



Eliminating base-editor-induced genome-wide and transcriptome-wide off-target mutations

Lijie Wang^{1,2,3,4,10}, Wei Xue^{5,10}, Hongxia Zhang^{6,7,10}, Runze Gao^{1,2,3,4,10}, Houyuan Qiu^{6,8,10}, Jia Wei⁵, Lina Zhou^{1,2,3,4}, Yun-Ni Lei^{1,5}, Xiaocheng Wu⁸, Xiao Li⁹, Chengfang Liu^{1,2,3,4}, Jing Wu^{1,2}, Qiubing Chen^{6,7}, Hanhui Ma^{1,2}, Xingxu Huang^{1,2,3}, Cheguo Cai⁹, Ying Zhang^{1,2}, Bei Yang^{1,2}, Hao Yin^{1,2}, Li Yang^{1,2} and Jia Chen^{1,2,3}

The fusion of CRISPR-Cas9 with cytidine deaminases leads to base editors (BEs) capable of programmable C-to-T editing, which has potential in clinical applications but suffers from off-target (OT) mutations. Here, we used a cleavable deoxycytidine deaminase inhibitor (dCDI) domain to construct a transformer BE (tBE) system that induces efficient editing with only background levels of genome-wide and transcriptome-wide OT mutations. After being produced, the tBE remains inactive at OT sites with the fusion of a cleavable dCDI, therefore eliminating unintended mutations. When binding at on-target sites, the tBE is transformed to cleave off the dCDI domain and catalyses targeted deamination for precise base editing. After delivery into mice through a dual-adeno-associated virus (AAV) system, the tBE system created a premature stop codon in *Pcsk9* and significantly reduced serum PCSK9, resulting in a ~30–40% decrease in total cholesterol. The development of tBE establishes a highly specific base editing system and its in vivo efficacy has potential for therapeutic applications.

Cytosine or adenine BEs (CBEs or ABEs) that fuse the apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) family of cytidine deaminases (CDAs) or in vitro evolved adenosine deaminases with CRISPR-Cas have been developed to induce targeted C-to-T or A-to-G conversions^{1–6}. However, whole-genome sequencing studies showed that the originally reported BE3 induced genome-wide off-target (OT) mutations in vivo^{7,8}. Furthermore, BE3 could induce OT mutations in transcriptomic RNA^{9,10}, as the rat APOBEC1 (rA1) moiety of BE3 can deaminate both deoxycytidine in DNA and cytidine in RNA^{11,12}. These genome-wide and transcriptome-wide OT mutations therefore hamper the applications of originally reported BEs, especially for therapies. As endogenous APOBECs can induce unexpected C-to-T mutations in single-stranded genomic DNA regions^{13–16}, the observed genome-wide and transcriptome-wide OT mutations by BE3 can be caused by the APOBEC moiety of BEs, but independent of the sgRNA or the Cas9 moiety of BEs^{6,17}.

Here, we developed a convenient method to quantitatively evaluate mutations at single-stranded OT (OTss) sites and confirmed that the APOBEC moiety of BEs could induce mutations in an sgRNA-independent manner. We further determined the role of each CDA domain of the APOBECs containing dual CDA domains and found that some inactive CDA domains exhibited dCDI activities. By linking a cleavable dCDI domain to the APOBEC moiety of BE, we set up a transformer BE (tBE) system that remains inactive

after being produced to avoid mutations at both sgRNA-independent OTss sites and the OT sites that have sequence similarity to on-target sites (sgRNA-dependent OT sites). Only after binding at on-target sites, the dCDI domain of tBE is removed and therefore activated for targeted deamination and editing. Whole-genome-wide and transcriptome-wide analyses revealed that tBE induces no observable OT mutation. Importantly, through a dual-AAV system, tBEs were delivered into mice to inhibit the expression of proprotein convertase subtilisin/kexin type 9 (*Pcsk9*), which resulted in significantly reduced PCSK9 protein and total cholesterol in the serum, indicating the potential of tBEs for therapies.

Results

BEs induced sgRNA-independent mutations at OTss sites. To evaluate BE-induced mutations at sgRNA-independent OTss sites, we set up a quantitative and convenient detection method by coexpressing *Staphylococcus aureus* and *Streptococcus pyogenes* Cas9 orthologues (CESSCO). In CESSCO, the expression of an orthogonal *S. aureus* nCas9 (nSaCas9)-Sa-sgRNA pair¹⁸ forms an R-loop that contains a single-stranded DNA (ssDNA) region^{19,20} (Extended Data Fig. 1a). Meanwhile, the expression of BE3 or hA3A-BE3 (refs. ^{1,21}) containing a *S. pyogenes* nCas9 moiety and a rA1 or a human APOBEC3A (hA3A) moiety was used to examine whether C-to-U deamination can be induced at the ssDNA region of the R-loop generated by nSaCas9-Sa-sgRNA (Extended

¹School of Life Science and Technology and Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China. ²Gene Editing Center, School of Life Science and Technology, ShanghaiTech University, Shanghai, China. ³CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. ⁴University of Chinese Academy of Sciences, Beijing, China. ⁵CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. ⁶Department of Urology, Frontier Science Center for Immunology and Metabolism Medical Research Institute, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China. ⁷Department of Pathology, Frontier Science Center for Immunology and Metabolism, Medical Research Institute, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China. ⁸Department of Rheumatology and Immunology, Frontier Science Center for Immunology and Metabolism Medical Research Institute, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China. ⁹Department of Orthopedics, Frontier Science Center for Immunology and Metabolism Medical Research Institute, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan, China. ¹⁰These authors contributed equally: Lijie Wang, Wei Xue, Hongxia Zhang, Runze Gao, Houyuan Qiu. [✉]e-mail: yangbei@shanghaitech.edu.cn; haoyin@whu.edu.cn; liyang@picb.ac.cn; chenjia@shanghaitech.edu.cn

