

Engineering a precise adenine base editor with minimal bystander editing

Received: 17 May 2022

Accepted: 6 September 2022

Published online: 13 October 2022

 Check for updates

Liang Chen^{1,4}, Shun Zhang^{1,4}, Niannian Xue^{1,4}, Mengjia Hong^{1,4}, Xiaohui Zhang^{1,4}, Dan Zhang¹, Jing Yang¹, Sijia Bai¹, Yifan Huang¹, Haowei Meng², Hao Wu², Changming Luan¹, Biyun Zhu¹, Gaomeng Ru¹, Hongyi Gao¹, Liping Zhong³, Meizhen Liu¹, Mingyao Liu¹, Yiyun Cheng¹, Chengqi Yi², Liren Wang¹, Yongxiang Zhao³✉, Gaojie Song¹✉ and Dali Li¹✉

Adenine base editors (ABEs) catalyze A-to-G transitions showing broad applications, but their bystander mutations and off-target editing effects raise safety concerns. Through structure-guided engineering, we found ABE8e with an N108Q mutation reduced both adenine and cytosine bystander editing, and introduction of an additional L145T mutation (ABE9), further refined the editing window to 1–2 nucleotides with eliminated cytosine editing. Importantly, ABE9 induced very minimal RNA and undetectable Cas9-independent DNA off-target effects, which mainly installed desired single A-to-G conversion in mouse and rat embryos to efficiently generate disease models. Moreover, ABE9 accurately edited the A₅ position of the protospacer sequence in pathogenic homopolymeric adenosine sites (up to 342.5-fold precision over ABE8e) and was further confirmed through a library of guide RNA–target sequence pairs. Owing to the minimized editing window, ABE9 could further broaden the targeting scope for precise correction of pathogenic single-nucleotide variants when fused to Cas9 variants with expanded protospacer adjacent motif compatibility. bpNLS, bipartite nuclear localization signals.

DNA base editors are innovative genome-editing tools catalyzing efficient base conversions without creating DNA double strand breaks (DSBs) or a requirement for donor DNA templates¹. Cytosine base editors (CBEs) are composed of Cas9 nickase (nCas9) and a cytosine deaminase domain to catalyze specific C•G-to-T•A transitions with the presence of a uracil glycosylase inhibitor (UGI)². Similarly, adenine base editors (ABEs) were developed by fusion of nCas9 to a wild-type or an evolved TadA (eTadA) (originally a transfer RNA (tRNA) adenine deaminase in *Escherichia coli*) to efficiently generate A•T-to-G•C conversions³. Unlike CBEs which also induce C to non-T side-products

and indels owing to activation of the base excision repair pathway, the first generation of ABEs (like ABE7.10) produces pure A-to-G conversions without inducing significant indels (typically $\leq 0.1\%$)³. More importantly, ABE7.10 rarely induces Cas9-independent off-target DNA editing, which has been reported in CBE-treated cells and embryos^{4,5}. These excellent features make ABEs promising tools for future clinical applications.

Great efforts have been made to improve the performance of ABEs. As TadA is a tRNA adenine deaminase, numerous occurrences of random RNA off-target editing have been reported^{6–8}. Through

¹Shanghai Frontiers Science Center of Genome Editing and Cell Therapy, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China. ²Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China.

³National Center for International Biotargeting Theranostics, Guangxi Key Laboratory of Biotargeting Theranostics, Collaborative Innovation Center for Targeting Tumor Theranostics, Guangxi Medical University, Guangxi, China. ⁴These authors contributed equally: Liang Chen, Shun Zhang, Niannian Xue, Mengjia Hong, Xiaohui Zhang. ✉e-mail: zhaoyongxiang@gxmu.edu.cn; gjsong@bio.ecnu.edu.cn; dlli@bio.ecnu.edu.cn

the introduction of point mutations in wild-type TadA and eTadA or using only an engineered eTadA, several versions of ABEs, such as ABEmax-F148A⁷ (an F148A mutation introduced to both TadA and eTadA), ABEmax-AW⁸ (with TadA E59A and eTadA V106W mutations) and SECURE-ABEs⁹ (with eTadA K20A/R21A or V82G mutations) exhibited minimized off-target edits. To improve the editing efficiency and targeting scope, two new groups of ABE variants, ABE8e¹⁰ and ABE8s¹¹, have been developed through molecular evolution of the eTadA monomer. ABE8e is the most efficient and compatible ABE variant whose activity exhibits a 3- to 11-fold improvement compared with ABE7.10, while it also expands the editing window¹⁰. ABE8e and ABE8s also showed quite high editing efficiencies in the livers of mice and non-human primates¹² or hemopoietic stem cells from patients with sickle cell anemia¹³, demonstrating their potential for gene therapeutics. However, with the increase of deamination activity, ABE8e exhibits significant Cas9-independent DNA and RNA off-target editing^{10,14,15}.

Although ABE8 variants are highly efficient, the editing window is also expanded with significant editing rates on the bystander adenines^{10,11,16}. Moreover, several studies have shown that ABE7.10 exhibits cytosine deamination activity which enables C-to-T/G/A conversions with a preference for TCN motif, demonstrating that ABEs also induce undesired bystander cytosine mutations in cell lines and animal embryos^{17–19}. It is critical to eliminate both adenine and cytosine bystander effects and Cas9-independent off-targeting editing of ABEs, especially for clinical applications. In this study, through structure-based engineering, we generated ABE9 which accurately catalyzed A-to-G conversions within a 1–2-nucleotide editing window without inducing C-to-T conversions in cells and rodent embryos. We also demonstrated it precisely corrected pathogenic single-nucleotide variants (SNVs), especially in homopolymeric adenosine sites with infinitesimally small rates of Cas9-independent RNA and DNA off-target effects.

Results

Structure-based molecular evolution of TadA-8e

ABE8e, whose deaminase component is a multiple-turnover enzyme with high processivity²⁰, edits more positions than previously reported ABEs¹⁰. We also confirmed that adenines in positions 3–12 were efficiently edited by ABE8e, suggesting a much wider editing window than ABEmax (Extended Data Fig. 1a). ABE8e also exhibited elevated cytosine bystander editing effects and increased Cas9-independent DNA off-target editing through a more sensitive orthogonal R-loop assay, which uses SaCas9 nickase instead of dSaCas9^{10,21,22} (Extended Data Fig. 1b,c). These elevated rates of undesired ABE8e editing effects encourage us to further optimize it for more accurate editing.

To increase its accuracy, we intended to evolve the TadA-8e based on its DNA-binding cryo-electron microscopy structure²⁰ (Protein Data Bank accession: 6VPC). The structure suggests that three nucleotides of the substrate, including the editing base (Fig. 1a) and the bases before and after it, are important for recognition by the deaminase. We hypothesized that mutating these residues that interacted with either the bases or the backbone of the substrate would change the environment of the binding pocket as well as the accessibility to the substrate. It might eventually reduce the non-specific binding and narrow down the editing window. Moreover, according to the apparently different size and electrophilicity of the purine ring (A) compared to the pyrimidine ring (C), these mutations would change the substrate selectivity of TadA-8e deaminase. Residues included the E27–V28–P29 loop and F148, which inserted into a valley formed by the '0' and '+1' bases; the F84, N108, L145, and Y149, which inserted into the other valley formed by the '0' and '-1' bases; and the P86/H57, which was adjacent to the editing base (Fig. 1a). Thus, ten residues were individually mutated to remove the large side chain (for example, F84T and F148A) or add a bulky residue (for example, V28F and P29W), or change between non-polar and polar residues (for example, L145T, V28N and N108V). Some highly conserved positions adjacent to the pocket (for example,

E27–P29, L145 and F148) were also included to maximize the possibility of developing a precise editing tool.

Following the above principles, 21 point mutations were constructed in TadA-8e and the activity was determined on three target sites. Deep-sequencing data of the first two targets with multiple adenines showed that the majority of the mutations reduced the editing window with a comparable or slightly decreased A-to-G efficiency compared to ABE8e, while H57D, H57Q, N108T and N108V mutations dramatically reduced the activity. In contrast to ABEmax⁷, the introduction of an F148A mutation in TadA-8e did not narrow the editing window (Fig. 1b). On the third target site previously used for evaluation of cytosine bystander mutations^{14,17}, ABEmax and ABE8e induced lots of cytosine mutations (8.83% and 45.20% in average) while V28F, V28N, N108Q, L145C, L145T and L145Q mutations exhibited high A-to-G activity on A₄ with greatly reduced cytosine conversions (ranging from 2.43% to 11.47%) (Fig. 1c). To evaluate the reduction of bystander editing and undesired cytosine conversion, the editing efficiency ratios of A₃/A₄ and C₆/A₄ were calculated. The ABE8e-N108Q construct was selected for further investigations, since it showed high A₄ editing efficiency (80.5%), relatively less A₃/A₄ and the lowest C₆/A₄ ratios (Fig. 1c).

ABE8e-N108Q reduces bystander adenine and cytosine editing

To further profile the performance of ABE8e-N108Q, 21 endogenous targets were tested in HEK293T cells by high-throughput sequencing (HTS). The first batch of 12 target sites contained multiple adenines and the other 9 targets contained mixed adenines and cytosines in the editing window. ABE8e was highly efficient (>50%) between positions A₃ to A₈ and considerable editing was also observed in a very lateral position such as A₂ or A₁₃, but ABE8e-N108Q mainly edited A₄–A₇ with almost no editing on A₉ to protospacer adjacent motif (PAM)-proximal positions (Fig. 2a and Extended Data Fig. 2a). In the remaining nine targets, we found that in addition to the TCN motif, ABE8e also edited cytosines in CCN, GCN and ACN motifs (Fig. 2b). ABE8e induced cytosine base conversions up to 39.23% (SSH2-sg10), with the highest efficiency on C₆ with an average rate of 18.02% (Fig. 2b,c and Extended Data Fig. 2b). By contrast, a significant decrease of cytosine conversions was observed in ABE8e-N108Q-treated cells with an average editing efficiency of 5.79% on C₆, although its cytosine deaminase activity was not fully eliminated (Fig. 2c). On the basis of all 21 target results, ABE8e-N108Q exhibited an identical A-to-G efficiency with ABE8e at the highest positions (82.1% versus 82.74% on A₅ and 83.62% versus 83.13% on A₆), but the major editing window was reduced to A₄–A₇ (Fig. 2d). Similar to ABE8e, ABE8e-N108Q minimally induced indels on the selected target sites (Fig. 2e). Together, ABE8e-N108Q is highly efficient with significantly reduced adenine and cytosine bystander mutation effects.

Single editing by further evolution of ABE8e-N108Q

Although ABE8e-N108Q exhibits a smaller editing window and fewer cytosine edits, we pursued a more accurate ABE featuring a single-nucleotide window and complete elimination of cytosine editing activity. We assumed that introducing more mutations in ABE8e-N108Q would further reduce bystander editing. As shown in Fig. 3a, the combination of N108Q with an additional single mutation on residues E27, P29, F84 or L145 exhibited a very stringent editing window even to a single adenine at the A₅ position. Although all three ABE8e-N108Q/L145 variants exhibited superb performance, we noticed that ABE8e-N108Q/L145T showed the most condensed editing window and high activity at these two sites (Source Data 3a), and we named it ABE9 as additional mutations were introduced into ABE8e. Since some of the mutations on N108 or L145 residue improved the performance of ABE8e (Fig. 1c), we next performed individual saturation mutation on these two residues to investigate whether other amino acid substitutions outperform ABE9. HTS data showed that ABE9 displayed the highest efficiency, lowest cytosine bystander editing and very narrow editing window compared to other 38 ABE variants (Extended Data Fig. 3a).

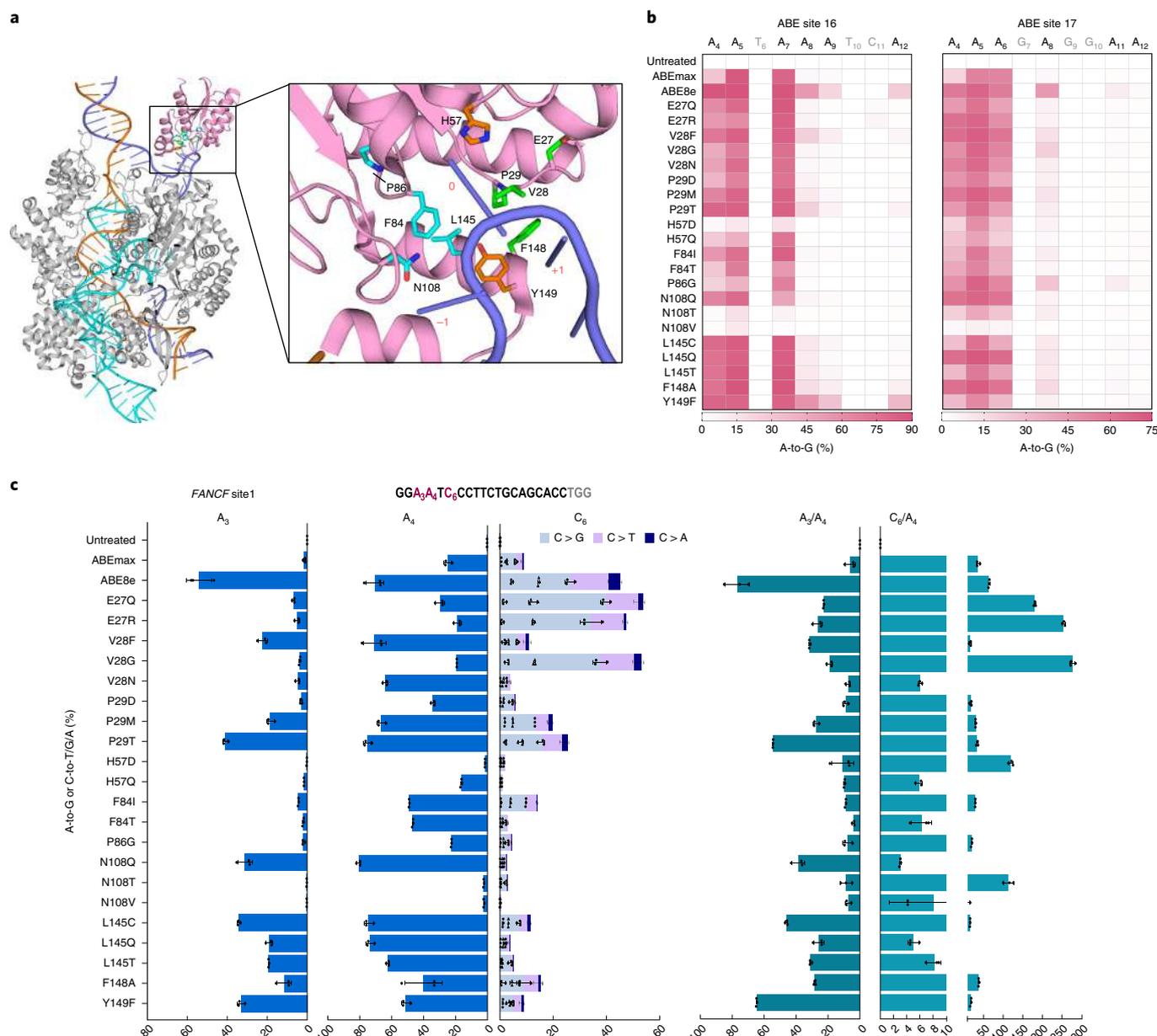


Fig. 1 | Structure-based molecular evolution of TadA-8e. a, The schematic diagram of the interplays of TadA-8e (pink) with the single-stranded DNA substrate (light blue sticks) (Protein Data Bank accession: 6VPC). Complementary strand DNA is in orange, non-complementary strand DNA is in light blue, Cas9n is in gray, and sgRNA is in cyan. Amino acids reacting with the substrate DNA are labeled on the enlarged image. The editing base is labeled '0' and the bases before and after it are labeled '-1' and '+1', respectively. **b**, The A-to-G base editing efficiency of ABE8e or ABE8e variants at two endogenous genomic loci containing multiple adenosines (ABE site 16 and ABE site 17) in

HEK293T cells. The heat map represents an average editing percentage derived from three independent experiments with editing efficiency determined by HTS. **c**, Base editing efficiency of ABE8e or ABE8e variants at an endogenous genomic locus (*FANCF* site 1) for both adenine and cytosine editing in HEK293T cells. A_3/A_4 means the ratio of undesired A_3 editing to desired A_4 editing, and C_6/A_4 means the ratio of undesired C_6 editing to desired A_4 editing. Data are mean \pm s.d. of $n = 3$ independent experiments. Statistical source data are available (Source Data Fig. 1).

