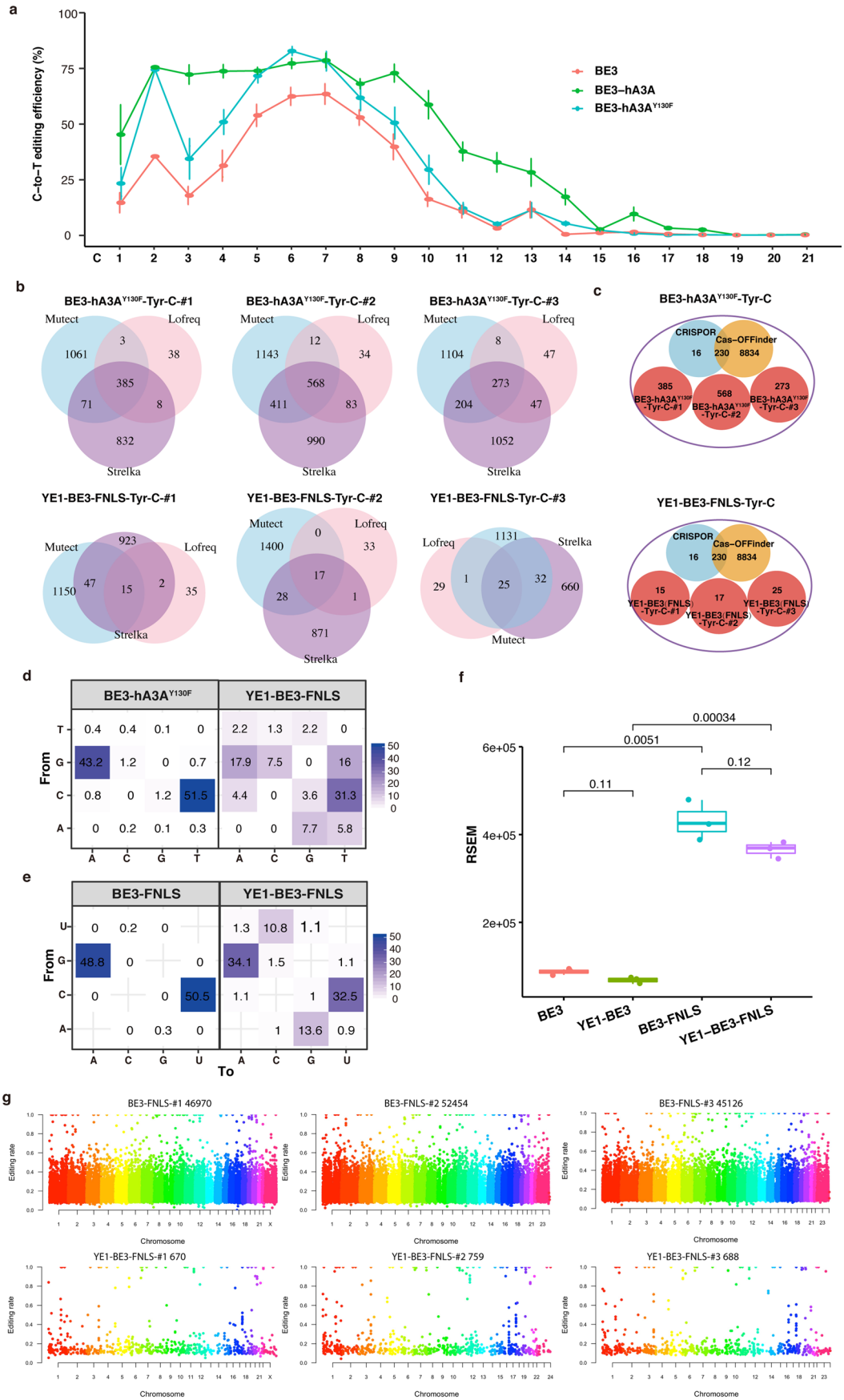


Extended Data Fig. 6 | Editing rate of RNA off-targets for BE3 variants at 36 h post-transfection. Editing rate of each variant across the chromosomes for each sample.