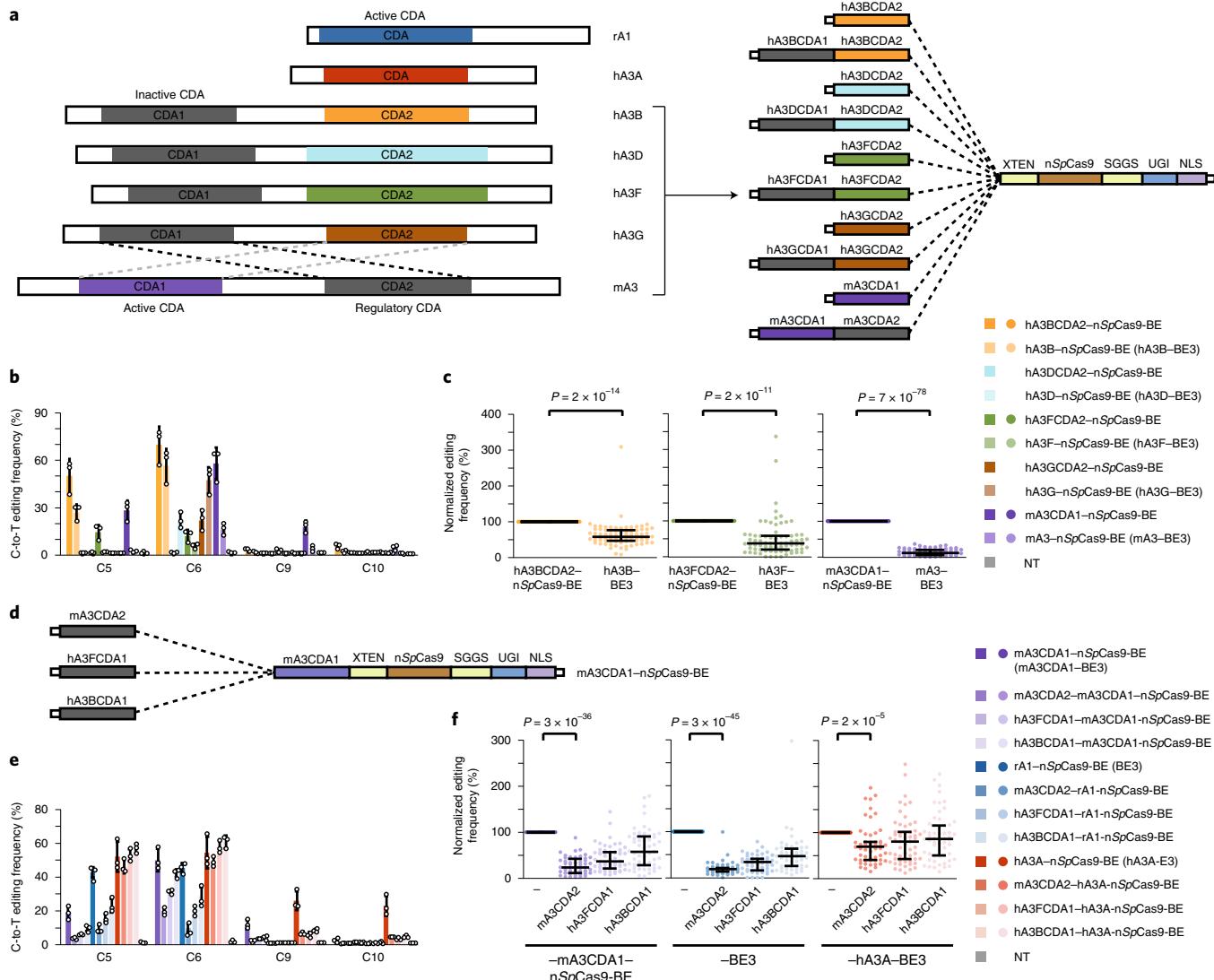


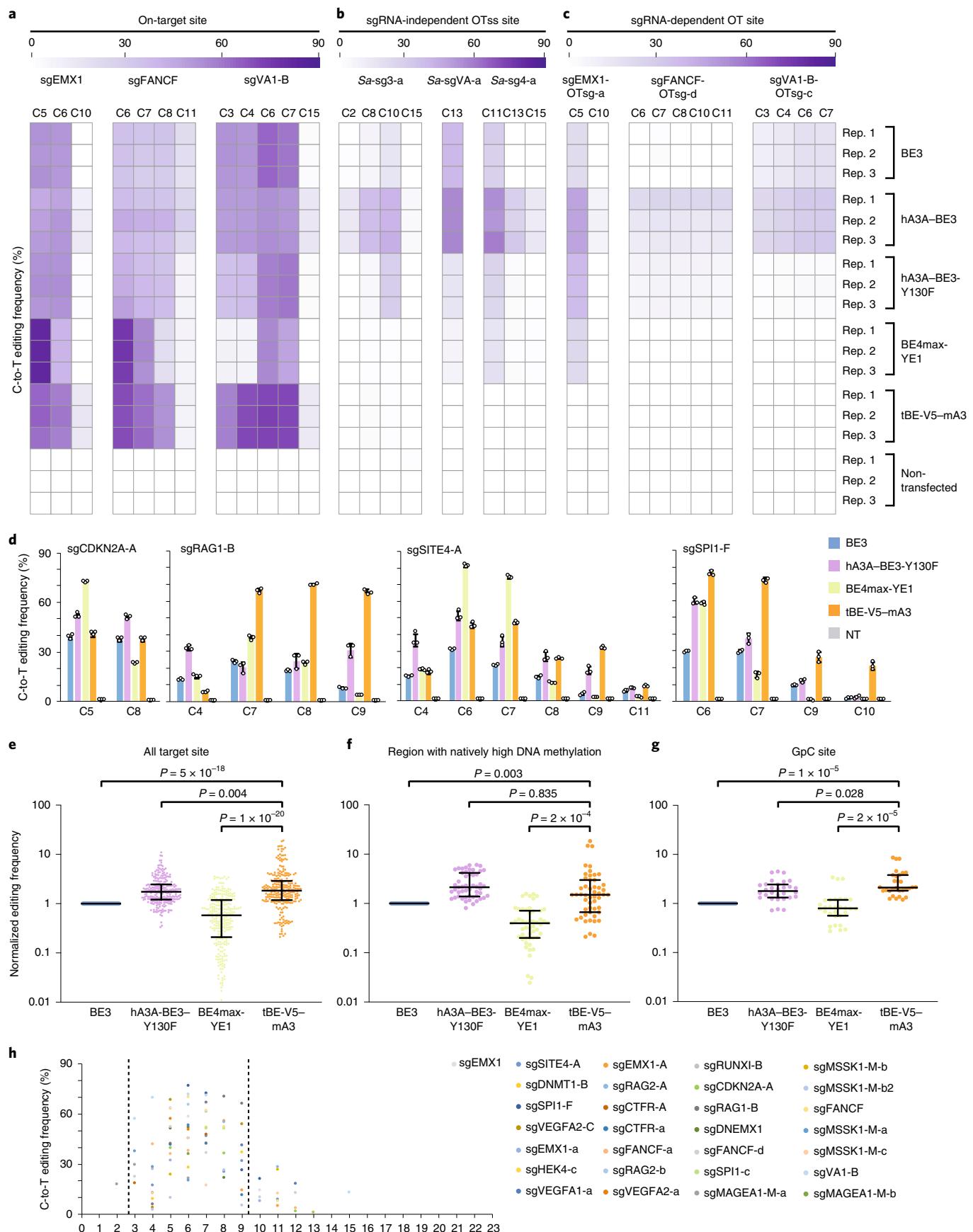
Fig. 1 | Identification of dCDI domains. **a**, Schematic of the APOBEC family members that have single or dual CDA domains (left) and BEs that were constructed with one or two CDA domains of dual-domain APOBECs (right). **b**, Editing frequencies induced by the indicated BEs at one representative genomic locus (sgTET1). **c**, Statistical analysis of normalized editing frequencies; the editing frequencies induced by the single-CDA-containing BEs were set as 100%. $n = 78$ (hA3BCDA2-nSpCas9-BE, hA3B-BE3, mA3CDA1-nSpCas9-BE and mA3-BE3) or $n = 74$ (hA3FCDA2-nSpCas9-BE and hA3F-BE3) edited cytosines at seven on-target sites from three independent experiments shown in **b** and Extended Data Fig. 2b. NT, non-transfected control. **d**, Schematic of the fusion of different dCDI domains to the N terminus of mA3CDA1-nSpCas9-BE (mA3CDA1-BE3). **e**, Editing frequencies induced by the indicated BEs at one representative genomic locus (sgTET1). **f**, Statistical analysis of normalized editing frequencies; the editing frequencies induced by the BEs without dCDI domain were set as 100%. $n = 57$ edited cytosines at five on-target sites from three independent experiments shown in **e** and Extended Data Fig. 2d. For **b** and **e**, data are mean \pm s.d. from three independent experiments. For **c** and **f**, statistical analysis was performed using two-tailed Student's *t*-tests. Data are median \pm interquartile range (IOR). Source data are available online.

Data Fig. 1a). Deep-sequencing showed that C-to-T mutations were induced by BE3 or hA3A-BE3 without the presence of *Sp*-sgRNA in nSaCas9-Sa-sgRNA-generated ssDNA regions (Extended Data Fig. 1b). These results confirmed that the rA1 or hA3A moiety of BE3 or hA3A-BE3 could induce OT mutations in a *Sp*-sgRNA (hereafter, sgRNA)-independent manner.

Identification of dCDI domains. Sequence analyses of APOBEC CDA showed that there are two groups^{22–24}. One group of APOBECs has only one CDA domain, such as rA1 and hA3A, whereas the other group has dual CDA domains, such as human APOBEC3B (hA3B) and mouse APOBEC3 (mA3) (Fig. 1a), of which only one CDA domain is thought to be catalytically active and the other is considered to be inactive but has regulatory functions^{22,23}. Thus, we

constructed ten BEs that have a catalytically active CDA domain without or with an inactive but regulatory domain (Fig. 1a and Extended Data Fig. 2a), and compared their C-to-T editing efficiencies. Surprisingly, the BEs containing the regulatory domains of certain APOBECs (hA3B, hA3F and mA3) induced significantly lower editing efficiencies compared with those that had only one active CDA domain (Fig. 1b,c and Extended Data Fig. 2b). This result suggested that the regulatory CDA domains from these dual-domain APOBECs, such as hA3B, hA3F and mA3, might have an inhibitory effect on the deoxycytidine deaminase activity of their corresponding active CDA domains.

To examine whether the inhibitory effect is general, we linked the regulatory CDA domain of mA3, hA3F or hA3B individually to the N terminus of mA3CDA1-BE3, BE3 and hA3A-BE3 (Fig. 1d



and Extended Data Fig. 2c). All of these regulatory CDA domains showed generally inhibitory effects on tested BEs. Among them, the CDA2 of mA3 (mA3CDA2) manifested the strongest inhibitory

effect (Fig. 1e,f and Extended Data Fig. 2d). Collectively, these regulatory domains exhibited generally inhibitory effects on deoxycytidine deaminase activity as dCDIs.

Fig. 2 | Comparison of tBE with other BEs. **a–c**, Comparison of the editing or mutation frequencies induced by the indicated BEs at on-target sites (**a**), sgRNA-independent OTss sites (**b**) and sgRNA-dependent OT sites (**c**). Rep., replicate. **d**, Editing efficiencies induced by the indicated BEs at four representative genomic loci. Data are mean \pm s.d. from three independent experiments. **e**, Statistical analysis of normalized editing frequencies at all 29 on-target sites shown in **a,d** and Extended Data Fig. 4a. $n=258$ edited cytosines from three independent experiments. **f**, Statistical analysis of normalized editing frequencies at six on-target sites with natively high DNA methylation levels shown in Extended Data Fig. 4a. $n=51$ edited cytosines from three independent experiments. **g**, Statistical analysis of normalized editing frequencies at 10 on-target sites containing GpC dinucleotides shown in Extended Data Fig. 4a. $n=30$ edited cytosines from three independent experiments. For **e–g**, the editing frequencies induced by the BE3 were set as 100%. Statistical analysis was performed using two-tailed Student's *t*-tests. Data are median \pm IQR. **h**, The editing window of tBE-V5-mA3. The region between two dashed lines is the major editing window of tBE-V5-mA3. Cytosines were counted with the base distal to PAM setting as position 1. Source data are available online.

Development of a BE system with a cleavable dCDI. We sought to take advantage of dCDI domains to develop a BE system to avoid sgRNA-independent mutations. Ideally, this BE would induce no observable OT mutation due to the linkage of dCDI, but can be transformed to be active by cleaving off the dCDI when binding at the on-target site; it is therefore referred to as tBE. To achieve this, we first used an sgRNA containing boxB hairpins to generate an R-loop region for the intended base editing and a helper sgRNA (hsgRNA) containing an MS2 hairpin to recruit an active APOBEC (Extended Data Fig. 3a, V1) or an APOBEC link to a dCDI domain through a TEV protease cleavage site (Extended Data Fig. 3a, V2 to V5). To cleave off the dCDI domain at the on-target site, we coexpressed a free TEV (Extended Data Fig. 3b, V2) or a N22p–TEV fusion, which can be recruited to the on-target site by the boxB-containing sgRNA (Extended Data Fig. 3b, V3) through a 2A self-cleaving peptide. Although the editing efficiencies at on-target sites indicated that the dCDI domain could be efficiently cleaved off by a free TEV or an N22p–TEV fusion (Extended Data Fig. 3c, compare V2 and V3 to V1), the mutations at sgRNA-independent OTss sites remained (Extended Data Fig. 3d). These results suggested that the intact TEV protease was able to access the TEV site and cleave the dCDI domain off disregarding on-target sites, probably due to free diffusion (Extended Data Fig. 3b).

To further reduce the mutations at sgRNA-independent OTss sites, we then took advantage of a split-TEV system²⁵ for constraining the activity of TEV protease at on-target sites. In the split-TEV system, each fragment of TEV is inactive but together reconstitutes an active protease. The C terminus (TEVc) was fused to N22p, which could be recruited by a boxB-containing sgRNA, and the N terminus (TEVn) was first fused to N22p (Extended Data Fig. 3a, V4). Although on-target editing could be induced (Extended Data Fig. 3c), unignored OT mutations could still be found at sgRNA-independent OTss sites (Extended Data Fig. 3d, tBE-V4-rA1). We reasoned that the recruitment of N22p-fused TEVn and N22p-fused TEVc by a boxB-containing sgRNA would form a complex containing an intact TEV protease that could still cleave the dCDI-linking TEV site through free diffusion (Extended Data Fig. 3b, V4). Thus, we altered the splitting strategy to express N22p-fused TEVc and free TEVn (Extended Data Fig. 3a, V5). The developed tBE-V5-rA1 induced efficient editing at on-target sites (Extended Data Fig. 3c, V5); meanwhile, the levels of sgRNA-independent OT mutations were similar to the background (Extended Data Fig. 3d, V5), demonstrating that the dCDI domain was cleaved off only at on-target sites but not at

sgRNA-independent OTss sites (Extended Data Fig. 3b, V5). We also replaced rA1 with mA3CDA1 to construct tBE-V5-mA3 (Extended Data Fig. 3a), which also induced efficient on-target editing (Extended Data Fig. 3c) with no observable mutation at sgRNA-independent OTss sites (Extended Data Fig. 3d). All tBEs induced no observable mutation at sgRNA-dependent OT sites (Extended Data Fig. 3e), as efficient editing requires the binding of both sgRNA and hsgRNA at proper positions^{26,27} (Extended Data Fig. 3b).

We next compared the efficiencies of tBE-V5-mA3 with three popularly used BEs—that is, BE3, hA3A–BE3-Y130F and BE4max–YE1 across ~30 on-target sites, three sgRNA-independent OTss sites and three sgRNA-dependent OT sites in 293FT cells (Fig. 2a–d and Extended Data Fig. 4a). At the tested on-target sites, tBE-V5-mA3 and hA3A–BE3-Y130F induced higher efficiencies compared with BE3 and BE4max–YE1 (Fig. 2e), including the sites with natively high DNA methylation levels¹⁴ (Fig. 2f) and the sites containing GpC dinucleotides (Fig. 2g). After analysing the editable positions at tested target sites, we found that the editing window of tBE-V5-mA3 spanned from position 3 to 9 (Fig. 2h, counting the protospacer adjacent motif (PAM) distal position in target site as 1). Recently, BE4max–YE1 was reported to induce efficient editing with only a background level of sgRNA-independent OT mutations^{19,28}, which has a narrower editing window (position 5 to 7)^{1,29}. Thus, we further compared their efficiencies in the narrow editing window (Extended Data Fig. 4b) and the wide window excluding the narrow one (Extended Data Fig. 4c) and found that tBE-V5-mA3 induced significantly higher efficiencies compared with BE4max–YE1 in both comparisons. At sgRNA-independent OTss sites and sgRNA-dependent OT sites, tBE-V5-mA3 manifested only background levels of OT mutations compared with other BEs (Fig. 2b,c). Furthermore, we compared the above four BEs in U2OS cells and obtained similar results (Extended Data Figs. 5a–c).