After evaluation at 12 endogenous sites, we found that ABE9 showed much higher activity than ABE7.10 and slightly compromised activity compared to ABE8e, but it significantly reduced adenine bystander edits and narrowed the editing window to 1–2 nucleotides (Fig. 3b). Importantly, its activity at the adjacent A_4 or A_7 position was dramatically reduced or even eliminated at 10 of these 12 targets, and single adenine editing was observed at half of tested sites (Fig. 3b). Using the editing rate of the most efficient position to divide by the second-highest position, we further confirmed that ABE9 was the most accurate variant and showed up to 8-fold (4.3-fold in average)

discrimination of the two most efficient adenine positions compared to ABE8e-N108Q (Fig. 3c). With the editing rate of the highest position divided by the cumulative efficiencies on each edited position, similar results were obtained, suggesting that ABE9 was the most precise of the tested ABE variants (Extended Data Fig. 3b). Collectively, these results show that we developed a more accurate ABE variant ABE9, which showed a stringent and steep editing window of 1–2 nucleotides at A_3 or A_6 (Fig. 3d and Extended Data Fig. 3c). As expected, the indel rates of ABE9 are comparable or even slightly reduced compared with ABE8e and ABE8e-N108Q (Extended Data Fig. 3d).

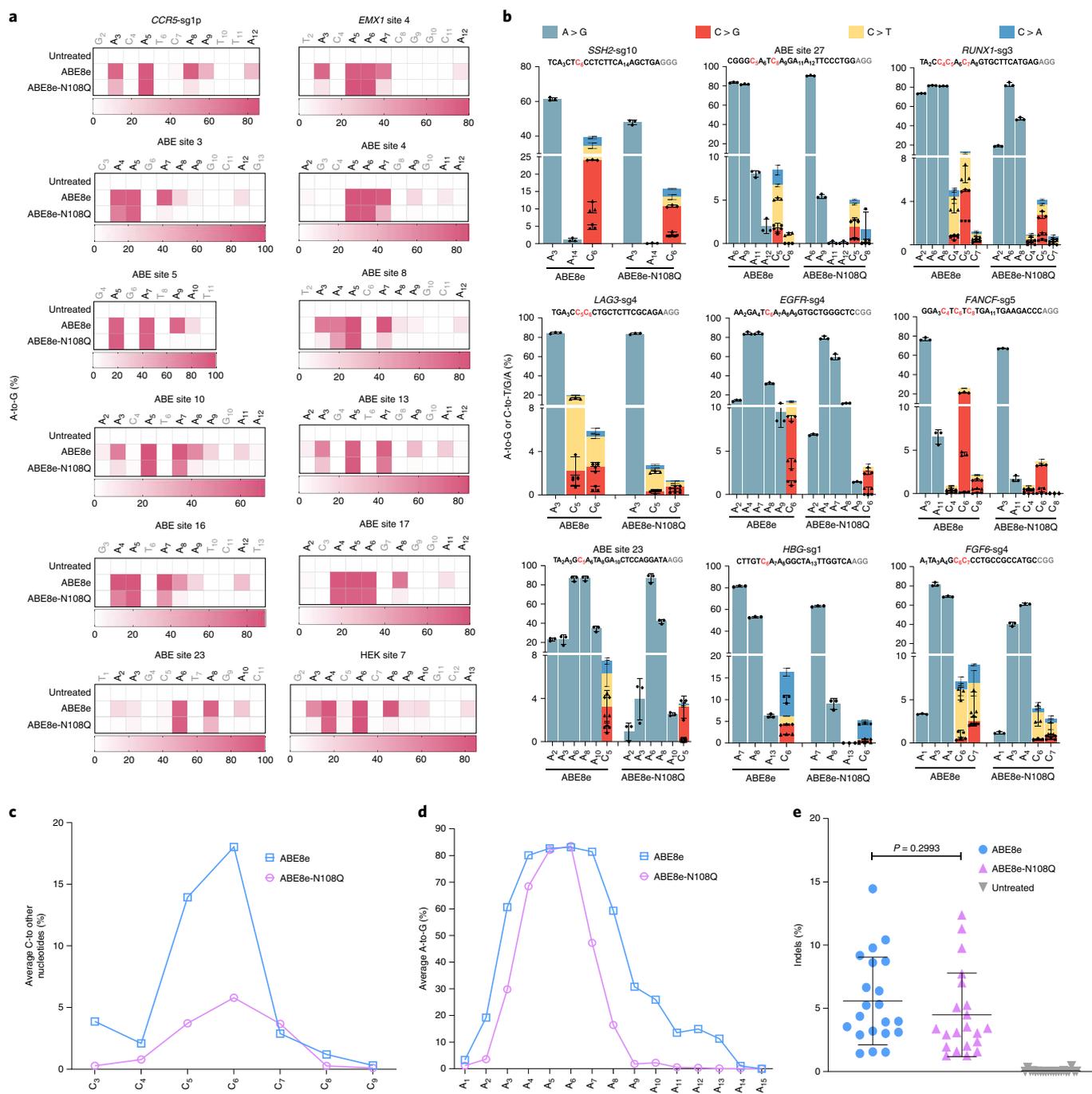


Fig. 2 | Characteristics of ABE8e-N108Q in HEK293T cells. a, The editing efficiency of ABE8e or ABE8e-N108Q was examined at 12 endogenous genomic loci containing multiple As in HEK293T cells. The heat map represents the average editing percentage derived from three independent experiments. **b**, The editing efficiency of ABE8e or ABE8e-N108Q was examined at nine endogenous genomic loci containing an NCN motif in HEK293T cells. Data are mean \pm s.d. of $n = 3$ independent experiments. **c**, Average C-to-T/G/A editing efficiency of ABE8e or ABE8e-N108Q at the nine target sites in **b**. **d**, Average A-to-G editing

efficiency of ABE8e or ABE8e-N108Q at the 21 target sites in **a**. **e**, Frequency of indel formation by ABE8e or ABE8e-N108Q at the 21 target sites in **a**, **b**. Each data point represents the average indel frequency at each target site calculated from three independent experiments. Error bar and P value are derived from these 21 data points. Data are mean \pm s.d. P value was determined by a two-tailed Student's t -test. **c**, **d**, Data represent averages from three independent experiments. Statistical source data are available (Source Data Fig. 2).

Next, 11 target sites were employed to evaluate the cytosine bystander mutation rate. As shown in Fig. 3e, ABE9 did not edit Cs in 10 of these 11 targets whereas ABE8e and ABE8e-N108Q induced considerable edits on all targets tested (efficiency lower than 1% considered as no editing). The highest cytosine conversion activity of ABE9 detected was 1.5% on C_7 of the *TIM3*-sg4 target but ABE8e and ABE8e-N108Q catalyzed 29.6% and 3.9% conversions on C_5 , respectively. According to statistical

analysis of the cytosine editing rate of the most efficient position, ABE9 strikingly decreased the cytosine bystander mutation rate by 13.2- to 147.5-fold (mean 56.2-fold) and 2.6- to 40.8-fold (10.2-fold on average) in comparison to ABE8e and ABE8e-N108Q, respectively (Fig. 3f). Moreover, we also found that ABE9 was very efficient in Hela cells and displayed a condensed editing window compared with ABE8e-N108Q, suggesting that ABE9 was suitable for variant cell lines (Extended Data Fig. 4).

Off-target analysis of ABEs in mammalian cells

To evaluate Cas9-dependent off-target activity, 44 potential off-target sites from 5 short guide RNA (sgRNA) targets were analyzed, including 17 known off-target sites identified by GUIDE-seq or ChIP-seq^{3,23} and 27 in silico-predicted off-target sites by Cas-OFFinder²⁴. We found that ABE8e induced mild off-target editing (1.04–12.29%, 4.15% on average) at 11 sites on HEK site 2, HEK site 3 and *PD-I*-sg4 loci, while ABE9 only edited two sites with comparable on-target activity and background level of indels (Fig. 4a and Extended Data Fig. 5a,b). The Cas9-independent DNA and RNA off-target editing caused by the deaminase were more unpredictable and intractable. Through an enhanced orthogonal R-loop assay^{22,23}, Cas9-independent DNA and RNA off-target effects of ABE9 with infinitesimal indels were greatly reduced compared to ABE8e (Fig. 4b,c and Extended Data Fig. 6a,b). Amazingly, the off-target activity of ABE9 was lowered to near-background levels (mean <0.3%) (Fig. 4b), indicating that it eliminated unpredictable DNA off-target activity. Through whole-genome mRNA profiling analysis, we found that RNA off-target effects of ABE9 were reduced to background level and displayed 726.1- and 117.1-fold reduction compared to ABE8e and ABE8e-N108Q, respectively (Fig. 4c). These results demonstrate that ABE9 is highly specific with infinitesimal rates of unpredictable DNA and RNA off-target activity.

Highly accurate editing by ABE9 in rodent embryos

Accurate base conversion is critical for modeling pathogenic SNVs, but ABEs or CBEs usually induce severe bystander mutations at the target sites in cells and embryos^{25–27}. To test whether ABE9 could generate precise single nucleotide conversion in embryos, ABE8e or ABE9 mRNA was co-injected with sgRNA targeting the splicing acceptor site of *Tyrosinase* gene intron 3 into mouse zygotes to model albinism. Once the splicing site was destroyed (A₅ position), exon skipping might occur to disrupt tyrosinase coding and lead to an albino phenotype (Fig. 5a). After deep sequencing of genomic DNA from F0 pups, all of the mice injected with ABE8e or ABE9 contained A₅ editing (Fig. 5b and Extended Data Fig. 7a), and almost no indels (<0.2% on average) were observed in embryos injected with ABEs (Extended Data Fig. 7b). Notably, ABE9 selectively edited A₅ in 88% (14 out of 16) of the pups and the other two pups bore very low (8.13% and 9.75%) simultaneous A₈ conversions, but only 5% (1 out of 19) of the pups generated by ABE8e injection bore the desired A₅ transition (Extended Data Fig. 7a,c). After analysis of total NGS reads from all F0 pups in the same group, ABE9 generated the desired A₅ transition in 54.32% of the reads, but only 5.1% of the reads induced by ABE8e was the desired mutation (Fig. 5c). The albino phenotype in the eyes and fur color of the founders suggested that tyrosinase activity was disrupted by ABE9-induced A₅ conversion (Fig. 5d).

We further inspected the efficiency and accuracy of ABE9 in rat embryos through targeting of a site with three adenines in an A₄–A₈ canonical editing window (Fig. 5e). As our previous data showed that only the A₆-to-G conversion, which caused D645 mutation in *Gaa* gene identified in patients with early-onset Pompe (glycogen storage disease type II) disease, lead to an obvious phenotype in rats²⁷. Through reanalysis of our published data, it showed that ABE7.10 only induced 6 of 28 (21%) pups bearing desired D645G mutation with the efficiency ranging from 6.04% to 27.94% (Extended Data Fig. 7d). By contrast,

ABE9 induced desired A₆ substitution in all 8 (100%) pups with the efficiency ranging from 36.08% to 62.41% (Fig. 5f,g and Extended Data Fig. 7e). Consistent with the data obtained in mice, ABE9 induced very limited indels similar to ABE7.10 in rats (Extended Data Fig. 7f). From HTS results of all 28 F0 rats treated with ABE7.10, the proportion of desired reads was only about 2.76% of all cumulative HTS reads, while ABE9 induced an 18.0-fold increase (49.59%) compared to that of ABE7.10-treated rats (Extended Data Fig. 7g), suggesting ABE9 was more efficient and accurate than ABE7.10. These data demonstrate that ABE9 is very efficient at generating highly accurate base installation in mouse and rat embryos.

Precise correction of pathogenic mutations by ABE9

ABE generates A-to-G conversions and potentially corrects approximately half of known pathogenic SNVs in the ClinVar database, irrespective of bystander mutations²⁸. To investigate the therapeutic potential of ABE9 for treating genetic diseases, 4 pathogenic SNPs with at least 4 consecutive adenines within positions 4–8 were tested, including missense mutations in *COL1A2* gene (causing autosomal-dominant osteogenesis imperfecta)^{29,30}, *CARD14* gene (causing psoriasis)³¹, *BVES* gene (causing muscular dystrophy)³² and *KCNAS* gene (causing common cardiac rhythm disorder)³³. ABEs were transfected into four stable cell lines containing the pathogenic variants described above. For the *COL1A2* locus, ABE8e or ABE8e-N108Q did not generate considerable conversions selectively on A₅, while ABE9 induced 34.25% desired single A-to-G conversion which was 342.5- and 21-fold higher than ABE8e and ABE8e-N108Q, respectively. Similarly, for the other three loci, ABE8e and ABE8e-N108Q only generated desired edits with frequencies of up to 2.06% and average 5.3% (0.3–11.6%), respectively, while ABE9 generated precisely corrected alleles in all four targets with an efficiency ranging from 15.53–37.22% (mean 30.19%), suggesting it was very accurate to generate single nucleotide transition (Fig. 6a and Extended Data Fig. 8a–d). These data demonstrate that ABE9 is a precise and efficient editor with the ability to correct genetic variants even in promiscuous homopolymeric sites.