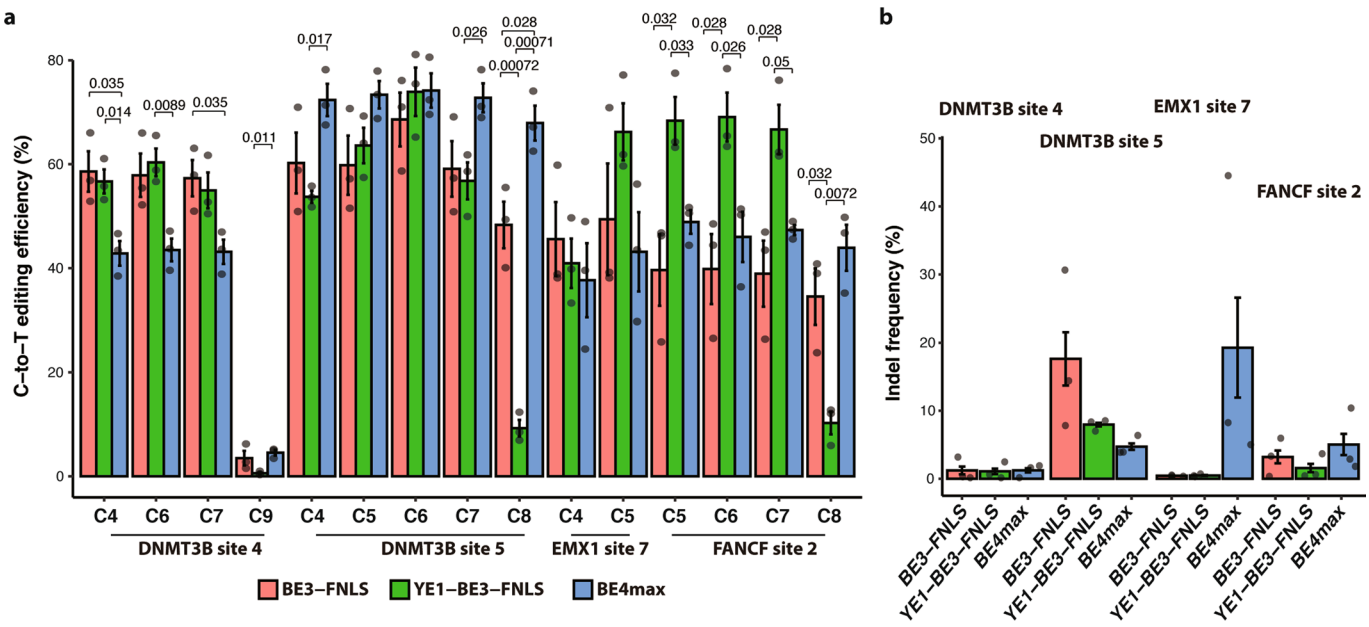
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | RNA off-target evaluation of engineered CBEs at 72h post-transfection. a, The comparison of the total number of detected RNA off-target SNVs at 72 h post-transfection. n = 6 biologically independent samples for GFP, n = 9 biologically independent samples for BE3, n = 7 biologically independent samples for BE3R126E and n = 2 biologically independent samples for YE1-BE3 groups. All values are presented as mean ± s.e.m. P values above each bar were calculated by comparing with GFP group with two-sided Student's t-tests. **b**, The distribution of mutation types for GFP, BE3, and BE3 variants-treated groups. **c**, Editing rate of RNA off-targets for BE3 variants at 72 h post-transfection.