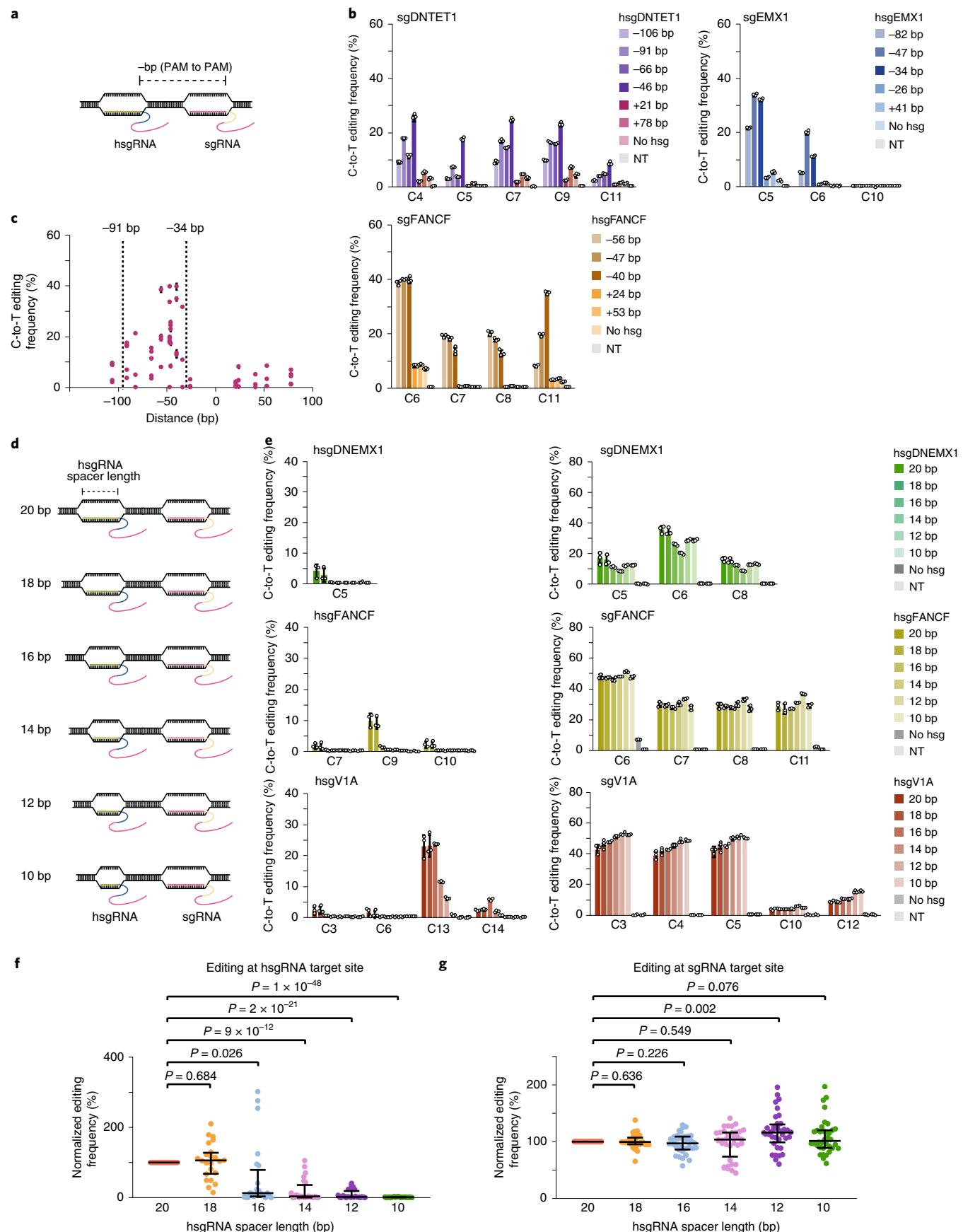
Improving the tBE system. When comparing editing efficiencies, we noticed that the distance between sgRNA and hsgRNA influenced on-target editing efficiencies of tBE. Thus, we determined the effect of the distance between sgRNA and hsgRNA (Fig. 3a,b) and found that tBE achieved optimal editing efficiency when the hsgRNA is in the –91 bp to –34 bp (upstream) region of the sgRNA (Fig. 3c, PAM to PAM distance).

As a hsgRNA also forms an R-loop region at its binding site (Fig. 3d), unintended C-to-T editing occasionally occurred at the target sites of some hsgRNAs (Fig. 3e, left). Previous studies showed that the

Fig. 3 | Optimization of hsgRNA in the tBE system. **a–c**, Characterization of the distance between hsgRNA and sgRNA for optimal editing efficiency. The schematic shows the distance between hsgRNA and sgRNA (**a**). The editing frequency at individual target sites (**b**) and summarized analysis (**c**) show that the distance range for optimal editing is –91 to –34 bp. **d–g**, Truncation of hsgRNA spacer eliminates unintended editing at hsgRNA target sites. The schematic illustrates the length of the hsgRNA spacer region (**d**). The editing frequency (**e**) and statistical analysis show that the truncation of the hsgRNA spacer to 10 bp removes unintended editing at hsgRNA target regions (**f**) while maintaining editing efficiency at sgRNA target regions (**g**); the editing frequencies induced by the 20 bp hsgRNA spacer were set as 100%. $n=24$ (**f**) or $n=36$ (**g**) edited cytosines at three on-target sites from three independent experiments. Statistical analysis was performed using two-tailed Student's *t*-tests. For **f** and **g**, data are median \pm IQR. For **b** and **e**, data are mean \pm s.d. from three independent experiments. Source data are available online.

use of guide RNA with a truncated spacer did not affect the binding of Cas9 (ref. ³⁰). We therefore examined whether unintended editing in hsgRNA target sites could also be suppressed by truncated

hsgRNAs, which form narrower R-loop regions (Fig. 3d). The unintended editing in hsgRNA target regions was reduced when truncated hsgRNAs were used (Fig. 3e,f). Furthermore, hsgRNAs with



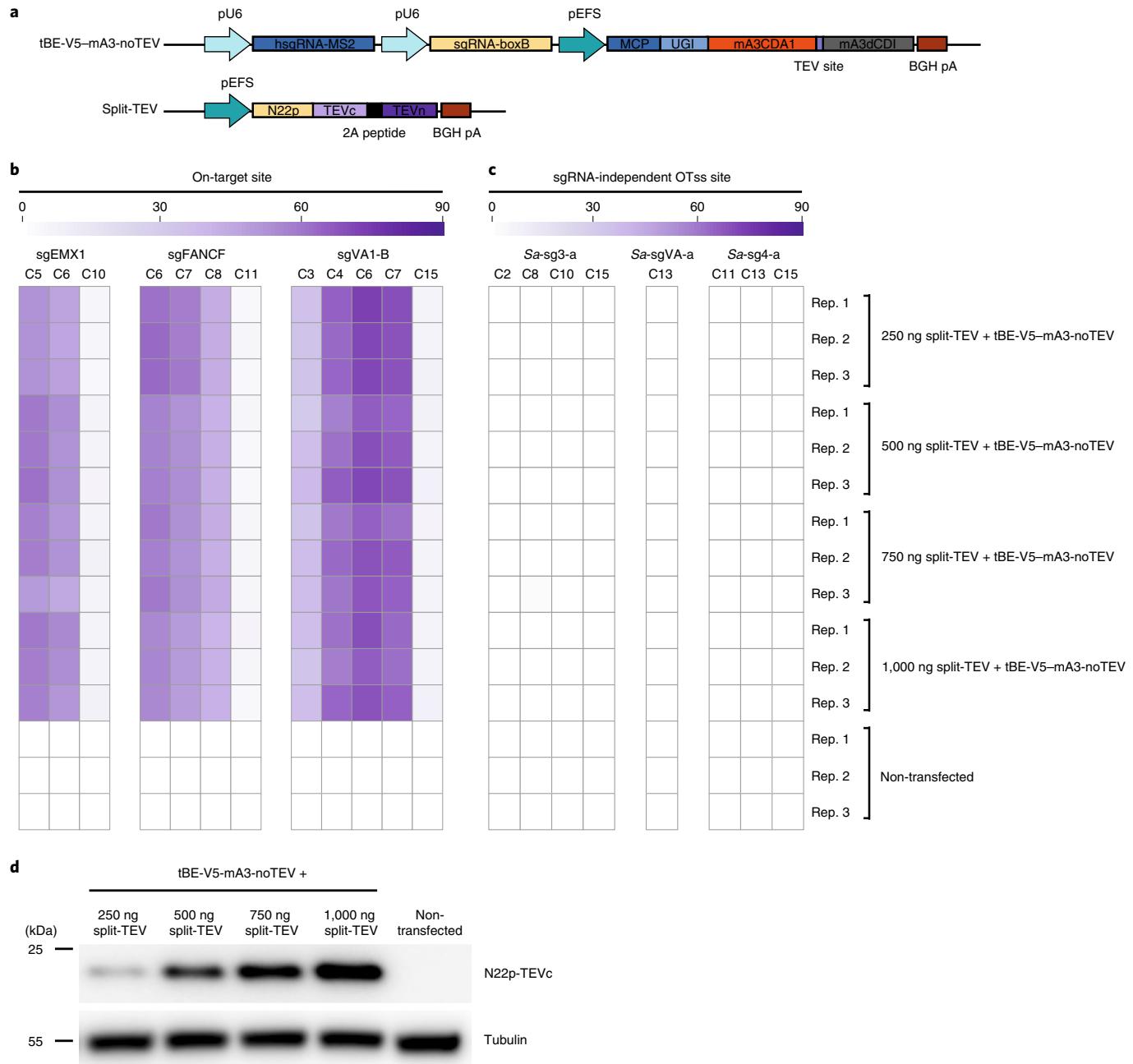


Fig. 4 | The effect of split-TEV protease concentration on editing efficiency and sgRNA-independent OT mutation. **a**, Schematic of the constructions for a tBE without the part of split-TEV protease (tBE-V5-mA3-noTEV) and for a split-TEV protease (split-TEV). **b,c**, Determining the editing or mutation efficiencies induced by the coexpression of split-TEV with different concentrations and tBE-V5-mA3-noTEV. The heat maps of editing efficiencies show that on-target editing efficiencies (**b**) and mutation frequencies at sgRNA-independent OTss sites (**c**) were not affected by the expression levels of the split-TEV components. **d**, The N22p-TEVc fusion protein expression level was examined with western blot using tubulin as the loading control. Western blot data are representative of three independent experiments. Source data are available online.

a 10 bp spacer eliminated unintended editing (Fig. 3e,f), while the on-target editing was unaffected (Fig. 3e,g). We also expressed the split-TEV system (Fig. 4a) with different amounts of expression vector and found that both on-target editing efficiencies (Fig. 4b) and mutation frequencies at sgRNA-independent OTss sites (Fig. 4c) were not affected at the examined protein levels of split-TEV components (Fig. 4d).

tBE-V5-mA3 corrects pathogenic mutations without OT mutations. Next, we sought to examine whether tBE induces OT mutations genome-wide when correcting pathogenic mutations. We

previously established a 293FT cell line harbouring a T-to-C substitution in the phosphoglucomutase 3 (*Pgm3*) gene (293FT^{PGM3(T248C)}), which is associated with hyper-IgE syndrome^{31,32}. We applied five BEs, that is, BE3, hA3A-BE3-Y130F, BE4max-YE1, tBE-V5-mA3 and tBE-V5-rA1, to correct this point mutation, and used Cas9 as the negative control. As endogenous APOBEC3 CDAs trigger the deamination of cytosine in genomic DNA^{15,23}, we further knocked out endogenously expressed APOBEC3 to reduce the background level of mutations on cytosine or guanine in 293FT^{PGM3(T248C)} cells (Fig. 5a and Extended Data Fig. 6a,b). APOBEC3-knockout 293FT^{PGM3(T248C)} cells from a single clone were expanded and then