Target library analysis of ABE9

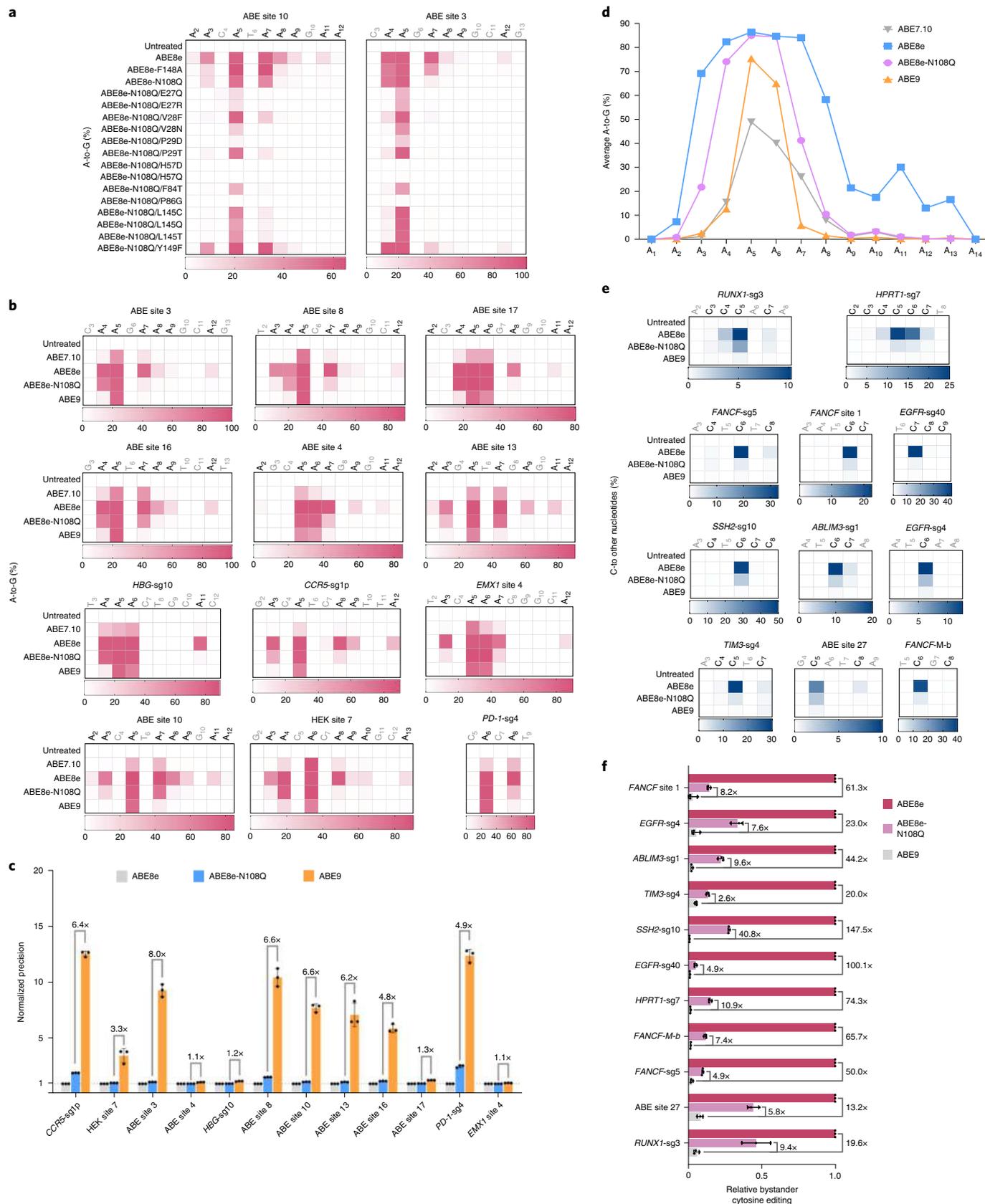
To unbiasedly characterize the performance of ABE9, we adapted the guide RNA–target pair strategy^{34,35} and synthesized a library of 9,120 oligonucleotides with all possible 6-mers containing at least an adenine and a cytosine across positions 4 to 9 of a protospacer (Methods). The oligonucleotide library was stably integrated into the genome of HEK293T cells via Tol2 transposon followed by stable transfection of a given base editor (Fig. 6b). We maintained an average 99% coverage of >300× per guide–target pair throughout the culturing process (Supplementary Table 4). Subsequently, the target region was amplified and sequenced at an average depth of 860 per target. The average editing efficiency of ABE8e, ABE8e-N108Q and ABE9 was 31.9%, 28.7%, 25.3% on position 5 respectively, suggesting the experiment was successful (Extended Data Fig. 9). The editing efficiency of the highest position in each target was considered as 100%, and the relative activity of other positions was determined comparing with the highest position. Analysis of the editing outcomes from three distinct base editors showed that ABE8e (evaluated 9,059 sgRNAs) had a wide editing window ranging from positions 3–12 with a major window (>50%) from 4–9,

Fig. 3 | Evolution and characterization of single A-to-G base editor. a, The A-to-G base editing efficiency of ABE8e-N108Q and its combination variants at 2 endogenous genomic loci containing multiple As (ABE site 10 and ABE site 3) in HEK293T cells. **b**, The A-to-G editing efficiency of ABE7.10, ABE8e, ABE8e-N108Q or ABE9 was examined at 12 endogenous genomic loci containing multiple As in HEK293T cells. **c**, The normalized precision (ABE8e is used for standardization) is defined as the highest or second-highest A-to-G base editing of ABE8e-N108Q or ABE9 at the 12 target sites in **b**. Data represent mean ± s.d. from three independent experiments. **d**, Average A-to-G editing efficiency of ABE7.10, ABE8e, ABE8e-N108Q or ABE9 at the 12 target sites in **b**. Data represent

mean from three independent experiments. **e**, The C-to-T/G/A editing efficiency of ABE9 was examined at 11 endogenous genomic loci containing multiple Cs in HEK293T cells. **f**, The normalized ratio (ABE8e is used for standardization) of the highest C-to-T/G/A editing efficiency of ABE8e-N108Q or ABE9 at 11 target sites in **e**. The numbers aside bars display the fold changes of ABE9 in reducing cytosine conversions compared with ABE8e and ABE8e-N108Q. Data represent mean ± s.d. from three independent experiments. In **a**, **b** and **e**, the heat map represents average editing percentage derived from two or three independent experiments. Statistical source data are available (Source Data Fig. 3).

while ABE8e-N108Q (evaluated 9,071 sgRNAs) narrowed the window to positions 4–7 (Fig. 6c). As expected, ABE9 (evaluated 8,954 sgRNAs) presented an extremely narrowed editing window of 1–2 nucleotides with the highest efficiency on position 5. Profiling of the motif preferences of the ABE9 showed that similar to ABE8e, they were suitable for a

wide range of accurate A-to-G editing without strict motif requirements, suggesting their accuracy was dependent on the position relative to the protospacer but not on sequence context (Fig. 6d). As determined by thousands of sgRNAs, it suggests that ABE9 is very accurate to preferentially edit adenines in position 5 of the protospacers.



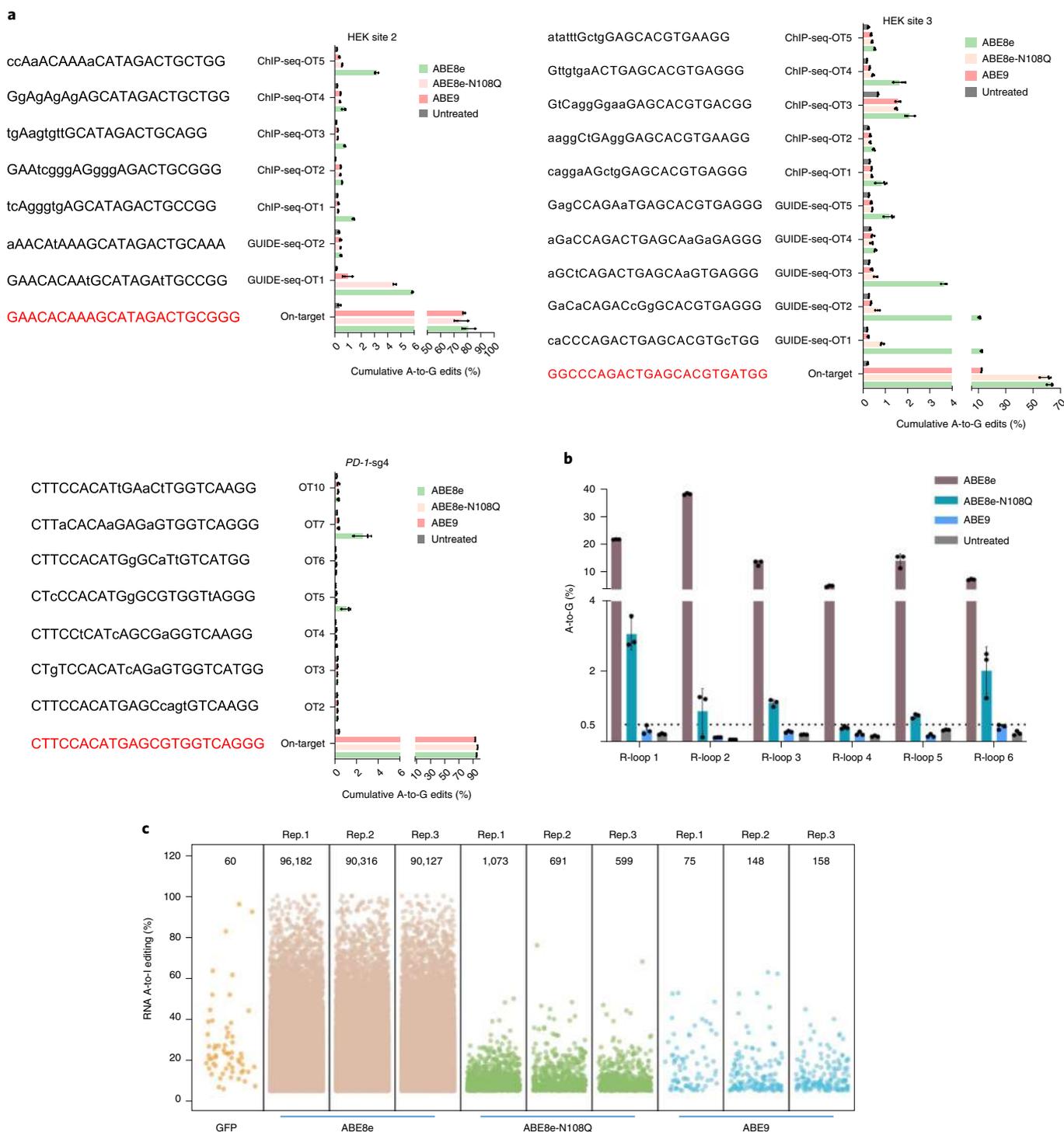


Fig. 4 | Off-target mutation assessment of ABE9. a, Cas9-dependent DNA on- and off-target analysis of the indicated targets (HEK site 2, HEK site 3 and *PD-1-sg4*) by ABE8e, ABE8e-N108Q and ABE9 in HEK293T cells. Data are mean \pm s.d. of $n = 2$ independent experiments for HEK site 2-GUIDE-seq-OT1 and 2 treated with ABE8e-N108Q, and $n = 3$ independent experiments for the other biological samples. **b**, Cas9-independent DNA off-target analysis of ABE8e, ABE8e-N108Q and ABE9 using the modified orthogonal R-loop assay at each R-loop site with

nSaCas9-sgRNA plasmid. Data are mean \pm s.d. of $n = 3$ independent experiments. **c**, RNA off-target editing activity by ABE8e, ABE8e-N108Q and ABE9 using RNA-seq. Jitter plots from RNA-seq experiments in HEK293T cells showing efficiencies of A-to-I conversions (y-axis) with ABE8e, ABE8e-N108Q and ABE9 or a GFP control. Each biological replicate (Rep.) and total numbers of modified bases are listed at the top. Statistical source data are available (Source Data Fig. 4).

Discussion

Highly efficient and precise correction of single-nucleotide pathogenic mutation is demanded for gene therapy to reach its potential. Using structure-based design and molecular evolution of Tada-8e, we have generated ABE9, which efficiently edits adenines in a 1–2-nucleotide

window without cytosine editing activity. To minimize the editing window of base editors, structure-based molecular evolution has been successfully leveraged to obtain new editors, such as BE4max-YE1 and YEE variants, which catalyze conversions within a 1–2-nucleotide window³⁶, and eA3A-BE preferentially editing in a TCN motif³⁷ and

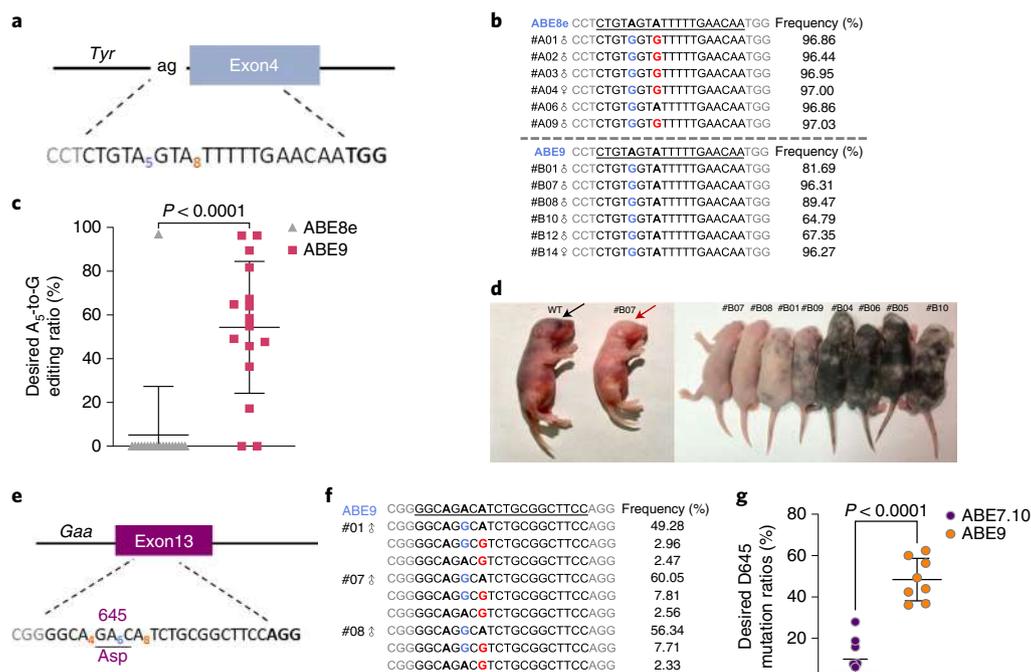


Fig. 5 | Examination of precision in rodent embryos with ABE9. a, The splicing acceptor sequence in intron 3 of the mouse *Tyr* gene. The 'ag' sequence of the splice acceptor site is shown in black. The sgRNA target and PAM sequence are both shown in black (PAM in bold). Target A₅ is in blue with bystander A₈ in red. **b**, Genotyping of representative F0 generation pups from mouse embryos microinjected with ABE8e or ABE9. The guanines converted from editable As indicate desired editing in A₅ (blue) or undesired in A₈ (red). **c**, Single A₅-to-G conversion ratio in F0 mice induced by ABE8e ($n = 19$) or ABE9 ($n = 16$). **d**, Phenotype of F0 generated by microinjection of sgRNA and ABEs. The photo on the left was taken when the mouse was 7 days old, while the right one was

at 14 days old. WT, wild type. **e**, The target sequence of exon 13 (dark purple) in the rat *Gaa* gene. The sgRNA target sequence where target A₅ is in blue with bystander A₅ and A₈ in red is shown in black (PAM in bold). The triplet codon of D645 is underlined. **f**, Genotyping of representative F0 generation pups from rat embryos microinjected with ABE9 (desired editing in blue or undesired in red). **g**, Desired D645 mutation ratios in F0 rats induced by ABE7.10 ($n = 28$) or ABE9 ($n = 8$). **b, f**, The percentage on the right represents the frequency determined by the rate of indicated mutant alleles to total alleles counts. **c, g**, Data are mean \pm s.d. and P values (3.6×10^{-6} in **c**, 8.7×10^{-16} in **g**) was determined by a two-tailed Student's t -test. Statistical source data are available (Source Data Fig. 5).

A3G-BEs selectively editing the second C in a CC motif³⁸. Although ABEmax-F148A has been shown to reduce the editing window⁷, very limited effects have been observed when it has been transferred to Tada-8e (Fig. 1b), indicating the experiences from Tada7.10 could not be directly transferred to Tada-8e.