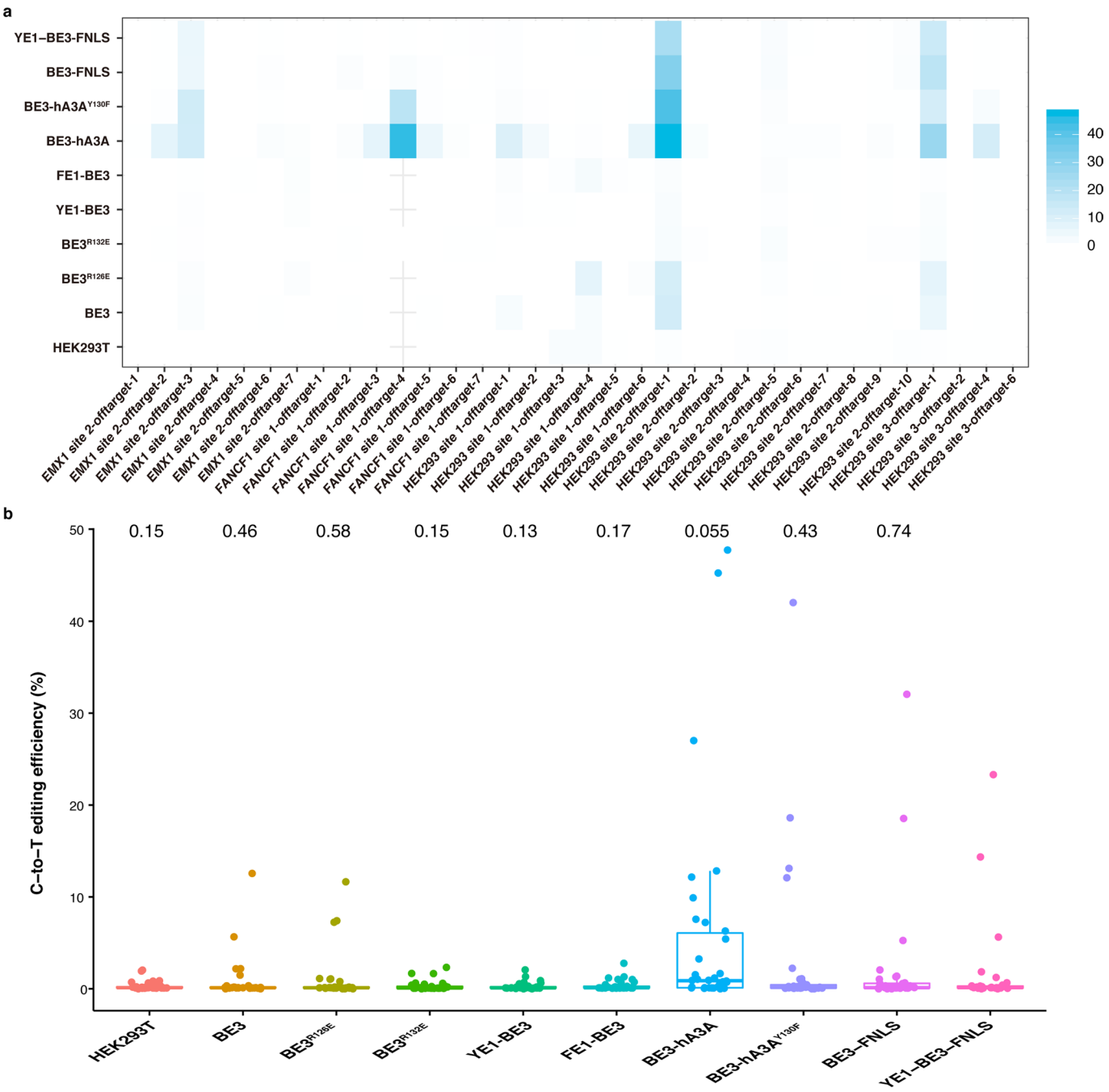


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | On-target editing efficiency and off-target effects of BE3-FNLS and BE3-hA3A. a, The C-to-T editing efficiency for the engineered variants at each C of the 21 target sites. n = 21 independent experiments for each group. All values are presented as mean ± s.e.m. The data for BE3 are also used in Fig. 3d. **b**, SNVs identified in BE3-hA3A^{Y130F} and YE1-BE3-FNLS-treated embryos. **c**, The overlap among SNVs detected from our analysis with predicted off-targets sites by Cas-OFFinder and CRISPOR. **d**, The distribution of mutation types of DNA off-target SNVs for BE3-hA3A^{Y130F} and YE1-BE3-FNLS-treated embryos. **e**, The distribution of mutation types of RNA off-target SNVs for BE3-hA3A^{Y130F} and YE1-BE3-FNLS-treated embryos. **f**, The expression level of APOBEC1 in BE3 and BE3-FNLS variants. n = 3 biologically independent samples for each group. Box-and-whisker plots: center line indicates median, the bottom and top lines of the box represents the first quartile and third quartile of the values, respectively. The bottom and top of the vertical line represent the minimum and maximum value. **g**, Editing rate of RNA off-targets for BE3 and BE3-FNLS variants at 36 h post-transfection. n = 3 biological replicates for each group. P value was calculated by two-sided Student's t-test.



Extended Data Fig. 9 | The comparation of BE3-FNLS, YE1-BE3-FNLS and BE4max. a, The C-to-T editing efficiency for BE3-FNLS, YE1-BE3-FNLS and BE4max at indicated target sites. **b**, Indel frequency for BE3-FNLS, YE1-BE3-FNLS and BE4max at indicated target sites. Data are shown as mean values ± SEM for n = 3 biological replicates performed at the same time. P value was calculated by two-sided Student's t-test.