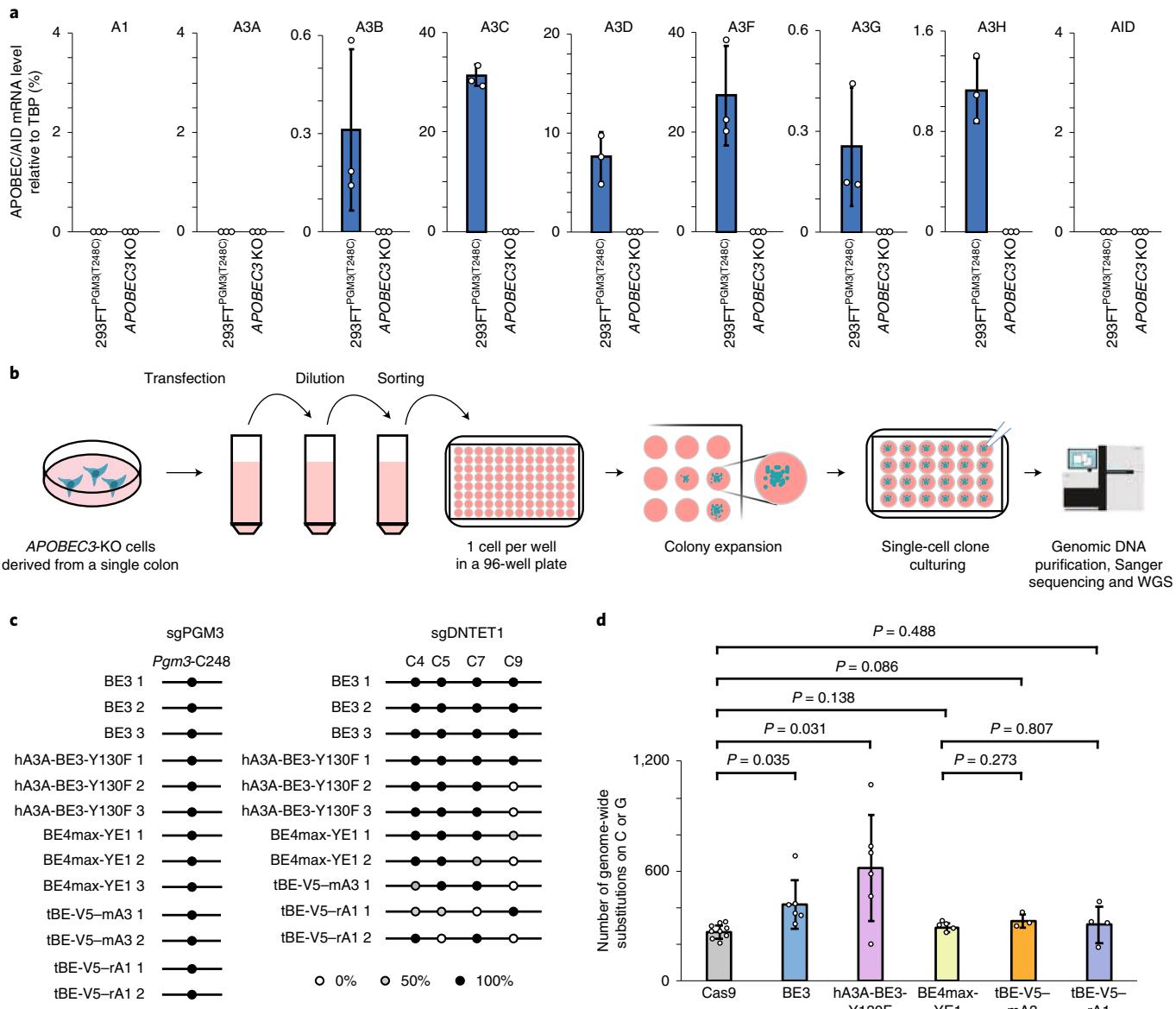


Fig. 5 | tBE eliminates OT mutations genome-wide. **a**, The mRNA expression level of endogenous APOBEC CDAs in 293FT^{PGM3(T248C)} or APOBEC3-knockout 293FT^{PGM3(T248C)} (APOBEC3 KO) cells. Data are mean \pm s.d. from three independent experiments. **b**, Schematic of the procedures to determine the genome-wide mutations induced by BEs. **c**, On-target editing efficiencies in the single-cell clones that were processed for whole-genome sequencing. **d**, The numbers of genome-wide substitutions of C or G in the single cell clones treated with Cas9 or the indicated BEs. Statistical analysis was performed using two-tailed Student's *t*-tests. Data are mean \pm s.d. from $n=9$ (Cas9), $n=6$ (BE3, hA3A-BE3-Y130F), $n=5$ (BE4max-YE1), $n=3$ (tBE-V5-mA3) or $n=4$ (tBE-V5-rA1) single-cell colonies. Source data are available online.

transfected with five BEs individually with the corresponding sgPGM3 (Fig. 5b). After transfection, single clones were sorted and then applied to whole-genome sequencing after colony expansion to detect OT mutations using the base/prime editor induced DNA off-target site identification unified toolkit (BEIDOU) (Fig. 5b and Methods). Another sgRNA (sgDNTET1) was also used in same procedures to detect genome-wide OT mutations by BEs (Fig. 5c,d). We first confirmed that using sgPGM3 corrected the *Pgm3*(T248C) mutation and using sgDNTET1 induced on-target editing (Fig. 5c). We next compared genome-wide OT mutations and found that only background levels of OT mutations on cytosine or guanine genome-wide were detected in tBE-V5-mA3- or tBE-V5-rA1-transfected single-cell colonies (Fig. 5d ($P>0.05$) and Extended Data Fig. 6c), similar to those in Cas9 or BE4max-YE1 (ref. 19). By contrast, BE3 and hA3A-BE3-Y130F induced a large

amount of genome-wide OT mutations (Fig. 5d ($P<0.05$) and Extended Data Fig. 6c), consistent with previous studies^{8,28}. Further analysis showed that sequence contexts of these genome-wide OT mutations induced by BE3 and hA3A-BE3-Y130F across different single-cell colonies were rarely overlapped, indicating they were generated randomly by the rA1 or hA3A moiety of BE3 or hA3A-BE3-Y130F in an sgRNA-independent manner (Extended Data Fig. 6d).

As recent studies also revealed BE-induced OT mutations in RNA^{9,10}, we performed whole-transcriptome sequencing and used the RADAR method³³ to determine RNA OT editing events in correcting the *Pgm3*(T248C) mutation (Fig. 6a). Only a background level of RNA OT mutations was observed in cells transfected with tBE-V5-mA3, similar to those in non-transfected control cells or the cells transfected with Cas9 or hA3A-BE3-Y130F¹⁰ (Fig. 6b-d).

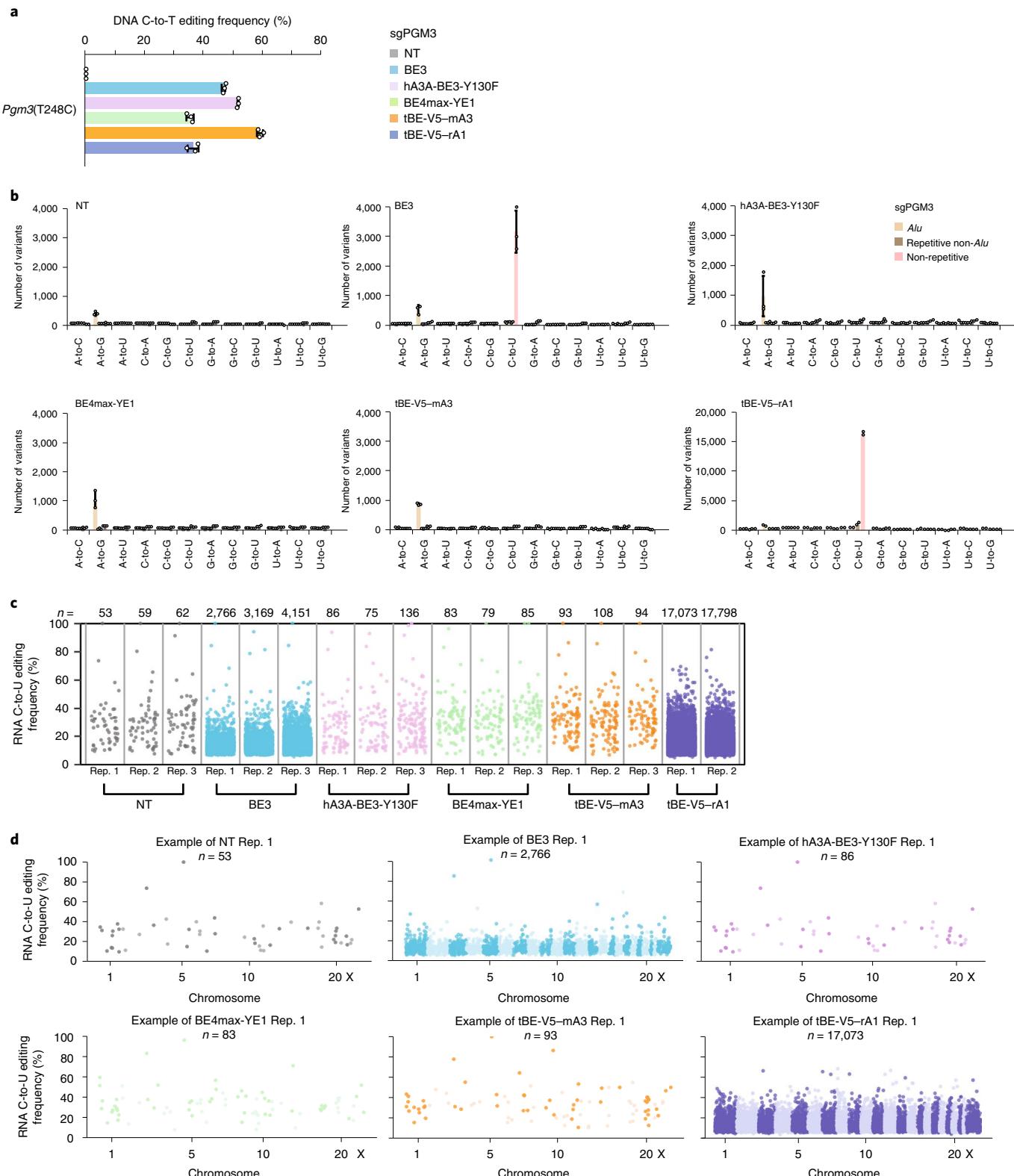


Fig. 6 | tBE eliminates OT mutations transcriptome-wide. **a**, DNA-editing frequencies induced by the indicated BEs at the *Pgm3*(248C) mutation site. Data are mean \pm s.d. from three independent experiments. **b**, The numbers of all 12 types of RNA editing in different genomic regions in the cells treated with the indicated BEs. Data are mean \pm s.d. from $n=3$ (NT, BE3, hA3A-BE3-Y130F, BE4max-YE1, tBE-V5-mA3) independent experiments or as the mean of $n=2$ (tBE-V5-rA1) independent experiments. **c**, Manhattan plot of RNA OT C-to-U editing frequencies shown in **b**. **d**, The frequencies and sites of RNA OT C-to-U editing induced by the indicated BEs in replicate 1. Source data are available online.

To further confirm that tBE-V5-mA3 did not induce RNA OT editing, we performed whole-transcriptome sequencing for the cells transfected with the BEs containing a mA3CDA1, a mA3dCDI

or a non-cleavable mA3CDA1–mA3dCDI fusion (Extended Data Fig. 7a). Consistently, no OT RNA editing was found in these cells (Extended Data Fig. 7b,c), showing that both tBE-V5-mA3

(Fig. 6b–d) and mA3CDA1–BE3 (in a non-V5 setting; Extended Data Fig. 7a–c) probably did not edit RNA³⁴. Although BE3 corrected the *Pgm3*(T248C) mutation (Fig. 6a), C-to-U RNA editing was substantially increased in BE3-treated cells (Fig. 6b–d), consistent with previous reports^{9,10}. Surprisingly, we observed that tBE-V5–rA1 induced higher levels of RNA OT mutations compared with BE3 (Fig. 6b–d). As rA1 is well known to induce RNA C-to-U editing¹¹, the increased RNA OT editing by tBE-V5–rA1 compared with rA1-conjugated BE3 is worthwhile for further investigation. Similarly, background levels of transcriptome-wide OT mutations with tBE-V5–mA3 transfection were observed when using sgDN-TET1 to perform base editing (Extended Data Fig. 7d,e). These results suggested that tBE-V5–mA3 induced highly efficient C-to-T base editing at targeted sites with no detectable genome-wide and transcriptome-wide OT mutations.