More complicated than CBEs, ABEs are capable of catalyzing both adenines and cytosines in a similar editing window¹⁷. Since the editing window of C-to-T is overlapped with that of A-to-G, it is impracticable to eliminate their cytosine deamination activity through reducing the editing window. While we were completing this project, Bae and colleagues reported introduction of D108Q in ABEmax or N108Q in ABE8e could reduce their cytosine deaminase activity¹⁴, which was consistent with our current study, suggesting that residue 108 was critical for the discrimination of substrates such as adenines and cytosines. The previous study also showed this residue was important for the recognition of single-stranded DNA substrates as the D108N mutation was pivotal for the generation of eTadA, the unnatural DNA adenine deamination³. Moreover, the combinational mutation in ABEmax (Tada-E59A + N108W/Q) displayed greatly reduced RNA editing and preferentially catalyzing adenine conversions at protospacer position 5 but the activity was compromised⁸. It is consistent with our findings that ABE8e-N108Q exhibited reduced editing window and RNA off-target effects (Figs 3b,d and 4c).

As for the discrimination between cytosines and adenines, we speculated that the mutation of N108 to a larger side chain residue (Q) would expel the backbone of its substrate. It apparently affected the deamination of cytosines greater than adenines since the pyrimidine ring of cytosines needs to be shifted further toward the pocket for the catalytic reaction to happen. However, Tada-8e-N108Q still

retained considerable cytosine deaminase activity (Fig. 3e,f) until the introduction of a second mutation, L145T, which nearly abolished cytosine conversions and further narrowed the adenine editing window to 1–2 nucleotides without apparently sacrificing on-target adenine conversion efficiency. We found that introduction of variant mutations at L145 had similar effects on reducing the editing window and cytosine bystander mutation as N108Q, suggesting the L145 position was a previously unnoticed residue, which was also critical for substrate discrimination. It was further supported by saturation mutation analysis on the L145 residue as most of the substitution exhibited compromised cytosine editing efficiency (Extended Data Fig. 3a). As the L145 residue is located relatively distal to the target base, the mutations may adjust the pocket indirectly by influencing the positions of its nearby essential residues, such as P29, F84, N108 and Y149. Especially, the L145T + N108Q double mutants performed the best on adenine editing while removing the bystander cytosine editing, suggesting that the combination of the two mutations within the pocket somehow precisely adopted adenine versus cytosine; however, the detailed mechanism still awaits further structural study. F84 is also a critical residue identified in the initial generation of eTadA³. It is located within the pocket right below the target base ring and it forms a triangle platform together with Y149 and V28 to hold the base ring of the substrate. Additionally, we found that V28 could be a critical position involved in the discrimination of cytosines and adenines, since whereas V28F and V28N showed a significant decrease of cytosine conversions, V28G had opposite effects (Fig. 1c), suggesting it is possible to innovate pure CBE, C-to-G base editors or dual-base editors which are capable of spontaneous adenine and cytosine conversions through further engineering of Tada-8e.

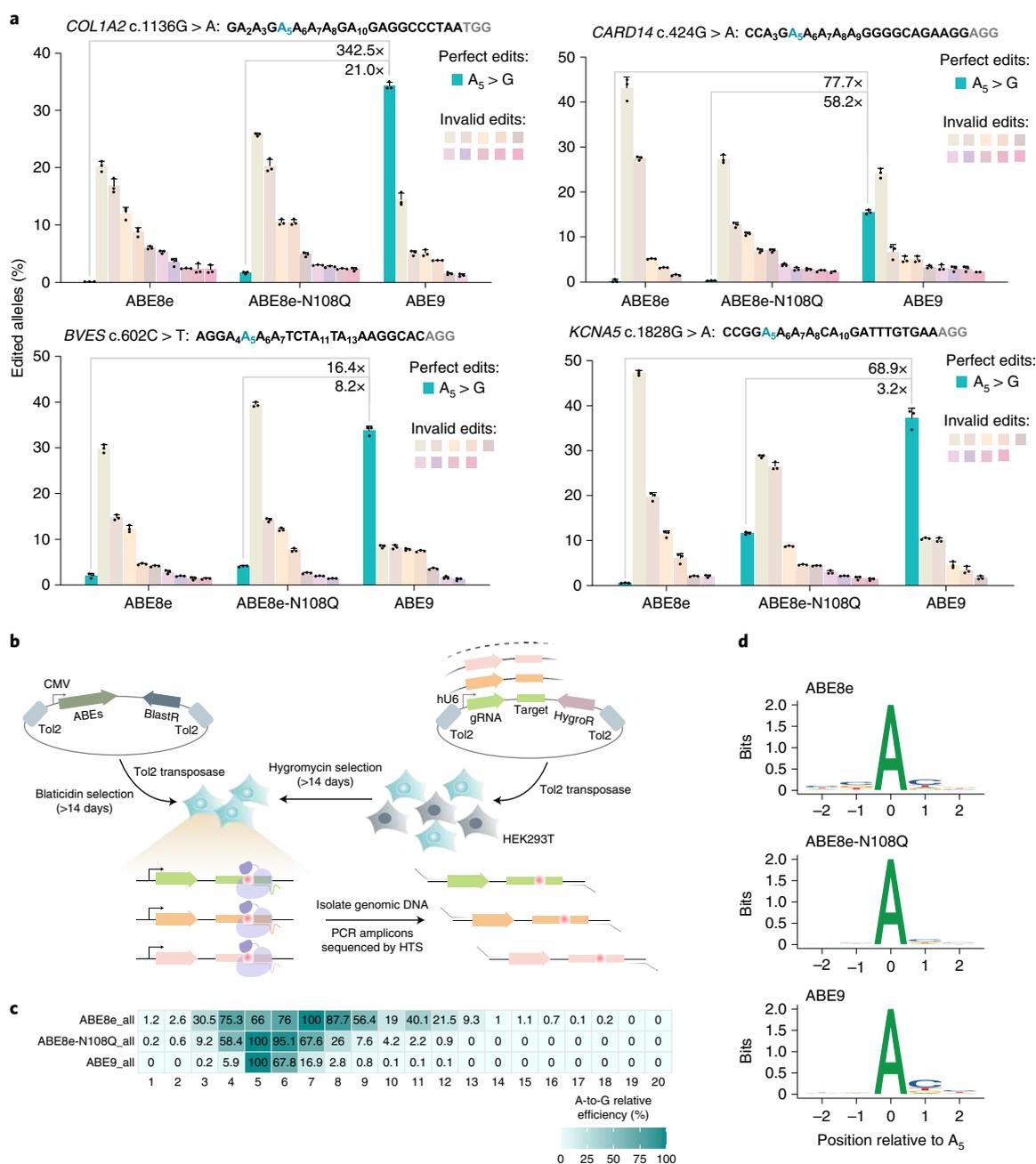


Fig. 6 | Correction of human pathogenic mutations in mammalian cells and target library analysis to unbiasedly characterize ABE9. a, Comparison of correcting pathogenic mutations induced by ABEs in four stable HEK293T cell lines, including *COL1A2* c.1136G > A ($n = 3$), *KCNA5* c.1828G > A ($n = 3$), *BVES* c.602C > T ($n = 3$), *CARD14* c.424G > A ($n = 2$ for the tenth or ninth invalid edits induced by ABE8e-N108Q or ABE9, $n = 3$ for the other samples). Base editing efficiency was determined by HTS. Data are mean \pm s.d. Desired A₅-to-G percentiles of alleles (green bar) are exhibited, while percentiles of the top ten invalid allele types are presented and percentiles of invalid allele types less

than 1% are omitted. The numbers above green bars display the fold changes of ABE9 in desired A₅-to-G percentiles compared with ABE8e and ABE8e-N108Q. **b**, Schematic of target library analysis. **c**, Analysis of relative editing efficiency of ABE8e, ABE8e-N108Q and ABE9. The heat map represents editing efficiency computed relatively to the highest A-to-G base editing of the protospacer. Positions of the protospacer are shown at the bottom of each heat map, counting the PAM as positions 21–23. **d**, Motif visualization of ABE8e, ABE8e-N108Q and ABE9 in fifth-adenine-containing cassettes. Statistical source data are available (Source Data Fig. 6).

Developing a base editor with a refined editing window is challenging, especially editing a specific base within promiscuous homopolymeric sites. Recently, a precise ABE-NG variant has been developed through engineering of TadA-8e³⁹, but its major window is A₄-A₇, which is much wider than 1–2-nucleotide editing window of ABE9. Moreover, questions remain about bystander cytosine editing effects and whether its 4-nucleotide major window could be adapted to SpCas9. Using selected target sites in cells and rodent embryos, we determined that

ABE9 was accurate with a very narrow editing window. More importantly, through a guide RNA–target pair library containing over 9,000 targets, the data showed that ABE9 could be considered as an ABE focusing on a 1–2-nucleotide editing window with the highest efficiency A₅ (Fig. 6c). To our knowledge, it is potentially the most accurate ABE to date. As SpRY almost does not require any PAM sequence, ideally ABE9-SpRY could precisely target any adenine through an appropriate sgRNA for broad targeting scope. Importantly, ABE9 induces almost

no off-target effects (either Cas9-dependent or -independent) at both DNA and RNA levels, which is important not only for basic research but also critically important for clinical 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/s41589-022-01163-8>.

References

- 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).
- 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).
- 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).
- Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
- Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
- Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
- Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
- Rees, H. A., Wilson, C., Doman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **5**, eaax5717 (2019).
- Grunewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).
- Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
- Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* **38**, 892–900 (2020).
- Musunuru, K. et al. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).
- Newby, G. A. et al. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature* **595**, 295–302 (2021).
- Jeong, Y. K. et al. Adenine base editor engineering reduces editing of bystander cytosines. *Nat. Biotechnol.* **39**, 1426–1433 (2021).
- Li, S., Liu, L., Sun, W., Zhou, X. & Zhou, H. A large-scale genome and transcriptome sequencing analysis reveals the mutation landscapes induced by high-activity adenine base editors in plants. *Genome Biol.* **23**, 51 (2022).
- Xu, L. et al. Efficient precise in vivo base editing in adult dystrophic mice. *Nat. Commun.* **12**, 3719 (2021).
- Kim, H. S., Jeong, Y. K., Hur, J. K., Kim, J. S. & Bae, S. Adenine base editors catalyze cytosine conversions in human cells. *Nat. Biotechnol.* **37**, 1145–1148 (2019).
- Lee, H. K. et al. Targeting fidelity of adenine and cytosine base editors in mouse embryos. *Nat. Commun.* **9**, 4804 (2018).
- Liu, Z. et al. Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat. Commun.* **9**, 2338 (2018).
- Lapinaite, A. et al. DNA capture by a CRISPR–Cas9-guided adenine base editor. *Science* **369**, 566–571 (2020).
- 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).
- Wang, L. et al. Eliminating base-editor-induced genome-wide and transcriptome-wide off-target mutations. *Nat. Cell Biol.* **23**, 552–563 (2021).
- Zhang, X. et al. Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. *Nat. Cell Biol.* **22**, 740–750 (2020).
- Hwang, G. H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinf.* **19**, 542 (2018).
- Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
- Ryu, S. M. et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* **36**, 536–539 (2018).
- Yang, L. et al. Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell* **9**, 814–819 (2018).
- Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
- Symoens, S. et al. Type I procollagen C-propeptide defects: study of genotype-phenotype correlation and predictive role of crystal structure. *Hum. Mutat.* **35**, 1330–1341 (2014).
- Forlino, A., Cabral, W. A., Barnes, A. M. & Marini, J. C. New perspectives on osteogenesis imperfecta. *Nat. Rev. Endocrinol.* **7**, 540–557 (2011).
- Jordan, C. T. et al. Rare and common variants in CARD14, encoding an epidermal regulator of NF-kappaB, in psoriasis. *Am. J. Hum. Genet.* **90**, 796–808 (2012).
- Schindler, R. F. et al. POPDC1^{S201F} causes muscular dystrophy and arrhythmia by affecting protein trafficking. *J. Clin. Invest.* **126**, 239–253 (2016).
- Yang, Y. et al. Novel KCNA5 loss-of-function mutations responsible for atrial fibrillation. *J. Hum. Genet.* **54**, 277–283 (2009).
- Arbab, M. et al. Determinants of base editing outcomes from target library analysis and machine learning. *Cell* **182**, 463–480 e430 (2020).
- Wang, D. et al. Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. *Nat. Commun.* **10**, 4284 (2019).
- 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).
- Gehrke, J. M. et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
- Lee, S. et al. Single C-to-T substitution using engineered APOBEC3G-nCas9 base editors with minimum genome- and transcriptome-wide off-target effects. *Sci. Adv.* **6**, eaba1773 (2020).
- Tu, T. et al. A precise and efficient adenine base editor. *Mol. Ther.* **30**, 2933–2941 (2022).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Molecular cloning

Plasmid DNA sequences and primers (Biosune) used can be found in Supplementary Tables 1–3 and the Supplementary Notes. ABE7.10 (#102919), ABE8e (#112095) and ABE8e (#138489) were obtained from Addgene. For ABE plasmid constructions, DNA Polymerase KOD-Plus-Neo (Toyobo, no. KOD-401) and MultiS One Step ClonExpress Kit (Vazyme, no. C113) were adopted. sgRNA expression plasmids were constructed as described previously^{23,40}. Specifically, oligonucleotides in Supplementary Table 1 were under incubating conditions of 95 °C for 5–8 min and then naturally cooled to room temperature. BbsI-linearized U6-sgRNA(sp)-EF1 α -GFP (Thermo Fisher Scientific) was ligated with annealed oligonucleotides.

Cell culture

Dulbecco's modified Eagle's medium (DMEM, Gibco) was used to foster both HEK293T cells (ATCC CRL-3216) and Hela cells (ATCC CCL-2), and DMEM was mixed with an antibiotic of 1% penicillin–streptomycin (Gibco) and fetal bovine serum (FBS; Gibco) of 10% (v/v). The cell line was maintained at 37 °C with 5% CO₂ in the incubator.

Cell transfection and genomic DNA extraction

Before cell transfections, 24-well plates (Corning) were used to culture HEK293T or Hela cells until the confluency was at approximately 80%. Next, 750 ng of ABEs and 250 ng of sgRNA plasmids were transfected simultaneously with polyethyleneimine (PEI; Polysciences) according to manufacturer instructions. After 72 h, transfected cells were digested with 0.25% trypsin (Gibco). For HEK293T cells, certain cell populations with positive GFP signals were gathered in sorting, whereas Hela cells were unsorted and directly subjected to DNA extraction. Genome DNA was isolated by the use of Genomic Kit (Tiangen Biotech, no. DP348-03) according to the manufacturer's instructions. For rodent DNA extractions, mouse or rat tail tip genomic DNA was isolated employing One-Step Mouse Genotyping Kit (Vazyme, no. PD101) according to the manufacturer's instructions.

In silico-predicted off-target sites by Cas-OFFinder

The selection principle of off-target sites was performed as depicted previously⁴¹. In brief, PAM type and target genome need to be determined at the first step on the Cas-OFFinder website, then put 20 bp target sequences of interest into the text box to initiate the searching program for potential off-target sites with normally setting parameters up to three nucleotides mismatches and one DNA bulge.

Modified R-loop assay

In this study, Cas9-independent DNA off-target analysis was using the modified orthogonal R-loop assay with nSaCas9-sgRNA plasmid at each R-loop site. For transfection, 250 ng of SpCas9 sgRNA, 300 ng of ABE8e, ABE8e-N108Q or ABE9 and 300 ng of nSaCas9-sgRNA were co-transfected into HEK293T cells using PEI. After 72 h, transfected cells were digested with 0.25% trypsin (Gibco). Genomic DNA was isolated using Genomic Kit (Tiangen Biotech, no. DP348-03) according to the manufacturer's instructions.