Extended Data Fig. 10 | Activities of CBE and CBE variants at the indicated Cas9-dependent off-target sites. a, The Cas9-dependent off-target effects of the CBE and CBE variants. **b**, The comparison of editing frequencies of CBE and CBE variants at 34 potential off-target sites. P values were calculated by two sided Student's t-tests, compared with YE1-BE3-FNLS group. Each cell represents the percentage of total sequencing reads with C to T conversion. n = 21 independent experiments for each group. Box-and-whisker plots: center line indicates median, the bottom and top lines of the box represents the first quartile and third quartile of the values, respectively. The bottom and top of the vertical line represent the minimum and maximum value. HEK293T cells were transfected with plasmids expressing BE3, BE3^{R126E}, BE3^{R132E}, YE1-BE3, FE1-BE3, BE3-hA3A, BE3-hA3A^{Y130F}, BE3-FNLS and YE1-BE3-FNLS and sgRNAs matching the indicated on-target sequence using Lipofectamine 3000. Three days after transfection, genomic DNA was extracted, amplified by PCR, and analyzed by high-throughput DNA sequencing at the on-target loci, plus the top ten known Cas9 off-target loci for these sgRNAs, as previously determined using the GUIDE-seq method23, 24 and ChIP-seq method25. Sequences of the on-target and off-target protospacers and primers were shown in Extended Data Table 5.

nature research

Corresponding author(s): Hui Yang

Last updated by author(s): Mar 30, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Next-generation sequencing data was collected and demultiplexed by Illumina HiSeq X-Ten platform. FACS data was generated using MoFlo XDP (Beckman Coulter) high-speed flow cytometry sorter.
Data analysis	fastqQC (v0.11.3), Trimmomatic (v0.36), BWA (v0.7.12), Picard-tools (v2.3.0), GATK (v3.5), Lofreq (v2.1.2), Strelka (v2.7.1), Scalpel (v0.5.3), STAR (v2.5.2b), PyMOL (v2.3.2), CRISPResso2 (release 20180918), FlowJo X 10.0.7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing data were deposited in NCBI Sequence Read Archive (SRA) under project accession PRJNA527003.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. Based on previous publications, experiments in cell lines were performed in triplicates n = 3, unless otherwise noted.
Data exclusions	No data was excluded.
Replication	We tested experimental conditions using different gRNAs to ensure robustness. We also performed biological replicates performed independently of each other.
Randomization	Due to the small sample, randomization was not relevant for this study. Covariates were controlled for by running controls in parallel whenever applicable.
Blinding	Blinding was not relevant to our study because in general, based on the prior experience of other groups in the field, these types of assays do not require blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells were obtained from Cell bank of Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences.
Authentication	Cell lines were authenticated by the supplier.
Mycoplasma contamination	Cell lines were tested and no contamination of mycoplasma.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used was listed in the database of ICLAC.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	he following mouse strains were used in the manuscript: C57BL/6J mice: female, 3-4 week-old; Ai9 mice: male , 8-15week-old; ICR mice: females, 8 week-old;
Wild animals	No wild animals were involved in this study.
Field-collected samples	The study did not involve samples collected from field.
Ethics oversight	The use and care of animals complied with the guideline of the Biomedical Research Ethics Committee of Shanghai Institutes for Biological Science, Chinese Academy of Sciences.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To isolate cells, the prepared tissues were dissociated enzymatically in an incubation solution of 5 mL Trypsin-EDTA (0.05%) at 37°C for 30min. The digestion was stopped by adding 5 ml of DMEM medium with 10% Fetal Bovine Serum (FBS). Fetal tissues were then homogenized by passing 30-40 times through a 1ml pipette tips. The cell suspension was centrifuged for 6 min (800 rpm), and the pellet was resuspended in DMEM medium with 10% FBS. Finally, the cell suspension was filtered through a 40-µm cell strainer, and tdtomato+/tdtomato- cells were isolated by FACS. Samples were found to be >95% pure when assessed with a second round of flow cytometry and fluorescence microscopy analysis.
Instrument	Cell Sorter (Beckman, MoFlo XDP)
Software	FlowJo X 10.0.7
Cell population abundance	Samples were found to be >95% pure when assessed with a second round of flow cytometry and fluorescence microscopy analysis.
Gating strategy	Positive boundaries were determined by E14.5 tdTomato+ embryos injected with Cre at zygote stage, and negative boundaries were determined by control E14.5 tdTomato- embryos that were not injected with Cre.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.