Dual-AAV delivery of tBE-V5–mA3 for in vivo editing. To demonstrate in vivo applications, we applied tBE-V5–mA3 to edit *Pcsk9*, a therapeutic target gene³⁵, by introducing a premature stop codon to repress PCSK9 protein expression. The inhibition of PCSK9 protein can reduce the levels of serum low-density lipoprotein cholesterol and is therefore associated with a low risk of cardiovascular disease³⁶. We first examined the efficiencies of tBE-V5–mA3 to generate premature stop codons at different positions in mouse *Pcsk9* in N2A cells and picked one pair of sgRNA and hsgRNA with a high efficiency for in vivo editing (Extended Data Fig. 8a, sgPCSK9-E4 and hsgPCSK9-E4). At the selected PCSK9-E4 site, tBE-V5–mA3 induced higher editing efficiencies compared with intein-BE3, intein-hA3A–BE3–Y130F and intein-BE4max–YE1 (Extended Data Fig. 8b), which were divided into halves using *trans*-splicing intein (Extended Data Fig. 8c). We also found that tBE-V5–mA3 induced efficient editing at target sites flanking an PAM of NG dinucleotides when coexpressed with nCas9-NG (Extended Data Fig. 8a).

As APOBEC is separated from nSpCas9 in the tBE system, we packaged sgRNA/hsgRNA/tBE-V5–mA3 and nSpCas9 in a dual-AAV8 system (Fig. 7a), which has high liver tropism, for in vivo base editing. We injected different titres of AAV into mice and analysed the editing efficiencies, PCSK9 protein and total cholesterol levels in mice two or four weeks after administration (Fig. 7a). Compared with the uninjected control mice, we found that, although low titres of AAV (1.6×10^{11} and 8×10^{11} viral genomes (vg)) induced less than 10% editing efficiencies, the highest titre (4×10^{12} vg) of AAV led to ~30% editing efficiencies after AAV expression for 4 weeks (Fig. 7b). Consistently, the highest titre of AAV resulted in significant reduction in both PCSK9 expression and cholesterol level (Fig. 7c; $P=0.002$ and 0.0006, respectively) without affecting serum glyceride level (Fig. 7c; $P=0.256$). As expected, a four-week expression period of AAV showed higher efficiencies and a greater reduction of PCSK9 protein and cholesterol compared with a two-week expression period (Fig. 7b,c). Furthermore, we compared the in vivo editing efficiency of tBE-V5–mA3 with three other BEs (intein-BE3, intein-hA3A–BE3–Y130F and intein-BE4max–YE1), which were also packaged in the dual-AAV system. Four weeks after injection of 4×10^{12} vg AAV, we found that no obvious editing was induced by intein-BE3, intein-hA3A–BE3–Y130F or intein-BE4max–YE1 in mouse livers, compared with the efficient editing induced by tBE-V5–mA3 (Fig. 7d), which is

consistent with the results in N2A cells (Extended Data Fig. 8b). No obvious reduction of PCSK9 protein and cholesterol was induced by intein-BE3, intein-hA3A–BE3–Y130F or intein-BE4max–YE1 (Fig. 7e). We further determined the in vivo OT effects of tBE-V5–mA3 at the genome-wide and transcriptome-wide scale. Whole-genome sequencing showed that the genome-wide mutations induced by tBE-V5–mA3 in mouse livers (AAV8 target organ) were similar to those in the kidneys (AAV8 non-target organ) of the same mice (Fig. 8a,b). Whole-transcriptome sequencing showed that the transcriptome-wide mutations in the livers of tBE-V5–mA3-treated mice were similar to those in the livers of uninjected mice (Fig. 8c,d). These results showed that tBE-V5–mA3 induced efficient in vivo editing with therapeutic benefits, with no observable genome-wide and transcriptome-wide OT mutation.

Discussion

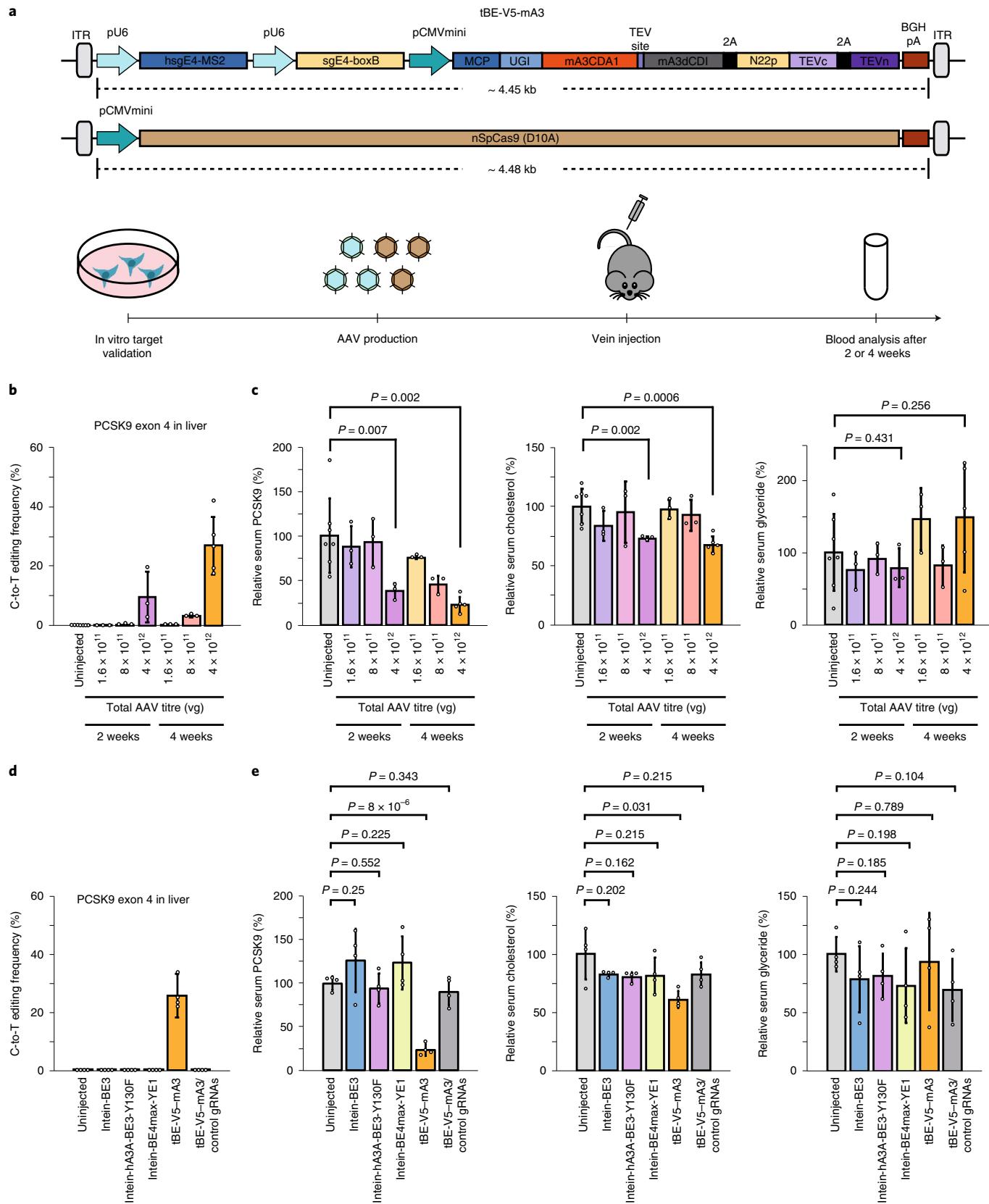
In this study, we developed a convenient and quantitative method (CESSCO) to evaluate sgRNA-independent OT mutations and found that BEs with a regular APOBEC–nCas9 backbone induced mutations at sgRNA-independent OTss sites (Extended Data Fig. 1). Recently, two groups applied similar strategies to demonstrate that the BE3 induces sgRNA-independent mutations in the ssDNA regions generated by an orthogonal dSaCas9 (ref. 19) or nSaCas9 (ref. 20). By using a nSaCas9 instead of dSaCas9 in CESSCO, the OT uracils induced by BE3-mediated cytidine deamination can be better preserved by endogenous mismatch repair³⁷. When comparing the sgRNA-independent OT mutations induced by BE3 in the ssDNA regions generated by dSaCas9 and nSaCas9, higher OT mutation frequencies were found in nSaCas9-generated ssDNA regions compared with those in dSaCas9-generated ones (Extended Data Fig. 1b), confirming the high sensitivity of CESSCO. Furthermore, as CESSCO requires only amplicon sequencing, it is cheaper and more convenient than using whole-genome sequencing to detect sgRNA-independent OT mutations for many laboratories.

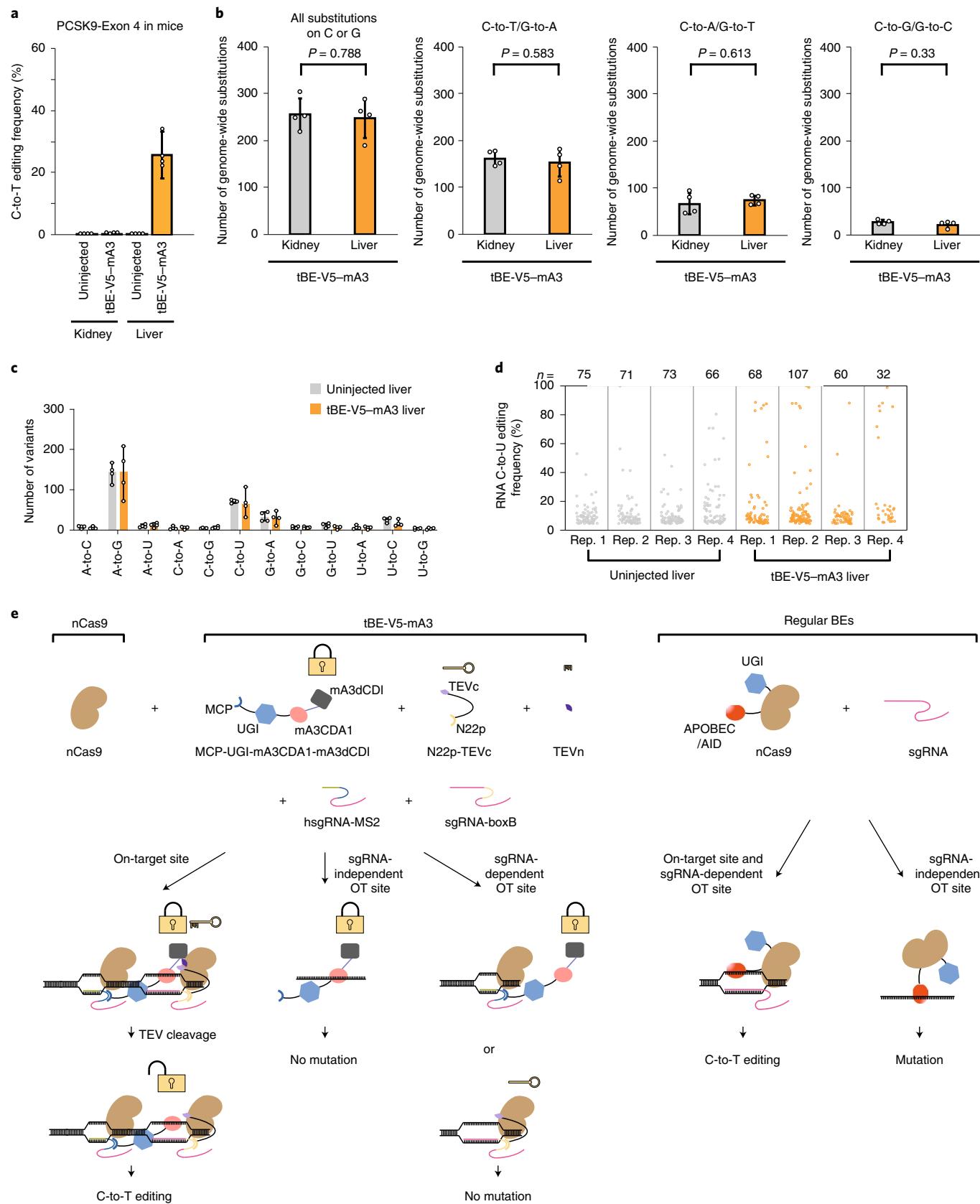
With CESSCO, we identified that inactive CDA domains of certain dual-domain APOBECs function as dCDIs and then took advantage of mA3dCDI to develop the tBE system (Figs. 1 and 2). The tBE system remains inert in general, therefore eliminating OT mutations (Fig. 8e). tBE can be transformed to be active for targeted base editing only after binding at on-target sites. Genome-wide and transcriptome-wide analyses demonstrated that tBE-V5–mA3 induced no detectable DNA and RNA OT mutations (Figs. 5 and 6). Importantly, the tBE system can be conveniently packaged into a dual-AAV system to maintain editing efficiency maximally. We used tBE to install a premature stop codon in *Pcsk9*, which significantly reduced the levels of both PCSK9 protein and total cholesterol in AAV-injected mice with no observable OT mutations (Figs. 7 and 8). This further demonstrated the potential of tBE in cell and gene therapies that require precise genome editing in primary or stem cells^{38,39}. We found that the width of the tBE-V5–mA3 editing window is around 7 bp (position 3–9), which is relatively bigger than previously reported BEs^{40,41} and is therefore suitable for generating premature stop codons. In the future, tBE versions with narrowed editing windows for precise base editing could be achieved by engineering mA3 or by adopting BEs with small editing windows, such as BE4max–YE1.