Total mRNA preparation

In the experimental process of RNA off-target, 10-cm dishes were employed to culture HEK293T cells until cell confluency was at approximately 80% and cells were transfected with 25 μ g of Cas9n-P2A-GFP, ABE8e-P2A-GFP, ABE8e-N108Q-P2A-GFP or ABE9-P2A-GFP using PEI. After 72 h, transfected cells were digested with 0.25% trypsin (Gibco) for sorting on FACS Aria III (BD Biosciences) using FACSDiva v.8.0.2 (BD Biosciences). For specific cell population gating conditions in sorting see Supplementary Fig. 1. Roughly 400,000 cells (top 15% of cell populations with positive GFP signals) were gathered, and RNA of all samples was extracted according to standard instructions.

RNA sequencing experiments

For the preparation of each RNA sequencing sample, the input material was derived from a total of 3 μ g RNA. Before sequencing, Ultra RNA Library Kit for Illumina (NEB) was used to generate RNA libraries under standard instructions. For each sample, index codes were added to attribute sequences. Specifically, magnetic beads of poly-T oligonucleotides were employed to purify mRNA from the total amount of RNA. Under the higher temperature solution of NEBNext First Strand Synthesis Buffer (5 \times), divalent cations were added to fragment samples. For the synthesis of first-strand cDNA, random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-) were employed. Subsequently, DNA Polymerase I and RNase H were added in the synthesis of second-strand cDNA. Exonuclease and polymerase activities converted overhangs of cDNA into blunt ends and 3' ends of DNA fragments were adenylated before a hairpin loop structure was ligated for the sake of hybridization. cDNA fragments of 250–300 base pairs were obtained via the purification of the AMPure XP system (Beckman Coulter) and subsequently were incubated with 3 μ l USER Enzyme (NEB) for 15 min at 37 °C and then 5 min at 95 °C before PCR. Next, PCR was conducted with the use of universal PCR primers, Index (X) Primer and Phusion High-Fidelity DNA polymerase. Lastly, through the AMPure XP system and Agilent 2100 Bioanalyzer system, libraries were purified and assessed in quality. On a cBot Cluster System, the clustering of index-added samples was generated with the use of TruSeq PE Cluster Kit v3-cBot-HS (Illumina) under the standard instructions. After cluster generation, paired-end reads of 125 base pairs/150 base pairs were obtained when libraries were sequenced on an Illumina NovaSeq 6000 platform.

Transcriptome-wide RNA analysis

For the RNA sequencing (RNA-seq) analysis, adapter sequences were first removed from reads in the high-throughput sequencing data using Trim Galore (v.0.6.6) (<https://github.com/FelixKrueger/TrimGalore>), and aligned to the hg38 genome (<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/>) using STAR⁴² (v.2.7.1a). Aligned BAMs were tag added, and sorted with SAMtools⁴³ (v.1.14). Duplication was removed using the Picard MarkDuplicates module (v.2.23.9) (<https://github.com/broadinstitute/picard>) and unmapped reads were filtered using SAMtools. BAMs were converted to mpileup format with SAMtools, which records integrated mutation information. The significant mutation information was extracted on the basis of mpileup files. The sites with coverage higher than 25, a mutation count at least 6 and mutation ratios over 5% were subsequently collected as filtered sites. As the edits found in mpileup files were filtered by removing the sites existing in the Cas9n-transfected condition, the sites only in Cas9n-transfected cases were the control.

Animal manipulations

The manipulation of rodent embryos was described previously⁴⁴. Specifically, 6–10-week-old female C57BL/6 J, ICR strain mice and Sprague–Dawley strain rats purchased from Laboratory Animal Center in Shanghai were housed at 20–22 °C with 40–60% humidity in facilities free of pathogens on a 12 h light–dark cycle. Water and food were offered ad libitum. All animal experiments permitted by the Center for Animal Research in East China Normal University obeyed rules drafted by the Association for Assessment and Accreditation of Laboratory Animal Care in Shanghai. sgRNA with chemical modification was synthesized by GenScript, and mRNA was prepared as previously described²⁷. The T7 promoter was introduced into ABE8e or ABE9 template with T7-mRNA (ABE8e/ABE9)-F/R primers (Supplementary Table 2). For the transcription of ABE8e or ABE9 mRNA, mMACHINE T7 Ultra Kit (Ambion, no. AM1345) was utilized and mRNA was subsequently subject to purification with MEGAclear Kit (Ambion, no. AM1908). With the Eppendorf TransferMan NK2 micromanipulator, mixed samples of ABE8e or ABE9 mRNA of (100 ng μ l⁻¹) and sgRNA (200 ng μ l⁻¹) were injected into zygotes after diluting with water free

of nuclease. Pseudo-pregnant female mice or rats served as receptors of injected zygotes that required prompt transfers for mice or after overnight culture in KSOM medium at 37 °C with sufficient humidity and 5% CO₂ for rats.

Generation of stable cell line disease models

The 150-base-pair fragments of G-C-to-A-T disease-associated genes from ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>) were assembled into a modified lenti-vector from lentiCRISPR v2 (#52961), obtaining transfer plasmids (Lenti *COL1A2*-EF1 α -DsRed-P2A-puro, Lenti *CARD14*-EF1 α -DsRed-P2A-puro, Lenti *BVES*-EF1 α -DsRed-P2A-puro or Lenti *KCN A5*-EF1 α -DsRed-P2A-puro). Twenty-four-well plates were employed to foster HEK293T cells. After 12–16 h, cells with approximate confluency of 80% were co-transfected with 300 ng transfer plasmids, 300 ng pMD2.G and 300 ng psPAX2 using PEI. After 48 h of transfection, the collection of virus-containing supernatant was performed, and then a 0.45- μ m low protein binding membrane (Millipore) was used to carry out filtration. HEK293T cells were seeded into 12-well plates at approximately 40–50% confluency and 50 μ l filtered virus-containing supernatant was added to the 12-well plates. After 24 h, when cells were fully transduced with lentivirus, new plate wells were supplemented with puromycin (1 μ g ml⁻¹) as selection pressure. After 72 h, cells that fulfilled the requirement of the single-copy integration were collected and consistently cultured for future transfections.

Library design

The architecture of the oligonucleotides in the guide–target pair library was designed as previously depicted³⁴. Each oligonucleotide contains a full-length sgRNA with a corresponding cassette targeted by the sgRNA. The spacers of the sgRNAs fulfilled the following criteria. (1) Each spacer is initiated with a guanine. (2) The 4th to 9th positions of the spacers are composed of all possible 6-mers with at least an adenine or a cytosine. The 6-mers were surrounded by random 2-mers and 11-mers at the 5' and 3' end, respectively. (3) Spacers with five consecutive thymines were avoided for it might impede the transcription. Each targeted cassette contained a 20-base-pair target sequence followed by an 'NGG' PAM. Wild-type human DNA sequences of random selections flanked the target sequence.

Integration of the library and cell culture

The oligonucleotides were assembled into a modified pBlueScript backbone containing a U6 promoter and a hygromycin-resistant gene (hygro). The U6-sgRNA and hygro cassette were flanked by Tol2 sites to ensure its integration by Tol2 transposon. The construction and amplification of the library were finished by GENEWIZ Biotechnology. For the library integration, 10-cm plates (Corning) were utilized to foster HEK293T cells. We co-transfected the Tol2 transposon plasmid (10 μ g) and library mixture (10 μ g) into HEK293T cells at approximately 90% confluency. To facilitate sgRNA integration, cells were selected with hygromycin B (25 μ g ml⁻¹) (Thermo Fisher Scientific, no. 10687010) 1 day after transfection, lasting for >14 days, during which over 90% of cells were screened out. The screening was performed on at least 20 plates to ensure library coverage. When the HEK293T cells were once again at approximately 90% confluency, the second-round transfection was conducted by co-transfection of a Tol2 transposon plasmid (10 μ g) and a base editor plasmid (ABE8e, ABE8e-N108Q or ABE9) (10 μ g) that contained a blasticidin resistance gene and Tol2-transposase binding sites. The next day, 10 μ g ml⁻¹ blasticidin S HCL (Thermo Fisher Scientific, no. A11139-03) was used for the second-round selection, lasting for >14 days. Again, the selection strength was adjusted so that over 90% of the cells were killed after 2–3 days of selection. As the density of cells reached 80–90%, cells were digested with 0.25% trypsin for subsequent DNA extraction (see above). After DNA extraction, 100 ng DNA was subjected to PCR using the primers listed in Supplementary Table 2 and amplicons of target regions were obtained. The resulting amplicons were sequenced using the GENEWIZ Biotechnology HTS

platform. The complete sequences of the mentioned plasmids and base editor sequences are appended in the Supplementary Notes.

Library genomic DNA extraction

Digest cells with 0.25% trypsin and centrifuge at 100g at room temperature for at least 3 min. Cell pellets were resuspended and washed with proper volumes of phosphate-buffered saline (PBS) once followed by lysing with protein K at 55 °C for at least 1 h until the lysate became relatively clear. An equal volume of phenol-chloride was added into the lysate, followed by vortexing for at least 1 min. The mixture was incubated at room temperature for at least 10 min for phase separation and was centrifuged at 14,000g for at least 15 min. An equal volume of chloride was added after carefully removing the aqueous layer into a new tube, and the mixture was vortexed for at least 1 min and incubated at room temperature for 10 min. The mixture was centrifuged again at 14,000g for at least 15 min. A 1/10th volume of NaOAc (3.5 M) was added along with 2.5 volumes of ethanol after carefully removing the aqueous layer into a new tube and the mixture was incubated at –20 °C overnight. After the overnight incubation, samples were subsequently centrifuged at 14,000g for at least 30 min at room temperature. The DNA pellet was rinsed twice with 75% ethanol and the DNA was dissolved with Nuclease-Free Water (Ambion, no. AM9932).

Editing efficiency calculation of the library and motif visualization

The JavaScript version of fastq-join (<https://github.com/brwnj/fastq-join>) firstly joins two fastq files from HTS. The combined fastq files were aligned to all of the amplicons in the library using BWA-mem (0.7.17-r1188) and the reads were divided for each amplicon to determine the connection between the amplicons and the sequenced reads. Reads with many equally plausible alignments were detected by the random mode. To minimize the influence of PCR amplification, targets with sequencing depths more than 20 times higher than the average depth of the library were abandoned for every library. Next, all of the reads were aligned to the corresponding amplicon pairwise using EMBOSS needle (v.6.6.0.0). Only the reads that matched the following criteria were chosen for analysis: 10-base-pair sequences upstream and downstream of the 20-base-pair target sites completely matched the consensus sequences; the target sites included no detectable indels or degenerate base Ns. The editing type, the total number of reads aligned to amplicons and the number of edited reads at each position were then analyzed to calculate the absolute editing efficiency of each type at each site. Specifically, the reads were filtered to merely include loci with editing greater than 0% when calculating the absolute mean efficiency. Besides, the relative editing efficiency was also computed relatively to the highest absolute editing efficiency. The matching sgRNA was accumulated once for each edited read when enriching motifs since the effectiveness of sgRNA varied greatly. The motifs edited at A_n were tallied and visualized using the 'ggseqlogo' package in R.

HTS sequencing and data processing

Genomic DNA (100–150 ng) was subjected to PCR with the primers listed in Supplementary Tables 2 and 3 and amplicons of interest in on- and off-target studies were obtained. In the preparation of HTS libraries, adapter sequences (5'-GGAGTGAGTACGGTGTGC-3' forward; 5'-GAGTTGGATGCTGGATGG-3' backward) were added at the 5' end of DNA via specific site primers and DNA polymerase (KOD-Plus-Neo). Subsequently, the second-round PCR was implemented with primers that feature distinct barcode sequences. The resulting mixtures of DNA libraries were sequenced on the Illumina HiSeq X Ten platform. BE-Analyzer²⁴ was employed to analyze HTS data to access adenine or cytosine conversion rates and indels.

Statistics and reproducibility

Error bars are derived from data that are calculated as mean \pm s.d. from biologically independent samples. The specific numbers of independent

experiments or biological samples (normally $n = 3$) are depicted in figure legends. P values were determined by a two-tailed Student's t -test in GraphPad Software (GraphPad Prism 9.3) and are indicated in the figures or figure legends. $P < 0.05$ was considered significant.

Reporting summary

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

Data availability

HTS data have been uploaded to the NCBI Sequence Read Archive (SRA) database under accession codes [PRJNA812697](#), [PRJNA812700](#) and [PRJNA862289](#). RNA-seq raw data have been uploaded into the SRA database under accession code [PRJNA811343](#). Data for rat embryos treated with ABE7.10 have already been posted in the SRA database under accession code [PRJNA471163](#). There are no restrictions on data availability. Source data are provided with this paper.

References

- Zhang, X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* **38**, 856–860 (2020).
- 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).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- Li, D. et al. Heritable gene targeting in the mouse and rat using a CRISPR–Cas system. *Nat. Biotechnol.* **31**, 681–683 (2013).

Acknowledgements

We are grateful to S. Siwko (Texas A&M University Health Science Center) for proofreading the manuscript and support from East China Normal University Public Platform for innovation (O11). We thank Y. Zhang from the Flow Cytometry Core Facility of School of Life Sciences in ECNU and Haixia Jiang from the Core Facility and Technical Service Center for SLSB of School of Life Sciences and Biotechnology in SJTU. We thank L. Ji (MedSci) for drawing schematic diagrams. This work was partially supported by grants from the National Key R&D Program of China (2019YFA0110802 to D.L. and 2019YFA0802800 to M.L.), the National Natural Science Foundation

of China (No.32025023 to D.L., No.32230064 to D.L. and No.31971366 to L.W.); grants from the Shanghai Municipal Commission for Science and Technology (20MC1920400 and 21CJ1402200 to D.L.), the Innovation Program of Shanghai Municipal Education Commission (2019-01-07-00-05-E00054 to D.L.), the Fundamental Research Funds for the Central Universities; and support from East China Normal University Outstanding Doctoral Students Academic Innovation Ability Improvement Project (YBNLTS2021-026 to L.C.).

Author contributions

L.C. and D.L. designed the project. L.C., S.Z., N.X., M.H., X.Z., J.Y., S.B., Y.H., C.L., B.Z. and L.W. performed ABE8e variant screening experiments. L.C., S.Z., N.X., M.H., J.Y., S.B. and Y.H. performed characterization of ABE9. L.C., S.Z. and N.X. assessed off-target effects of ABE9. L.C., M.H., S.Z. and M.L. performed animal studies. M.H., G.R. and H.G. generated stable cell lines and other assays. S.Z., N.X. and D.Z. performed target library analysis. L.C., S.Z., N.X., M.H., J.Y., S.B., Y.H., H.M., H.W., C.Y., M.L., L.Z., Y.C. and D.L. analyzed the sequencing data, D.L., L.C., Y.Z. and G.S. wrote the manuscript with input from all the authors. D.L. supervised the research.

Competing interests

L.C., D.L., S.Z., N.X. and M.L. have submitted patent applications on the basis the results reported in this study. The remaining authors declare no competing interests.

Additional information

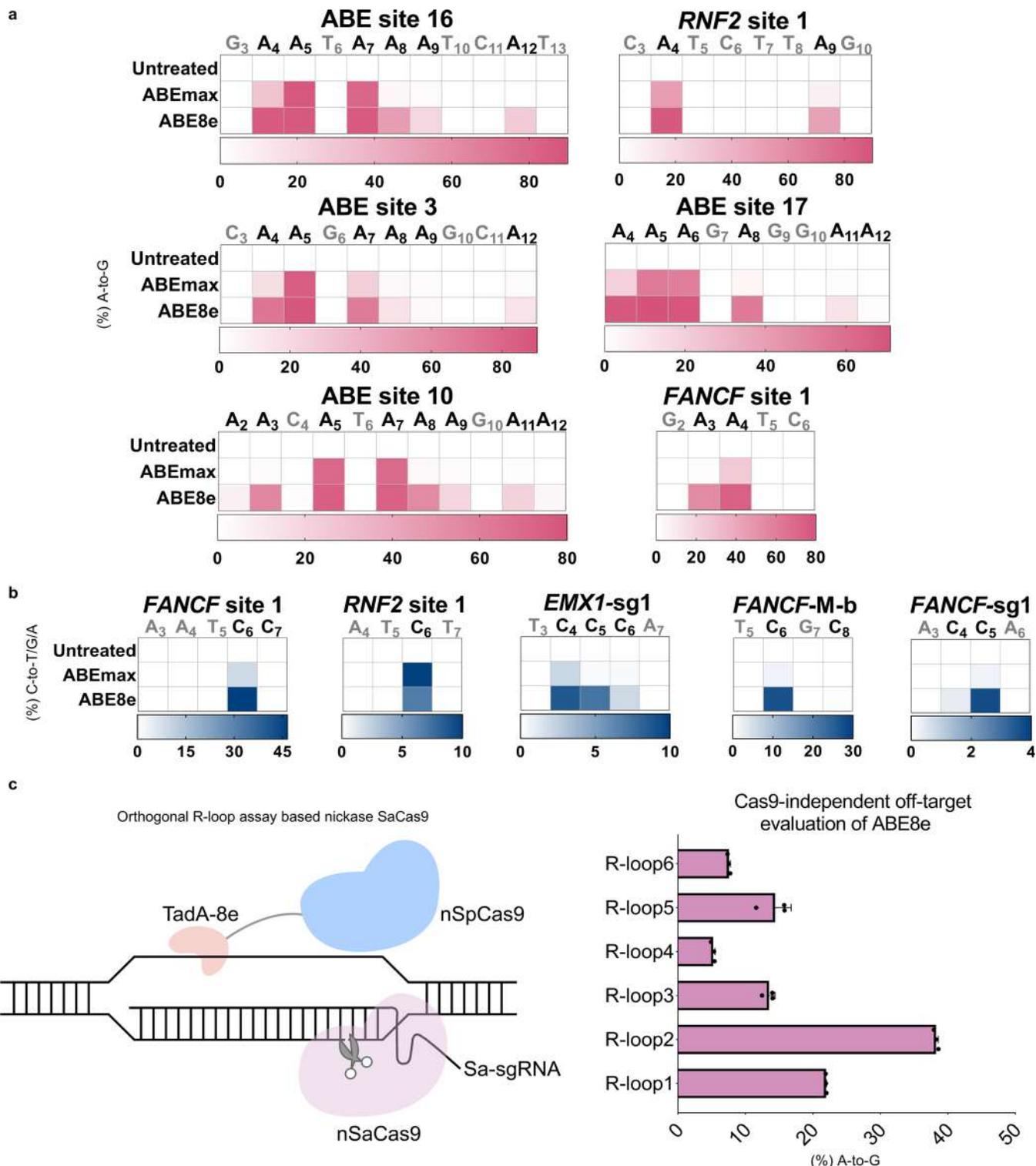
Extended data is available for this paper at <https://doi.org/10.1038/s41589-022-01163-8>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-022-01163-8>.

Correspondence and requests for materials should be addressed to Yongxiang Zhao, Gaojie Song or Dali Li.

Peer review information *Nature Chemical Biology* thanks Hui Yang, Wen Xue and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

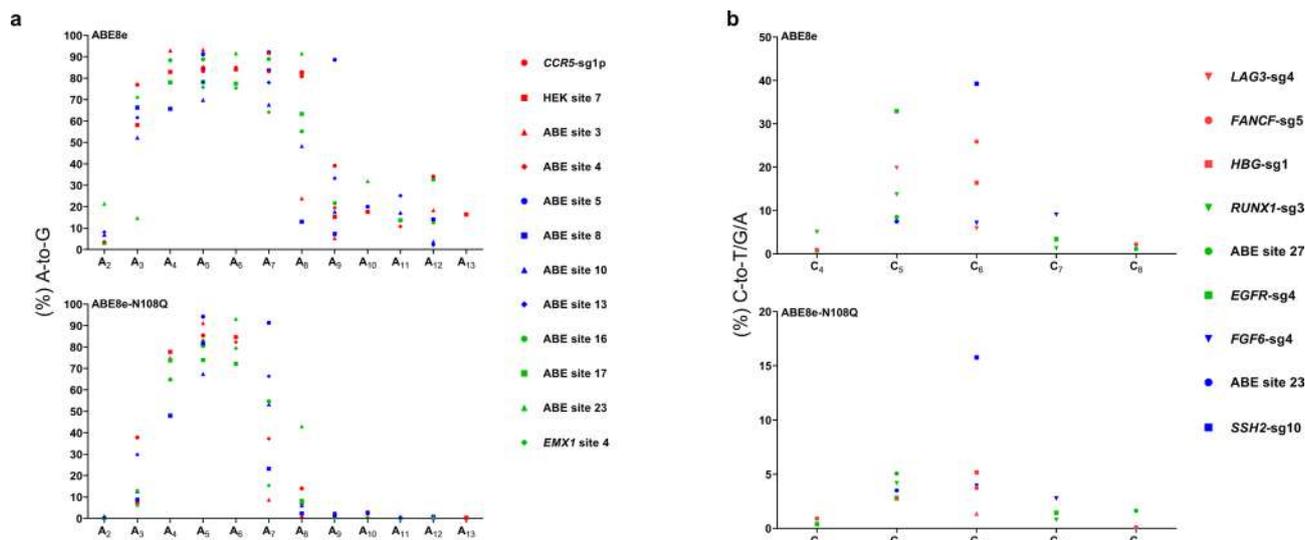
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | ABE8e induces severe bystander mutations and global random off-target editing.

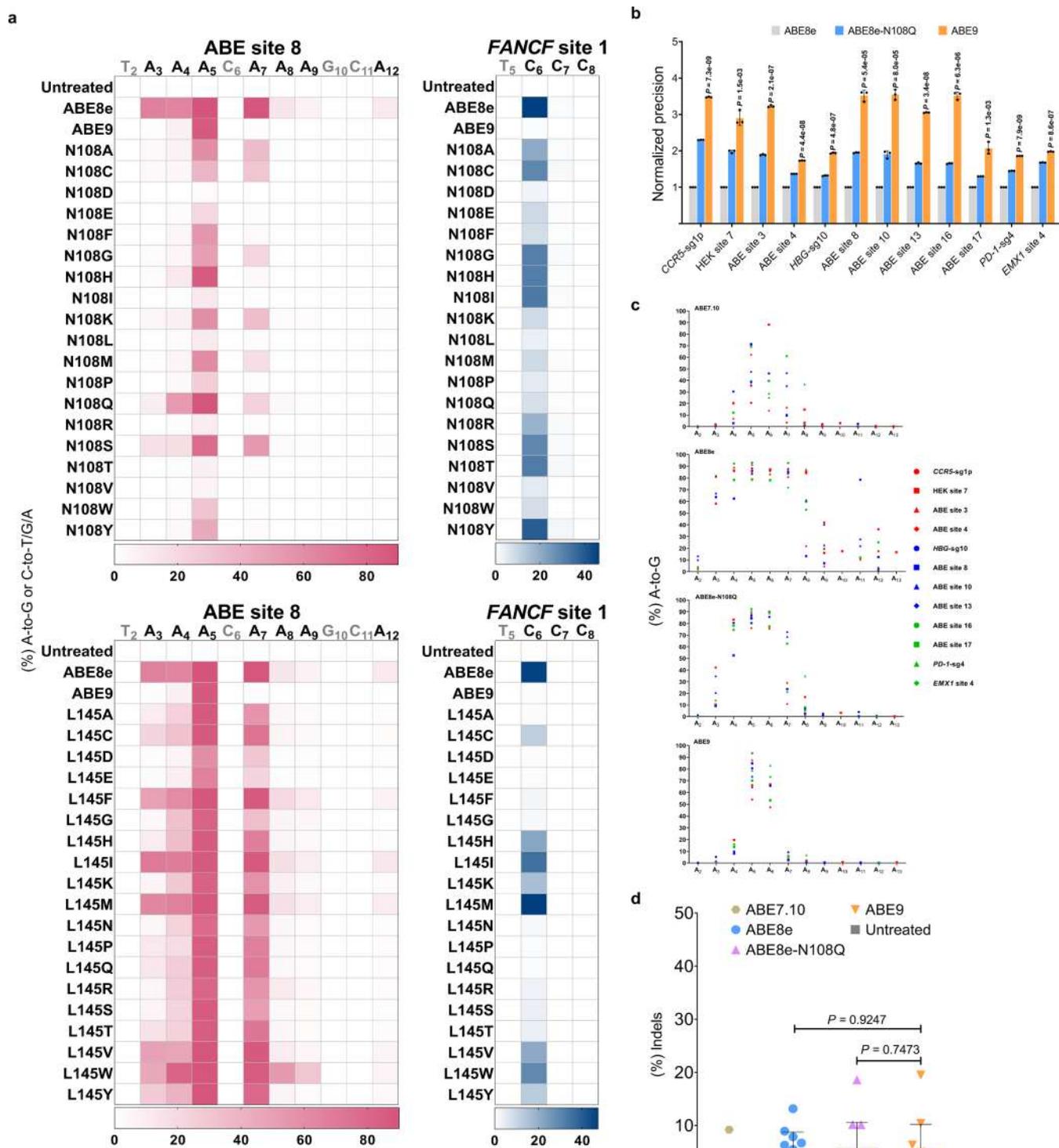
a, Comparison of base editing efficiency of ABEmax or ABE8e at 6 endogenous genomic loci in HEK293T cells. Partial data are derived from Fig. 1b. **b**, The C-to-T/G/A editing efficiency of ABEmax or ABE8e was examined at 5 endogenous genomic loci in HEK293T cells. Partial data are derived from Fig. 1c. **c**, The schematic diagram of orthogonal R-loop assay-based nickase SaCas9 (nSaCas9) (left panel); Cas9-independent DNA

off-target analysis of ABE8e using the modified orthogonal R-loop assay at each R-loop site with nSaCas9-sgRNA plasmid (right panel). Data are mean \pm s.d. ($n = 3$ independent experiments). Data are derived from Fig. 4b. In **a** and **b**, the heatmap represents average editing percentage derived from three independent experiments and editing efficiency was determined by HTS. Statistical source data are provided online.



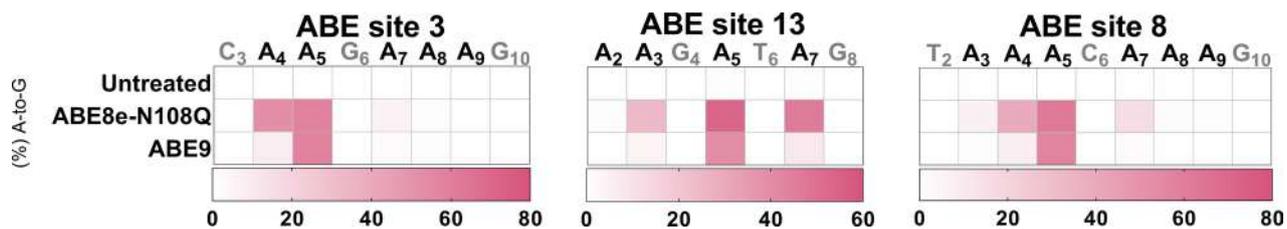
Extended Data Fig. 2 | Comparison of the editing window between ABE8e and ABE8e-N108Q. a, Comparison of A-to-G base editing window of ABE8e or ABE8e-N108Q at 12 target sites in HEK293T cells. **b**, Comparison of C-to-T/G/A base

editing window of ABE8e or ABE8e-N108Q at 9 target sites in HEK293T cells. In **a** and **b**, data are from Fig. 2a (**a**) and Fig. 2b (**b**), and each point represents mean from three independent experiments. Statistical source data are provided online.



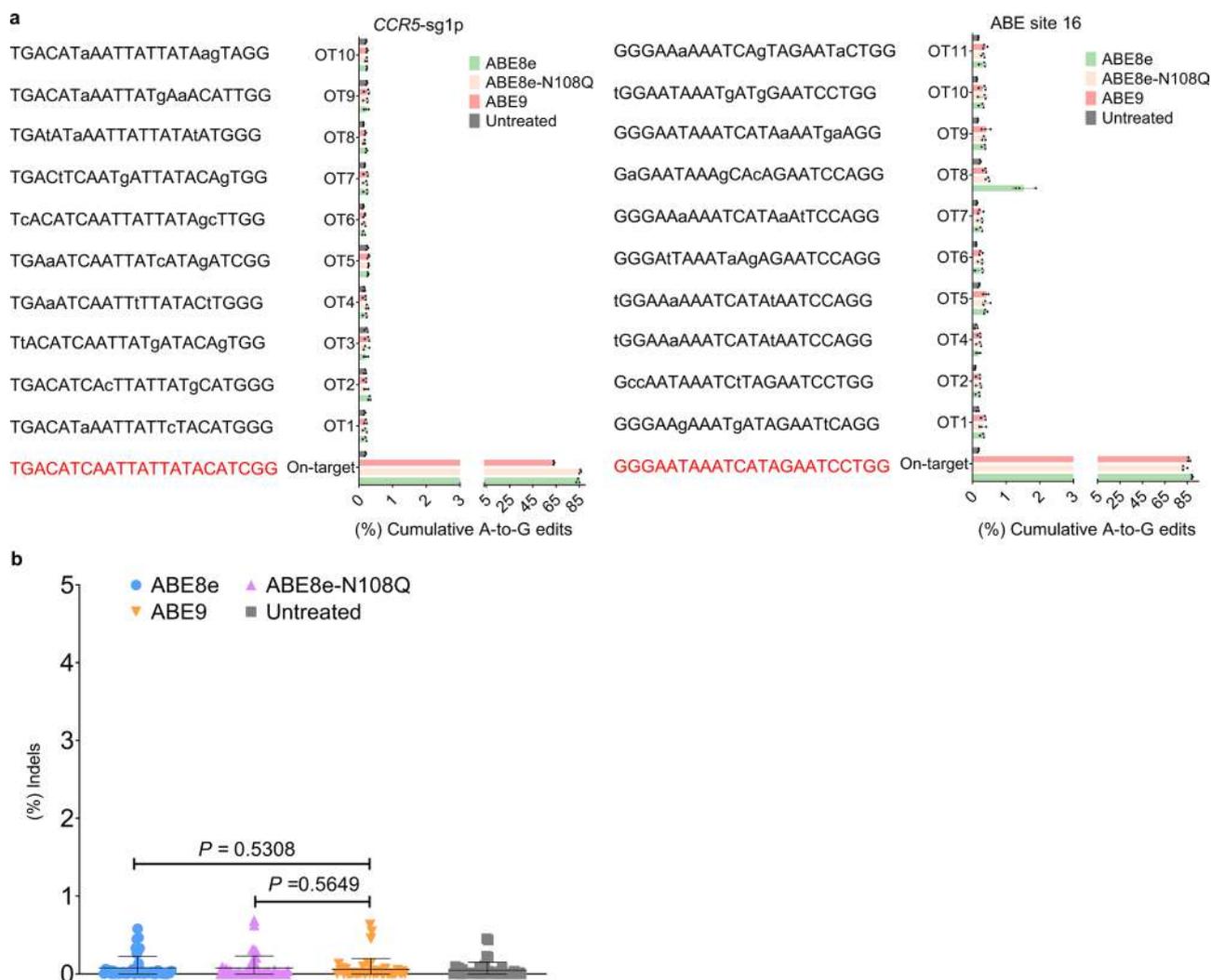
Extended Data Fig. 3 | Evaluation of ABE8e, ABE9 and individual saturation variants. **a**, Base editing efficiency of N108- or L145-saturated variants at 2 endogenous genomic loci in HEK293T cells. The heatmap represents an average editing percentage derived from two or three independent experiments with editing efficiency determined by HTS. **b**, The normalized precision (ABE8e is used for standardization) is defined as the highest / all other A-to-G base editing of ABE8e-N108Q or ABE9 at the 12 target sites in Fig. 3b. Data represent mean \pm s.d. from three independent experiments. *P* values above each group indicated the comparison of ABE8e-N108Q and ABE9. **c**, Comparison of A-to-G

base editing window of ABE7.10, ABE8e, ABE8e-N108Q or ABE9 at 12 target sites from Fig. 3b. In HEK293T cells. Data points represent mean from three independent experiments. **d**, Comparison of indels induced by ABE7.10, ABE8e, ABE8e-N108Q or ABE9 at 12 target sites from Fig. 3b. Each data point represents the average indel frequency at each target site calculated from 3 independent experiments. Error bar and *P* value are derived from these 12 data points. Data are mean \pm s.d. In **b** and **d**, *P* value was determined by two-tailed Student's *t* test. Statistical source data are provided online.