Fig. 7 | Dual-AAV delivery of tBE for in vivo base editing. **a**, Schematic of the dual-AAV8 system for tBE-V5–mA3 and the key experimental steps. **b**, In vivo editing frequencies induced by tBE-V5–mA3 with different titres and expression time periods. **c**, The levels of serum PCSK9 protein, cholesterol and glyceride levels in the mice injected with dual-AAV8 expressing tBE-V5–mA3. Statistical analysis was performed using two-tailed Student's *t*-tests. For **b** and **c**, data are mean \pm s.d. from $n=7$ (uninjected), $n=3$ (1.6×10^{11} vg, 2 weeks), $n=3$ (8×10^{11} vg, 2 weeks), $n=3$ (4×10^{12} vg, 2 weeks), $n=3$ (1.6×10^{11} vg, 4 weeks), $n=3$ (8×10^{11} vg, 4 weeks) or $n=5$ (4×10^{12} vg, 4 weeks) mice. **d**, Comparison of in vivo editing efficiencies of the indicated BEs using dual-AAV8 delivery. **e**, The levels of serum PCSK9 protein, cholesterol and glyceride levels in mice injected with dual-AAV8 expressing indicated BEs. Statistical analysis was performed using two-tailed Student's *t*-tests. For **d** and **e**, data are mean \pm s.d. from $n=4$ mice. Source data are available online.

In addition to the original *SpCas9*, tBE also worked well with Cas9-NG (Extended Data Fig. 8a), which recognizes more flexible PAMs, expanding the target scope of tBE to correct pathogenic

mutations⁴² (Extended Data Fig. 8d). Thus, the combination of tBE with other Cas9 orthologues⁴³ could further expand the targeting scope of tBE system. After characterizing the distance between





hsgRNA and sgRNA, we discovered a range of 57 bp for efficient editing, which allows a high flexibility in selecting hsgRNA (Fig. 3). Finally, as there are abundant members within APOBEC family^{23,44}, other dCDIs may be identified in the future (Extended Data Fig. 8e)

that could further enrich the tBE system. As editing precision is essential for BEs, we believe that the tBE system developed here for eliminating OT mutations sheds light on therapy-related applications.

Fig. 8 | Determining the in vivo OT effects of tBE-V5-mA3. **a**, Editing frequencies in the kidneys and livers of uninjected mice and the mice injected with dual-AAV8 expressing tBE-V5-mA3. The data in livers are the same as those shown in Fig. 7d. **b**, The numbers of genome-wide substitutions of C or G in the kidneys and livers of the mice injected with dual-AAV8 expressing tBE-V5-mA3. Statistical analysis was performed using two-tailed Student's *t*-tests. **c**, The numbers of all 12 types of RNA editing in the livers of uninjected mice and the mice injected with dual-AAV8 expressing tBE-V5-mA3. **d**, Manhattan plot of the RNA OT C-to-U editing frequencies shown in c. **e**, Schematic illustrating that tBE-V5-mA3 induces base editing at on-target (ON), but not sgRNA-independent or sgRNA-dependent OT sites. Only the reconstitution of an intact TEV protease (key) at an on-target site cleaves off the mA3dCDI domain and then unlocks the mA3CDA1 for the intended editing. The mA3CDA1 remains locked at sgRNA-independent or sgRNA-dependent OT sites at which the covalently linked mA3dCDI domain is not cleaved, as an intact TEV protease cannot be reconstituted. For a–c, data are mean \pm s.d. from $n = 4$ mice. Source data are available online.

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/s41556-021-00671-4>.

Received: 1 April 2020; Accepted: 26 March 2021;

Published online: 10 May 2021

References

- 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).
- Li, X. et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
- 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).
- Yang, B., Yang, L. & Chen, J. Development and application of base editors. *CRISPR J.* **2**, 91–104 (2019).
- Yang, L. & Chen, J. A tale of two moieties: rapidly evolving CRISPR/Cas-based genome editing. *Trends Biochem. Sci.* **45**, 874–888 (2020).
- Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
- Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (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).
- Lau, P. P., Chen, S. H., Wang, J. C. & Chan, L. A 40 kilodalton rat liver nuclear protein binds specifically to apolipoprotein B mRNA around the RNA editing site. *Nucleic Acids Res.* **18**, 5817–5821 (1990).
- Saraconi, G., Severi, F., Sala, C., Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol.* **15**, 417 (2014).
- Chen, J., Miller, B. F. & Furano, A. V. Repair of naturally occurring mismatches can induce mutations in flanking DNA. *eLife* **3**, e02001 (2014).
- Lei, L. et al. APOBEC3 induces mutations during repair of CRISPR-Cas9-generated DNA breaks. *Nat. Struct. Mol. Biol.* **25**, 45–52 (2018).
- Burns, M. B. et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366–370 (2013).
- Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **45**, 977–983 (2013).
- Chen, J., Yang, B. & Yang, L. To BE or not to BE, that is the question. *Nat. Biotechnol.* **37**, 520–522 (2019).
- Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
- 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).
- Jin, S. et al. Rationally designed APOBEC3B cytosine base editors with improved specificity. *Mol. Cell* **79**, 728–740 (2020).
- Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
- Salter, J. D., Bennett, R. P. & Smith, H. C. The APOBEC protein family: united by structure, divergent in function. *Trends Biochem. Sci.* **41**, 578–594 (2016).
- Yang, B., Li, X., Lei, L. & Chen, J. APOBEC: from mutator to editor. *J. Genet. Genomics* **44**, 423–437 (2017).
- Olson, M. E., Harris, R. S. & Harki, D. A. APOBEC enzymes as targets for virus and cancer therapy. *Cell Chem. Biol.* **25**, 36–49 (2018).
- Gray, D. C., Mahrus, S. & Wells, J. A. Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* **142**, 637–646 (2010).
- Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
- Shen, B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
- Zuo, E. et al. A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nat. Methods* **17**, 600–604 (2020).
- 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).
- Kiani, S. et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Methods* **12**, 1051–1054 (2015).
- Sassi, A. et al. Hypomorphic homozygous mutations in phosphoglucomutase 3 (PGM3) impair immunity and increase serum IgE levels. *J. Allergy Clin. Immunol.* **133**, 1410–1419 (2014).
- Wang, Y. et al. Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. *Genome Biol.* **20**, 218 (2019).
- Wang, X. et al. Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* **31**, 107723 (2020).
- Nair, S., Sanchez-Martinez, S., Ji, X. & Rein, A. Biochemical and biological studies of mouse APOBEC3. *J. Virol.* **88**, 3850–3860 (2014).
- Ding, Q. et al. Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ. Res.* **115**, 488–492 (2014).
- Horton, J. D., Cohen, J. C. & Hobbs, H. H. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem. Sci.* **32**, 71–77 (2007).
- Kunkel, T. A. & Erie, D. A. Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.* **49**, 291–313 (2015).
- Frangoul, H. et al. CRISPR-Cas9 gene editing for sickle cell disease and beta-thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).
- Koblan, L. W. et al. In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature* **589**, 608–614 (2021).
- Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
- Levy, J. M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat. Biomed. Eng.* **4**, 97–110 (2020).
- Stenson, P. D. et al. The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136**, 665–677 (2017).
- Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
- Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **4**, 868–877 (2004).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Plasmid construction. A primer set (hA3BCDA2_BE3_PCR_F/hA3BCDA2_BE3_PCR_R) was used to amplify the human hA3BCDA2 fragment using the template pUC57-humanAPOBEC3B (synthesized by Genscript). The amplified human hA3BCDA2 fragment was then cloned into NotI- and SmaI-linearized pCMV-BE3 (ref. 1) using the Clone Express plasmid recombination kit (C112-02, Vazyme) to generate the vector pCMV-hA3BCD A2-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS for the expression of hA3BCDA2-nSpCas9-BE. The vectors pCMV-hA3DCDA2-XTEN-nSpCas9 (D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-hA3FCDA2-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-hA3CDCA2-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS and pCMV-mA3CDA1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS for the expression of hA3DCDA2-nSpCas9-BE, hA3FCDA2-nSpCas9-BE, hA3GCDA2-nSpCas9-BE, mA3CDA1-nSpCas9-BE and mA3dCDI-nSpCas9-BE were constructed using the same strategy, respectively.

Two primer sets (mA3CDA2-mA3CDA1_BE3_PCR_F/mA3CDA2_PCR_R and mA3CDA1_PCR_F/mA3CDA1_BE3_PCR_R) were used to amplify the fragments mA3CDA2 and mA3CDA1, respectively. The fragment mA3CDA2-mA3CDA1 was next amplified from the fragments mA3CDA2 and mA3CDA1 using a primer set (mA3CDA2-mA3CDA1_BE3_PCR_F/mA3CDA1_BE3_PCR_R) and cloned into NotI- and SmaI-linearized pCMV-BE3 to generate the vector pCMV-mA3CDA2-mA3CD A1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS for the expression of mA3CDA2-mA3CDA1-nSpCas9-BE. The vectors pCMV-hA3FCDA1-mA3CD A1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-hA3BCDA1-m A3CDA1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-mA3C DA2-rA1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-hA3F CDA1-rA1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-hA3 BCDA1-rA1-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-mA 3CDA2-hA3A-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS, pCMV-h A3FCDA1-hA3A-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS and pCMV-hA3BCDA1-hA3A-XTEN-nSpCas9(D10A)-SGGS-UGI-SGGS-SV40NLS for the expression of hA3FCDA1-mA3CDA1-nSpCas9-BE, hA3BCDA1-mA3CDA1-nSpCas9-BE, mA3CDA2-rA1-nSpCas9-BE, hA3BCDA1-rA1-nSpCas9-BE, mA3CDA2-hA3A-nSpCas9-BE, hA3FCDA1-hA3A-nSpCas9-BE and hA3BCDA1-hA3A-nSpCas9-BE were constructed using the same strategy, respectively.