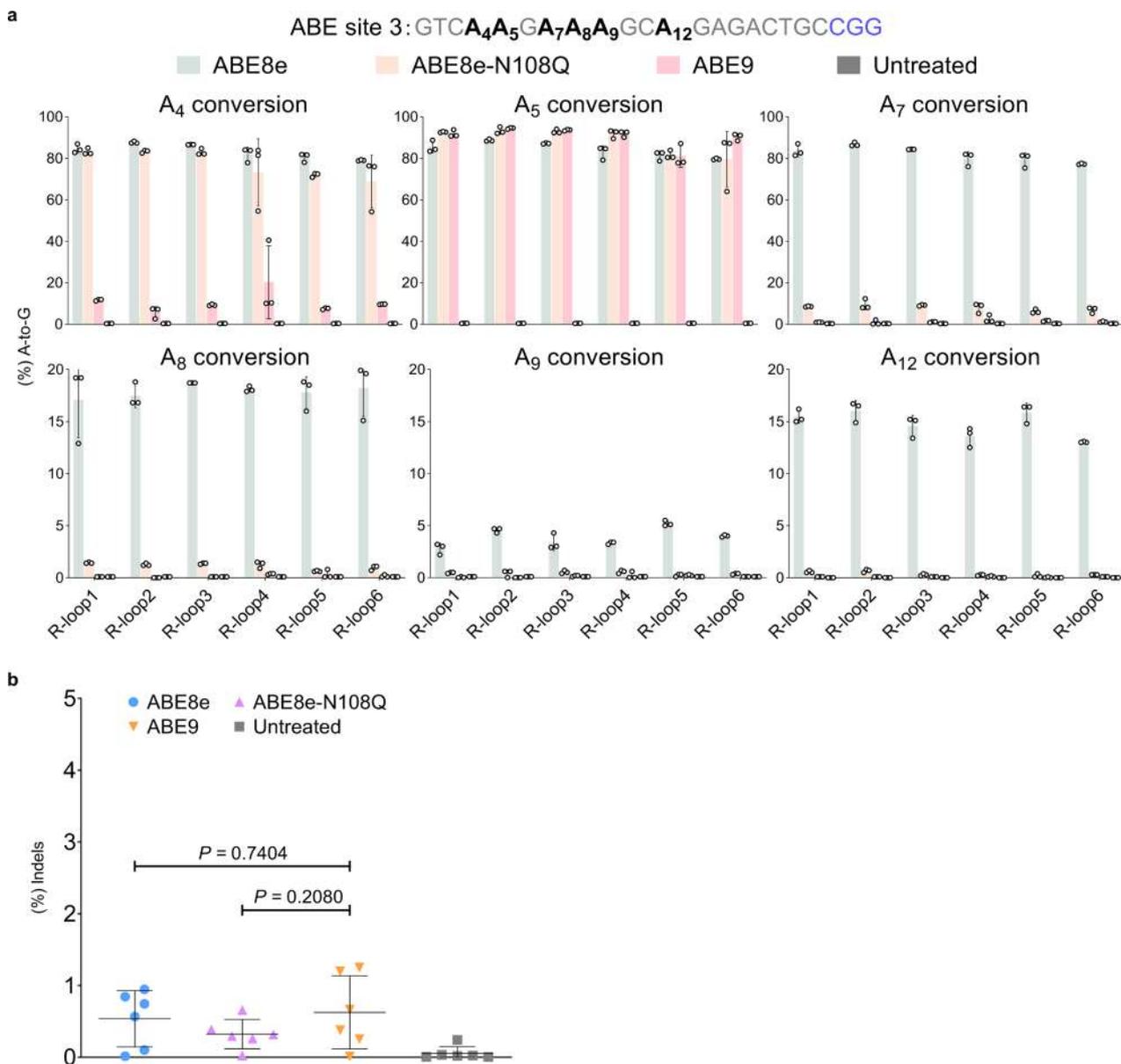
Extended Data Fig. 4 | Editing activities comparison of ABE8e-N108Q and ABE9 in HeLa cells. The A-to-G editing efficiency of ABE8e-N108Q or ABE9 was examined at 3 endogenous genomic loci containing multiple As. The heatmap

represents an average editing percentage derived from three independent experiments with editing efficiency determined by HTS. Statistical source data are provided online.



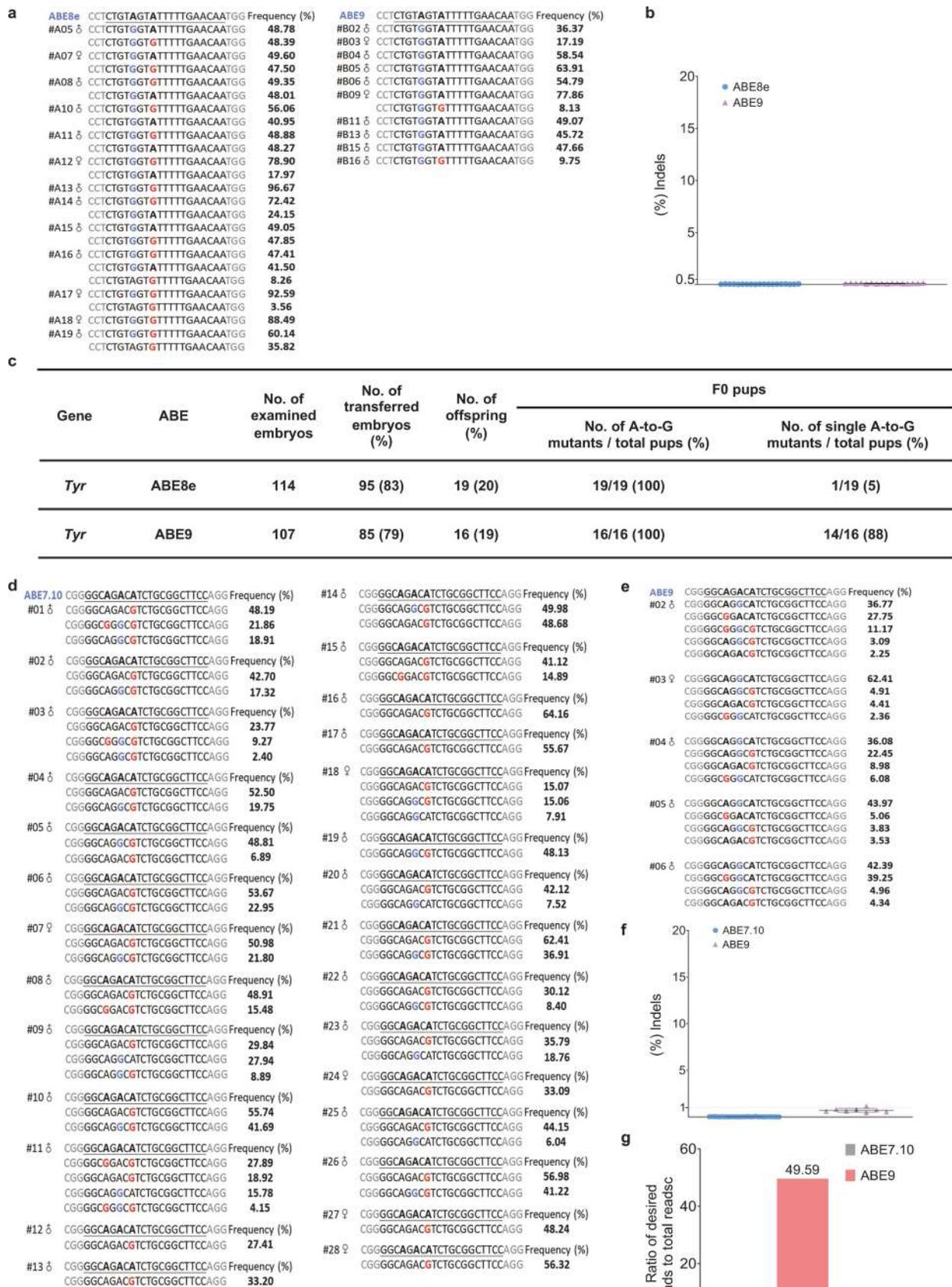
Extended Data Fig. 5 | Cas9-dependent off-target assessment of ABE9. a, Cas9-dependent DNA on- and off-target analysis of the indicated targets (*CCR5*-sg1p and ABE site 16) by ABE8e, ABE8e-N108Q and ABE9 in HEK293T cells. Data are mean \pm s.d. ($n = 3$ independent experiments). On-target data are derived from Fig. 3b. **b,** Comparison of indels induced by ABE8e, ABE8e-N108Q or ABE9 at 44 Cas9-dependent DNA off-target target sites from **a** and Fig. 4a. Each data

point represents the average indel frequency at each target site calculated from 2 independent experiments merely in HEK site 2-GUIDE-seq-OT1, 2 treated with ABE8e-N108Q and from 3 independent experiments in the rest of biological samples. Error bar and P value are derived from these 44 data points. Data are mean \pm s.d. P value was determined by two-tailed Student's t test. Statistical source data are provided online.



Extended Data Fig. 6 | Cas9-independent off-target assessment in the R-loop assay. a. On-target base editing induced by ABE8e, ABE8e-N108Q or ABE9 using the modified orthogonal R-loop assay at each R-loop site with nSaCas9-sgRNA plasmid. Data are mean \pm s.d. (n = 3 independent experiments). **b.** Comparison of indels induced by ABE8e, ABE8e-N108Q or ABE9 at six R-loop sites from Fig.

4b. Each data point represents the average indel frequency at each target site calculated from 3 independent experiments. Error bar and *P* value are derived from these 6 data points. Data are mean \pm s.d. *P* value was determined by two-tailed Student's *t* test. Statistical source data are provided online.



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

Extended Data Fig. 7 | Highly efficient and precise editing by ABE9 in rodent embryos. **a**, Genotyping of F0 generation pups treated by ABE8e or ABE9. **b**, Comparison of indels induced ABE8e (n = 19) or ABE9 (n = 16) in the target sequence of the splicing acceptor site in intron 3 of the mouse *Tyr* gene. Data are mean \pm s.d. **c**, Summary of the numbers of embryos used and the pups generated after microinjection of ABE8e/sgRNA or ABE9/sgRNA. **d, e**, Genotyping of F0 rats induced by ABE7.10 (**d**) and ABE9 (**e**) (desired editing in blue or undesired in

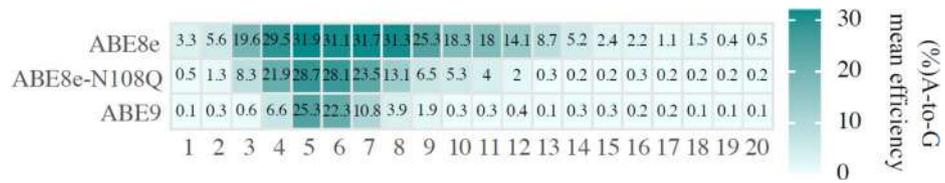
red). **f**, Comparison of indels induced by ABE7.10 (n = 28) or ABE9 (n = 8) in the target sequence of exon 13 in the rat *Gaa* gene. Data are mean \pm s.d. **g**, Ratio of desired reads to total reads in F0 rats induced by ABE7.10 or ABE9. In **a, d** and **e**, the percentage on the right represents the frequency determined by the rate of indicated mutant alleles to total alleles counts. Percentiles of each allele reads <1% are omitted. Statistical source data are provided online.