Oligonucleotides hTET1_FOR/hTET1_REV were annealed and ligated into BsaI-linearized pGL3-U6-sgRNA-PGK-puromycin to generate the vector psgTET1 for the expression of sgTET1. The oligonucleotides hSa-sg3-a_FOR/hSa-sg3-a_REV were annealed and ligated into BsaI-linearized pGL3-U6-Sa-sgRNA-PGK-puromycin to generate the vector pSa-sg3-a for the expression of Sa-sg3-a. Other expression vectors for sgRNA or Sa-sgRNA were constructed using a similar strategy.

Primer sets (BamHI_2U6_MS2_F/Sall_2U6_MS2_R) were used to amplify the fragment scaffold-MS2 with the template pLH-sgRNA1-1XMS2. The amplified fragment scaffold-MS2 was next cloned into BamHI- and Sall-linearized pUC57kan-T7-gRNA-U6 V2 to generate the vector pUC57-sgRNA-MS2-U6.

Two primer sets (MluI_Assembly_F/boxB_BsmBI_ccdB_R and boxB_F/HindIII_sgRNA_scaffold_R) were used to amplify the fragments U6-ccdB and sgRNA-boxB with the template pGL3-U6-XsgRNA-ccdB-EF1a-Puromycin and pLH-sgRNA1-2XboxB, respectively. The fragment U6-ccdB-boxB was then amplified from the fragments U6-ccdB and sgRNA-boxB using the primer set (MluI_Assembly_F/HindIII_sgRNA_scaffold_R) and cloned into MluI- and HindIII-linearized pHAGE-EFS-MCP-3XBFPNls to generate the vector pU6-ccdB-boxB-EFS-MCP-3XBFPNls. Two primer sets (BamHI_SV40NLS_UGI_F/ratA1_UGI_R and ratA1_F/XhoI_ratA1-R) were used to amplify the fragments SV40NLS-UGI and SG-rA1, respectively. The fragment SV40NLS-UGI-ratA1 was then amplified from the fragments SV40NLS-UGI and SG-rA1 using the primer set (BamHI_SV40NLS_UGI_F/XhoI_ratA1-R) and cloned into BamHI- and XhoI-linearized pU6-ccdB-boxB-EFS-MCP-3XBFPNls to generate the vector pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SG GS-rA1-TS-mA3dCDI-SV40NLS-T2A-TEV-SV40NLS for the expression of U6-ccdB-boxB-EFS-MCP-UGI-ratA1 (U6-ccdB-boxB-tBE-V1-rA1).

Two primer sets (ratA1_TS_mA3dCDI_F/T2A_R and T2A_TEVn_F/XhoI_SV40_TEV_R) were used to amplify the fragments TS-mA3dCDI and T2A-TEV, respectively. The fragment TS-mA3dCDI-T2A-TEV was then amplified from the fragments TS-mA3dCDI and T2A-TEV using the primer set (ratA1_TS_mA3dCDI_F/XhoI_SV40_TEV_R) and cloned into XhoI-linearized U6-ccdB-boxB-tBE-V1-rA1 to generate the vector pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SG GS-rA1-TS-mA3dCDI-SV40NLS-T2A-TEV-SV40NLS for the expression of U6-ccdB-boxB-EFS-MCP-UGI-ratA1-TS-mA3dCDI-T2A-TEV (U6-ccdB-boxB-tBE-V2-rA1). The vectors pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-rA1-TS-mA3d CDI-SV40NLS-T2A-N22p-(SGGS)₂-XTEN-(SGGS)₂-TEV-SV40NLS, pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-rA1-TS-mA3dCDI-SV40NLS-P2A-N22p-(SGGS)₂-XTEN-

(SGGS)₂-TEVc-SV40NLS-T2A-N22p-(SGGS)₂-XTEN-(SGGS)₂-TEVn-SV40NLS, pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-rA1-TS-mA3dCDI-SV40NLS-P2A-N22p-(SGGS)₂-XTEN-(SGGS)₂-TEV-SV40N LS-T2A-TEVn-SV40NLS, pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-mA3CDA1-TS-mA3dCDI-SV40NLS-P2A-N22p-(SGGS)₂-XTEN-(SGGS)₂-TEVc-SV40NLS-T2A-TEVn-SV40NLS and pU6-ccdB-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-mA3CDA1-TS-mA3dCDI-SV40NLS for the expression of U6-ccdB-boxB-EFS-MCP-UGI-ratA1-TS-mA3dCDI-T2A-N22p-TEV (U6-ccdB-boxB-tBE-V3-rA1), U6-ccdB-boxB-EFS-MCP-UGI-ratA1-TS-mA3dCDI-P2A-N22p-TEVc-T2A-TEVn (U6-ccdB-boxB-tBE-V4-rA1), U6-ccdB-boxB-EFS-MCP-UGI-ratA1-TS-mA3dCDI-P2A-N22p-TEVc-T2A-TEVn (U6-ccdB-boxB-tBE-V5-rA1), U6-ccdB-boxB-EFS-MCP-UGI-mA3CDA1-TS-mA3dCDI (U6-ccdB-boxB-tBE-V5-mA3-notEV) were constructed using the same strategy, respectively.

Primer sets (hsgEMX1_MS2_FOR/sgEMX1_boxB_REV) were used to amplify the fragment hsgEMX1-MS2-U6-sgEMX1 using the template pUC57-sgRNA-MS2-U6. The fragment hsgEMX1-MS2-U6-sgEMX1 was then ligated into BsmBI-linearized U6-ccdB-boxB-tBE-V1-rA1 to generate the vector pU6-hsgEMX1-MS2-U6-sgEMX1-boxB-EFS-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-rA1-SV40NLS for the expression of U6-hsgEMX1-MS2-U6-sgEMX1-boxB-EFS-MCP-UGI-ratA1 (tBE-V1-rA1-EMX1). tBE-V1-rA1, tBE-V2-rA1, tBE-V3-rA1, tBE-V4-rA1, tBE-V5-rA1, tBE-V5-mA3 and tBE-V5-mA3-noTEV with different on-target helper sgRNA and sgRNA were constructed using the same strategy, respectively.

Primer sets (NcoI_N22p_F/XhoI_TEVn_R) were used to amplify the fragment N22p-TEVc-2A-TEVn with the template tBE-V5-mA3. The fragment N22p-TEVc-2A-TEVn was then cloned into NcoI- and XhoI-linearized pHAGE-EFS-MCP-3XBFPNls to generate the vector pEFs-N22p-(SGGS)₂-XTEN-(SGGS)₂-TEVc-SV40NLS-T2A-TEVn-SV40NLS for the expression of Split-TEV.

Primer sets (ITR_HindIII_U6_F/ITR_RsrII_BGH_R) were used to amplify the fragment HindIII-tBE-V5-mA3-PCSK9-E4-RsrII using the template tBE-V5-mA3-E4. The fragment HindIII-tBE-V5-mA3-PCSK9-E4-RsrII was then ligated into HindIII- and RsrII-linearized AAV2-CAG-GFP to generate the vector ITR-pU6-hsgPCSK9-E4-MS2-U6-sgPCSK9-E4-boxB-CMV-MCP-(SGGS)₂-XTEN-(SGGS)₂-SV40NLS-UGI-SGGS-mA3CDA1-TS-mA3dCDI-SV40NLS-P2A-N22p-(SGGS)₂-XTEN-(SGGS)-TEVc-SV40NLS-T2A-TEVn-SV40NLS for the expression of ITR-tBE-V5-mA3-PCSK9-E4. The vectors ITR-pCMV-SV40NLS-rA1-XTEN-nSpCas9(1-573 aa)-Nintein, ITR-pCMV-SV40NLS-hA3A-Y130F-XTEN-nSpCas9(1-573 aa)-Nintein, ITR-pCMV-SV40NLS-rA1-(SGGS)₂-XTEN-(SGGS)₂-nSpCas9(1-573 aa)-Nintein, ITR-pCMV-Cintein-nSpCas9(574-1368 aa)-SGGS-UGI-SV40NLS-U6-sgPCSK9-E4 and ITR-pCMV-Cintein-nSpCas9(574-1368 aa)-SGGGSGSGGS-UGI-SGGSGSGGS-UGI-SV40NLS-U6-sgPCSK9-E4, for the expression of ITR-intein-BE3-N, ITR-intein-hA3A-BE3-Y130F-N, ITR-intein-BE4max-YE1-N, ITR-intein-BE3-C-PCSK9-E4 and intein-BE4max-YE1-C-PCSK9-E4 were constructed using the same strategy.

A list of the sequences of the oligos used for plasmid construction is provided in Supplementary Table 1 and a list of the amino acid sequences of plasmids is provided in the Supplementary Information. The step-by-step protocol for plasmid construction of the tBE system is available at the Protocol Exchange⁴⁵.

Cell culture and transfection. 293FT (Thermo Fisher Scientific, R70007), HEK293T/17 (ATCC CRL-11268), U2OS (ATCC, HTB-96) and N2A (ATCC CCL-131) cells were maintained in DMEM (10566, Gibco/Thermo Fisher Scientific) + 10% FBS (16000-044, Gibco/Thermo Fisher Scientific) and regularly tested to exclude mycoplasma contamination.

For base editing using regular BEs, 293FT and U2OS cells were seeded in a 24-well plate at a density of 1×10^5 per well and transfected with 250 μ l serum-free Opti-MEM that contained 3.75 μ l LIPOFECTAMINE LTX (Life, Invitrogen), 1.5 μ l LIPOFECTAMINE plus (Life, Invitrogen), 0.5 μ g pCMV-BE3 (or hA3B-BE3, hA3BCDA2-nSpCas9-BE, hA3D-BE3, hA3DCDA2-nSpCas9-BE, hA3F-BE3, hA3FCDA2-nSpCas9-BE, hA3G-BE3, hA3GCDAA2-nSpCas9-BE, mA3-BE3, mA3CDA1-nSpCas9-BE, mA3CDA2-mA3CDA1-nSpCas9-BE, hA3FCDA1-mA3CDA1-nSpCas9-BE, hA3BCDA1-mA3CDA1-nSpCas9-BE, mA3CDA2-rA1-nSpCas9-BE, hA3FCDA1-rA1-nSpCas9-BE, hA3BCDA1-rA1-nSpCas9-BE, hA3A-BE3, mA3CDA2-hA3A-nSpCas9-BE, hA3FCDA1-hA3A-nSpCas9-BE, hA3BCDA1-hA3A-nSpCas9-BE, hA3GCDAA2-nSpCas9-BE, mA3dCDI-nSpCas9-BE, hA3A-BE3, hA3A-BE3-Y130F, BE4max- YE1 and pCMV-dSpCas9) expression vector, 0.32 μ g sgRNA expression vector without or with 0.5 μ g Sa-sg-nSaCas9 expression vector. For base editing with transformer BEs, 293FT, U2OS and N2A cells were transfected with 250 μ l serum-free Opti-MEM containing 3.75 μ l LIPOFECTAMINE LTX, 1.5 μ l LIPOFECTAMINE plus, 0.5 μ g tBE-V1-rA1 (or tBE-V2-rA1, tBE-V3-rA1, tBE-V4-rA1, tBE-V5-rA1 and tBE-V5-mA3) expression vector, 0.5 μ g pEFS-nSpCas9 or pEFS-nSpCas9-NG expression vector without or with 0.5 μ g Sa-sg-nSaCas9 expression vector.