a COL1A2 c.1136G>A				b CARD14 c.424G>A			
ABE8e-Rep.1		ABE8e-Rep.2		ABE8e-Rep.1		ABE8e-Rep.2	
GAAGAAAAGAGGCCCTAA	Frequency (%)	GAAGAAAAGAGGCCCTAA	Frequency (%)	CCAGAAAAGGGGCGAAGG	Frequency (%)	CCAGAAAAGGGGCGAAGG	Frequency (%)
GGGGGGGAGAGGCCCTAA	21.09	GGGGGGGAGAGGCCCTAA	19.28	CCGGGGGAGGGCGAAGG	44.96	CCGGGGGAGGGCGAAGG	40.22
GGGGGGGAGAGGCCCTAA	16.36	GGGGGGGAGAGGCCCTAA	18.13	CCGGGGGAGGGCGAAGG	27.61	CCGGGGGAGGGCGAAGG	27.06
GGGGGGGAGAGGCCCTAA	12.44	GGGGGGGAGAGGCCCTAA	10.84	CCGGGGGAGGGCGAAGG	5.10	CCGGGGGAGGGCGAAGG	5.22
GGGGGGGAGAGGCCCTAA	8.88	GGGGGGGAGAGGCCCTAA	7.90	CCGGGGGAGGGCGAAGG	3.23	CCGGGGGAGGGCGAAGG	3.15
GGGGGGGAGAGGCCCTAA	5.90	GGGGGGGAGAGGCCCTAA	6.29	CCGGGGGAGGGCGAAGG	1.41	CCGGGGGAGGGCGAAGG	1.67
GGGGGGGAGAGGCCCTAA	5.43	GGGGGGGAGAGGCCCTAA	4.98	CCAGAAAAGGGGCGAAGG	0.02	CCAGAAAAGGGGCGAAGG	0.57
GGGGGGGAGAGGCCCTAA	3.96	GGGGGGGAGAGGCCCTAA	3.31	CCAGAAAAGGGGCGAAGG	7.02	CCAGAAAAGGGGCGAAGG	7.18
GGGGGGGAGAGGCCCTAA	2.38	GGGGGGGAGAGGCCCTAA	3.06	CCGGGGGAGGGCGAAGG	27.04	CCGGGGGAGGGCGAAGG	28.33
GGGGGGGAGAGGCCCTAA	1.81	GGGGGGGAGAGGCCCTAA	2.71	CCGGGGGAGGGCGAAGG	12.76	CCGGGGGAGGGCGAAGG	12.06
GGGGGGGAGAGGCCCTAA	1.69	GGGGGGGAGAGGCCCTAA	2.47	CCAGAAAAGGGGCGAAGG	10.28	CCAGAAAAGGGGCGAAGG	10.69
GGGGGGGAGAGGCCCTAA	0.11	GGGGGGGAGAGGCCCTAA	0.08	CCAGAAAAGGGGCGAAGG	7.34	CCAGAAAAGGGGCGAAGG	7.18
ABE8e-N108Q-Rep.1	Frequency (%)	ABE8e-N108Q-Rep.2	Frequency (%)	ABE8e-N108Q-Rep.1	Frequency (%)	ABE8e-N108Q-Rep.2	Frequency (%)
GAAGAAAAGAGGCCCTAA	25.84	GAAGAAAAGAGGCCCTAA	25.93	CCAGAAAAGGGGCGAAGG	23.05	CCAGAAAAGGGGCGAAGG	25.18
GGGGGGGAGAGGCCCTAA	21.60	GGGGGGGAGAGGCCCTAA	19.42	CCAGAAAAGGGGCGAAGG	16.02	CCAGAAAAGGGGCGAAGG	15.10
GGGGGGGAGAGGCCCTAA	11.02	GGGGGGGAGAGGCCCTAA	10.47	CCAGAAAAGGGGCGAAGG	5.75	CCAGAAAAGGGGCGAAGG	7.79
GGGGGGGAGAGGCCCTAA	9.48	GGGGGGGAGAGGCCCTAA	10.24	CCAGAAAAGGGGCGAAGG	4.90	CCAGAAAAGGGGCGAAGG	4.74
GGGGGGGAGAGGCCCTAA	4.41	GGGGGGGAGAGGCCCTAA	5.13	CCAGAAAAGGGGCGAAGG	2.54	CCAGAAAAGGGGCGAAGG	3.05
GGGGGGGAGAGGCCCTAA	2.95	GGGGGGGAGAGGCCCTAA	2.90	CCAGAAAAGGGGCGAAGG	2.00	CCAGAAAAGGGGCGAAGG	2.19
GGGGGGGAGAGGCCCTAA	2.54	GGGGGGGAGAGGCCCTAA	2.67	CCAGAAAAGGGGCGAAGG	0.22	CCAGAAAAGGGGCGAAGG	0.27
GGGGGGGAGAGGCCCTAA	2.29	GGGGGGGAGAGGCCCTAA	2.32	CCAGAAAAGGGGCGAAGG	23.05	CCAGAAAAGGGGCGAAGG	25.18
GGGGGGGAGAGGCCCTAA	1.92	GGGGGGGAGAGGCCCTAA	2.22	CCAGAAAAGGGGCGAAGG	15.02	CCAGAAAAGGGGCGAAGG	15.10
GGGGGGGAGAGGCCCTAA	1.41	GGGGGGGAGAGGCCCTAA	1.74	CCAGAAAAGGGGCGAAGG	5.75	CCAGAAAAGGGGCGAAGG	7.79
ABE9-Rep.1	Frequency (%)	ABE9-Rep.2	Frequency (%)	ABE9-Rep.1	Frequency (%)	ABE9-Rep.2	Frequency (%)
GAAGAAAAGAGGCCCTAA	33.89	GAAGAAAAGAGGCCCTAA	33.87	CCAGAAAAGGGGCGAAGG	3.93	CCAGAAAAGGGGCGAAGG	4.74
GAAGAAAAGAGGCCCTAA	13.91	GAAGAAAAGAGGCCCTAA	13.57	CCAGAAAAGGGGCGAAGG	4.29	CCAGAAAAGGGGCGAAGG	4.65
GAAGAAAAGAGGCCCTAA	5.24	GAAGAAAAGAGGCCCTAA	5.42	CCAGAAAAGGGGCGAAGG	3.93	CCAGAAAAGGGGCGAAGG	3.08
GAAGAAAAGAGGCCCTAA	4.61	GAAGAAAAGAGGCCCTAA	4.86	CCAGAAAAGGGGCGAAGG	3.69	CCAGAAAAGGGGCGAAGG	3.08
GAAGAAAAGAGGCCCTAA	3.83	GAAGAAAAGAGGCCCTAA	3.75	CCAGAAAAGGGGCGAAGG	2.65	CCAGAAAAGGGGCGAAGG	2.75
GAAGAAAAGAGGCCCTAA	1.71	GAAGAAAAGAGGCCCTAA	1.44	CCAGAAAAGGGGCGAAGG	2.25	CCAGAAAAGGGGCGAAGG	2.17
GAAGAAAAGAGGCCCTAA	1.17	GAAGAAAAGAGGCCCTAA	1.42	CCAGAAAAGGGGCGAAGG	1.47	CCAGAAAAGGGGCGAAGG	1.31

c BVES c.602C>T				d KCNA5 c.1828G>A			
ABE8e-Rep.1		ABE8e-Rep.2		ABE8e-Rep.1		ABE8e-Rep.2	
AGGAAAATCTATAAGGCAC	Frequency (%)	AGGAAAATCTATAAGGCAC	Frequency (%)	CCGAAAACAGATTTGTGAA	Frequency (%)	CCGAAAACAGATTTGTGAA	Frequency (%)
AGGGGGCTCTATAAGGCAC	30.71	AGGGGGCTCTATAAGGCAC	28.87	CCGGGGGACAGATTTGTGAA	47.26	CCGGGGGACAGATTTGTGAA	47.79
AGGGGGCTCTATAAGGCAC	15.39	AGGGGGCTCTATAAGGCAC	14.62	CCGGGGGACAGATTTGTGAA	20.49	CCGGGGGACAGATTTGTGAA	20.11
AGGGGGCTCTATAAGGCAC	12.21	AGGGGGCTCTATAAGGCAC	11.62	CCGGGGGACAGATTTGTGAA	11.77	CCGGGGGACAGATTTGTGAA	11.90
AGGGGGCTCTATAAGGCAC	4.79	AGGGGGCTCTATAAGGCAC	4.56	CCGGGGGACAGATTTGTGAA	6.81	CCGGGGGACAGATTTGTGAA	6.52
AGGGGGCTCTATAAGGCAC	3.96	AGGGGGCTCTATAAGGCAC	4.27	CCGGGGGACAGATTTGTGAA	2.18	CCGGGGGACAGATTTGTGAA	2.23
AGGGGGCTCTATAAGGCAC	2.90	AGGGGGCTCTATAAGGCAC	3.21	CCGGGGGACAGATTTGTGAA	2.03	CCGGGGGACAGATTTGTGAA	2.18
AGGGGGCTCTATAAGGCAC	2.25	AGGGGGCTCTATAAGGCAC	2.45	CCGGGGGACAGATTTGTGAA	0.51	CCGGGGGACAGATTTGTGAA	0.48
AGGGGGCTCTATAAGGCAC	1.86	AGGGGGCTCTATAAGGCAC	1.95	CCGAAAACAGATTTGTGAA	28.86	CCGAAAACAGATTTGTGAA	28.37
AGGGGGCTCTATAAGGCAC	1.50	AGGGGGCTCTATAAGGCAC	1.66	CCGAAAACAGATTTGTGAA	25.96	CCGAAAACAGATTTGTGAA	26.04
AGGGGGCTCTATAAGGCAC	1.44	AGGGGGCTCTATAAGGCAC	1.49	CCGAAAACAGATTTGTGAA	11.88	CCGAAAACAGATTTGTGAA	11.59
ABE8e-N108Q-Rep.1	Frequency (%)	ABE8e-N108Q-Rep.2	Frequency (%)	ABE8e-N108Q-Rep.1	Frequency (%)	ABE8e-N108Q-Rep.2	Frequency (%)
AGGAAAATCTATAAGGCAC	40.04	AGGAAAATCTATAAGGCAC	39.05	CCGGGGGACAGATTTGTGAA	8.67	CCGGGGGACAGATTTGTGAA	8.87
AGGAAAATCTATAAGGCAC	14.33	AGGAAAATCTATAAGGCAC	13.77	CCGGGGGACAGATTTGTGAA	4.52	CCGGGGGACAGATTTGTGAA	4.73
AGGAAAATCTATAAGGCAC	12.01	AGGAAAATCTATAAGGCAC	11.88	CCGGGGGACAGATTTGTGAA	4.30	CCGGGGGACAGATTTGTGAA	4.44
AGGAAAATCTATAAGGCAC	7.46	AGGAAAATCTATAAGGCAC	8.07	CCGGGGGACAGATTTGTGAA	2.74	CCGGGGGACAGATTTGTGAA	2.70
AGGAAAATCTATAAGGCAC	4.20	AGGAAAATCTATAAGGCAC	3.92	CCGGGGGACAGATTTGTGAA	2.08	CCGGGGGACAGATTTGTGAA	2.11
AGGAAAATCTATAAGGCAC	2.64	AGGAAAATCTATAAGGCAC	2.45	CCGGGGGACAGATTTGTGAA	1.84	CCGGGGGACAGATTTGTGAA	1.86
AGGAAAATCTATAAGGCAC	2.05	AGGAAAATCTATAAGGCAC	1.97	CCGGGGGACAGATTTGTGAA	1.44	CCGGGGGACAGATTTGTGAA	1.56
AGGAAAATCTATAAGGCAC	1.38	AGGAAAATCTATAAGGCAC	1.45	CCGAAAACAGATTTGTGAA	38.30	CCGAAAACAGATTTGTGAA	38.72
ABE9-Rep.1	Frequency (%)	ABE9-Rep.2	Frequency (%)	ABE9-Rep.1	Frequency (%)	ABE9-Rep.2	Frequency (%)
AGGAAAATCTATAAGGCAC	34.32	AGGAAAATCTATAAGGCAC	32.60	CCGGGGGACAGATTTGTGAA	10.23	CCGGGGGACAGATTTGTGAA	10.63
AGGAAAATCTATAAGGCAC	8.54	AGGAAAATCTATAAGGCAC	8.77	CCGGGGGACAGATTTGTGAA	9.79	CCGGGGGACAGATTTGTGAA	9.83
AGGAAAATCTATAAGGCAC	8.27	AGGAAAATCTATAAGGCAC	8.08	CCGGGGGACAGATTTGTGAA	4.42	CCGGGGGACAGATTTGTGAA	3.96
AGGAAAATCTATAAGGCAC	7.56	AGGAAAATCTATAAGGCAC	7.72	CCGGGGGACAGATTTGTGAA	3.15	CCGGGGGACAGATTTGTGAA	2.79
AGGAAAATCTATAAGGCAC	7.23	AGGAAAATCTATAAGGCAC	7.60	CCGGGGGACAGATTTGTGAA	1.43	CCGGGGGACAGATTTGTGAA	1.56
AGGAAAATCTATAAGGCAC	3.40	AGGAAAATCTATAAGGCAC	3.77	CCGAAAACAGATTTGTGAA	30.32	CCGAAAACAGATTTGTGAA	30.63
AGGAAAATCTATAAGGCAC	1.45	AGGAAAATCTATAAGGCAC	1.93	CCGAAAACAGATTTGTGAA	10.63	CCGAAAACAGATTTGTGAA	10.65
AGGAAAATCTATAAGGCAC	1.34	AGGAAAATCTATAAGGCAC	0.93	CCGAAAACAGATTTGTGAA	9.83	CCGAAAACAGATTTGTGAA	10.43
				CCGAAAACAGATTTGTGAA	5.37	CCGAAAACAGATTTGTGAA	5.37
				CCGAAAACAGATTTGTGAA	4.32	CCGAAAACAGATTTGTGAA	4.32
				CCGAAAACAGATTTGTGAA	2.18	CCGAAAACAGATTTGTGAA	2.18

Extended Data Fig. 8 | Allele tables for ABE9 in four stable HEK293T cell lines. a-d, Allele tables for ABE8e, ABE8e-N108Q and ABE9 in four stable HEK293T cell lines: COL1A2 c.1136 G > A (a), CARD14 c.424 G > A (b), BVES c.602 C > T (c) and KCNA5 c.1828G > A (d). The percentile and sequencing reads of each allele from

two or three independent experiments are listed on the right. Desired A₂-to-G percentiles of alleles are exhibited, while percentiles of top ten invalid allele types are presented and percentiles of invalid allele types less than 1% are omitted.



Extended Data Fig. 9 | Unbiased analysis of target library of ABE9. Analysis of absolute mean editing efficiency of ABE8e, ABE8e-N108Q and ABE9. Positions of the protospacer are shown at the bottom of the heatmap, counting the protospacer adjacent motif (PAM) as positions 21–23.

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 our [Editorial Policies](#) 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

Targeted amplicon sequencing data were collected and demultiplexed by an Illumina HiSeq X Ten instrument.
RNA-seq data were collected and demultiplexed by an Illumina NovaSeq 6000 instrument.
Library amplicon data were collected and demultiplexed by an Illumina NovaSeq 6000 instrument.
FACS gating data were collected on a FACSAria III (BD Biosciences) using FACSDiva version 8.0.2 (BD Biosciences).

Data analysis

High-throughput sequencing data were analyzed by BE-Analyzer (<http://www.rgenome.net/be-analyzer/#!>) (Hwang G-H et al, BMC Bioinformatics, 2018) for base editing (A>G, C>T, C>G and C>A) and indel efficiencies.
Potential DNA off-target sites for ABE8e, ABE8e-N108Q and ABE9 were predicated using Cas-OFFinder web (<http://www.rgenome.net/cas-offinder/>).
RNA-seq data were analyzed using Trim Galore (version 0.6.6), STAR (version 2.7.1a), SAMtools (version 1.14), Picard MarkDuplicates module (version 2.23.9) software.
Library amplicons data were analyzed using BWA (version 0.7.17-r1188), EMBOSS needle software (version 6.6.0.0).
GraphPad Prism 9.3 was also used to analyze data.
The JavaScript version of fastq-join (<https://github.com/brwnj/fastq-join>) joined two fastq files from high-throughput sequencing.

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 and 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 outcome of HTS has been uploaded into the NCBI Sequence Read Archive (SRA) database under accession codes PRJNA812697, PRJNA812700 and PRJNA862289. RNA-seq raw data have been uploaded into the SRA database under accession code PRJNA811343 and hg38 genome (<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/>) was used to be aligned in the process of analysis. Data for rat embryos treated with ABE7.10 have already been posted in the SRA database under accession code PRJNA471163 from the previous study. Source data and supplementary information for main figures and extended data figures are all electronically available. There are no restrictions on data availability.

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. Experiments were performed in biological triplicate n=3 unless otherwise noted. Sample sizes were opted to display the range and consistency of differences and three biological replicates made it sufficient to support the conclusions in this research.
Data exclusions	No data were excluded from the analyses.
Replication	Three independent biological replicates were performed on different days. All replications were successful.
Randomization	Samples were randomly distributed into groups.
Blinding	Investigators were not blinded to group allocation in this research since experimental conditions were evident and all samples of treatment were consistent throughout experiments.

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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293T cells (source: ATCC CRL-3216) and HeLa cells (source: ATCC CCL-2).
Authentication	Cell lines in this research were not undergone authenticated procedures.

Mycoplasma contamination	All cell lines used were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	6-10 weeks old female C57BL/6J, ICR strain mice and Sprague-Dawley strain rats purchased from Shanghai Laboratory Animal Center were housed in standard cages at 20-22°C with 40-60% humidity in a specific pathogen-free facility on a 12 h light/dark cycle. Water and food were offered ad libitum.
Wild animals	No studies with wild animals were performed.
Field-collected samples	No studies with field-collected samples were performed.
Ethics oversight	All animal experiments conformed to the regulations drafted by the Association for Assessment and Accreditation of Laboratory Animal Care in Shanghai and were approved by the East China Normal University Center for Animal Research.

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	Cell culture and transfection procedures are described in the methods. HEK293T cells were washed and filtered through a 45µm cell strainer cap before sorting (72 h after transfection).
Instrument	FACSAria III (BD Biosciences)
Software	FACSDiva version 8.0.2 (BD Biosciences)
Cell population abundance	HEK293T cell numbers gated for target populations were similar in different biology replicates.
Gating strategy	For HEK293T cells, gates were established using uninfected control cells and GFP positive control. Gates were drawn to collect subsets of GFP-expressing cells. For specified transcriptome profiling, cells with top 15% of GFP signals were collected. Detailed gating strategy is provided in the Supplementary Figure 1.

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