For split-TEV concentration analysis only, 293FT cells were transfected with 250 µl serum-free Opti-MEM containing 6.25 µl LIPOFECTAMINE LTX (Life, Invitrogen), 2.5 µl LIPOFECTAMINE plus (Life, Invitrogen), 0.5 µg tBE-V5-mA3-noTEV expression vector, 0.5 µg pEF3-nSpCas9 expression vector, 0.5 µg *Sa*-sg-n*Sa*Cas9 expression vector with 250 ng, 500 ng, 750 ng or 1,000 ng split-TEV expression vector. After 24 h, puromycin (ant-pr-1, InvivoGen) was added to the medium at a final concentration of 4 µg ml⁻¹. After another 48 h, the genomic DNA was extracted from the cells using QuickExtract^T DNA Extraction Solution (QEP09050, Epicentre) for subsequent sequencing analysis or the cells were lysed by 2× SDS loading buffer for western blot analysis.

To establish APOBEC3-knockout cell lines, the 293FT cells containing *Pgm3*(T248C) were seeded into a 60-mm plate at a density of 4 × 10⁵ per well and cultured for 24 h. Cells were transfected with plasmids expressing Cas9 and two APOBEC-targeting sgRNAs (sghA3A and sghA3H), according to the manufacturer's instructions. After 48 h, 10 µg ml⁻¹ puromycin was added to the medium for the next two days. The knockout of APOBECs in the cells expanded from single colonies was validated by genomic DNA PCR and quantitative PCR with reverse transcription (RT-qPCR).

Western blot. Transfected cells were lysed in protein loading buffer and incubated at 95 °C for 20 min, separated by SDS-PAGE (Genscript) in loading buffer and proteins were transferred to nitrocellulose membranes (Thermo Fisher Scientific). After blocking with TBST (25 mM Tris pH 8.0, 150 mM NaCl and 0.1% Tween-20) containing 5% (w/v) non-fat dry milk for 2 h, the membrane was reacted overnight with the indicated primary antibodies. After extensive washing, the membranes were reacted with horseradish-peroxidase-conjugated secondary antibodies for 1 h. Reactive bands were developed in ECL (Thermo Fisher Scientific) and detected with Amersham Imager 680. Uncropped blots are provided.

Antibodies. Primary antibodies were purchased from the following sources: anti-alpha-tubulin (T6199, clone DM1A), anti-2A peptide (MABS2005, clone 3H4) at 0.5 µg ml⁻¹ for western blot from Sigma-Aldrich. Secondary antibodies were purchased from the following sources: horseradish-peroxidase-conjugated goat anti-mouse (115-035-146) for western blot at 1:5,000 dilutions from Jackson ImmunoResearch.

AAV vector production. HEK293T/17 cells were maintained in 150-mm dishes in DMEM supplemented with 10% FBS (Gibco), 100 U ml⁻¹ penicillin and 100 mg ml⁻¹ streptomycin. For each transfection, 2 fmol of each RC8 packaging plasmid (15.2 µg), helper plasmid (24 µg) and AAV2-ITR plasmid carrying the construct of interest (the amount of DNA was calculated by the size of construct) were transfected into cells using PEI solution. The pellet and supernatant were collected 72 h after transfection, and the AAV8 viral particles were purified using iodixanol gradient ultracentrifugation. The purified viral particles were then titred using qPCR using specific primers against the transgene, and the purity of AAV8 vector was further verified by SDS-PAGE. The AAV8-nSpCas9 (D10A) and AAV8-tBE-V5-mA3 targeting PCSK9-E4 were provided by PackGene Biotech and the titres were 1 × 10¹³ vg per ml.

Animal studies. Mice were given free access to food and water, and were maintained under a 12 h–12 h light–dark cycle with controlled temperature (20–25 °C) and humidity (50 ± 10%). All mouse studies were approved by the Animal Care and Ethical Committee at Wuhan University. The step-by-step protocol for tBE-mediated base editing *in vivo* is available at the Protocol Exchange⁴⁵. AAV vector was delivered to at least three female C57BL/6 mice (aged 8 weeks) intravenously through lateral tail vein injection. Each AAV was adjusted to 150–200 µl per mouse with sterile PBS before injection. The total doses of AAV were 1.6 × 10¹¹, 8 × 10¹¹ or 4 × 10¹² vg per mouse. Two or four weeks after injection, the blood was collected and the serum was separated by centrifugation. Serum levels of PCSK9 and total cholesterol were measured using the Mouse PCSK9 ELISA Kit (MPC-900, R&D Systems) and total cholesterol kit (1040280, Shanghai ShenSuo UNF Medical Diagnostic Articles), respectively. The tissues were cryogenically ground into powder. The genomic DNA and total RNA were isolated using the E.Z.N.A. Tissue DNA Kit (D3396-01, Omega Bio-tek) and the HiPure Total RNA Plus Mini Kit (R4121-02, Magen), respectively.

DNA library preparation and sequencing. Target genomic sequences were PCR-amplified using high-fidelity DNA polymerase PrimeSTAR HS (Clontech) with primer sets flanking the examined sgRNA target sites. A list of the sgRNA target sites and PCR primers is provided in Supplementary Table 2. Indexed DNA libraries were prepared using the TruSeq ChIP Sample Preparation Kit (Illumina) with minor modifications. In brief, the PCR products amplified from genomic DNA regions were fragmented by Covaris S220, and then PCR-amplified using the TruSeq ChIP Sample Preparation Kit (Illumina). After quantification using the Qubit High-Sensitivity DNA kit (Invitrogen), PCR products with different tags were pooled together for deep sequencing using the Illumina HiSeq X10 (2 × 150) or NextSeq 500 (2 × 150) system at the CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China. Raw read qualities were evaluated using FastQC (v.0.11.5, <http://www.bioinformatics.babraham.ac.uk/>

projects/fastqc/). For paired-end sequencing, only R1 reads were used. Adaptor sequences and read sequences on both ends with a Phred quality score lower than 30 were trimmed. Clean reads were then mapped using the BWA-MEM algorithm (v.0.7.17-r1188) to target sequences. After being piled up with Samtools (v.1.9), base substitutions were further calculated according to previously published literature^{3,14}.

Base substitution calculation. Base substitutions were selected at each position of the examined sgRNA target sites that were mapped with at least 1,000 independent reads, and obvious base substitutions were observed only at the targeted base editing sites. Base substitution frequencies were calculated by dividing base substitution reads by total reads using CFBI pipeline (<https://github.com/YangLab/CFBI/>; v.1.0.0).

Whole-genome sequencing and data analysis. The step-by-step protocol for determining genome-wide off-target mutations is available at the Protocol Exchange⁴⁵. Genomic DNA was extracted from human 293FT cells and mouse livers/kidneys tissues using the FastPure cell DNA isolation kit (DC102-01, Vazyme). Indexed DNA libraries were prepared using the NEBNext Ultra II FS DNA Library Prep kit for Illumina. A total of 5.4 Tb WGS data of human 293FT cells and 2.9 Tb WGS data of mouse livers/kidneys tissues were obtained using the Illumina HiSeq X Ten (2 × 150) system at the CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China. The average coverage sequencing data generated for each 293FT sample and mouse tissue was 40× and 30×, respectively. After quality control using FastQC (v.0.11.5; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>; default parameters), WGS DNA-sequencing reads were trimmed using Trimmomatic (v.0.38; parameters: ILLUMINACLIP:TruSeq3-PE-2.6:2:20:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36)⁴⁶ to remove low-quality read sequences. The BWA-MEM algorithm (v.0.7.17-r1188; default parameters) was used to map clean reads of human 293FT cells to the human reference genome (GRCh38/hg38) and clean reads of mouse livers/kidneys tissues to mouse reference genome (CRCm38/mm10), respectively. Samtools (v.1.9; parameters: -bh -F 4 -q 30) was performed to select reads with a mapping quality score ≥ 30 and to sort BAM files. After marking duplicate reads identified by Picard (v.2.21.2; parameters: REMOVE_DUPLICATES=false) in the BAM file, GATK (v.4.1.3; default parameters)⁴⁷ was used to correct systematic bias using a two-stage process (BaseRecalibrator and ApplyBQSR; default parameters). To identify genome-wide de novo variants with high confidence, we conducted single-nucleotide variation calculation using three algorithms with the germline calling methods, GATK (v.4.1.3), LoFreq (v.2.1.3.1, default parameters)⁴⁸ and Strelka2 (v.2.9.10; default parameters)⁴⁹, separately. For GATK, genome-wide de novo variant calling was determined from the BAM file using three GATK (v.4.1.3) commands: HaplotypeCaller (default parameters), VariantRecalibrator (parameters: -an QD -an MQ -an MQRankSum -an ReadPosRankSum -an FS -an SOR -an DP--max-gaussians 4) and ApplyVQSR (parameters: -mode SNP -ts-filter-level 95). Notably, single-nucleotide variants (SNVs) that overlapped for the three algorithms were considered to be the reliable variants. To further obtain de novo SNVs, we filtered out the background variants in human 293FT cells or mouse livers/kidneys tissues, including: (1) for human 293FT cells, SNVs between the non-transfected 293FT cells of this study and that existed in the NCBI dbSNP (v.151; <http://www.ncbi.nlm.nih.gov/SNP/>) database; and, for mouse livers/kidneys tissues, SNVs that existed in the EMBL-EBI EVA (European Variation Archive release_1; <https://www.ebi.ac.uk/eva/>) database; (2) the overlapped SNVs with allele frequencies of less than 10% were removed from the following analysis; (3) variants overlapped with the UCSC repeat regions were removed; and (4) for human 293FT cells, further analyses focused solely on canonical (chromosomes 1–22, X, Y and M) chromosomes; for mouse livers/kidneys tissues, further analyses focused solely on canonical (chromosomes 1–19, X, Y and M) chromosomes. These WGS datasets were individually analysed using BEIDOU (<https://github.com/YangLab/BEIDOU/>; v.1.0.0) to call high-confidence base substitutions (SNVs) that could be identified by all three different callers—GATK, LoFreq and Strelka2.

Whole-transcriptome sequencing and RNA editing analysis. The step-by-step protocol for determining transcriptome-wide off-target mutations is available at the Protocol Exchange⁴⁵. Total RNAs were extracted from human 293FT cells and mouse livers tissues using the TransZol Up Plus RNA Kit (ER501-01, TransGen). Indexed RNA libraries were prepared using the TruSeq Stranded Total RNA with Ribo-Zero Globin. RNA-sequencing libraries were sequenced using the Illumina HiSeq X Ten (2 × 150) system at the CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China. RNA editing was analysed and visualized using the RADAR pipeline (<https://github.com/YangLab/RADAR/>; v.1.0.0)³³.

Statistics and reproducibility. All statistical analyses were performed using R (v.3.6.2; <http://www.R-project.org/>). P values were calculated using two-tailed Student's *t*-tests in this study. Experiments were performed three times independently unless indicated otherwise. The exact replication numbers are stated in figure legends. The experimental findings in all of the figures were reproduced successfully. No statistical method was used to predetermine sample size. No data

were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

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

Data availability

Deep-sequencing data, whole-transcriptome sequencing data and whole-genome sequencing data can be accessed in Gene Expression Omnibus under the accession codes [GSE164837](#) and [GSE164477](#), at the NCBI BioProject under the accession code [PRJNA692761](#) and in the National Omics Data Encyclopedia under the accession codes OEP001688, OEP001689 and OEP001690. All other data supporting the finding of this study are available from the corresponding authors on reasonable request. Source data are provided with this paper.

Code availability

The custom Perl and Shell scripts for calculating frequencies of base substitution and indels (CFBI) are available at GitHub (<https://github.com/YangLab/CFBI>). The computational pipeline of Base/Prime editor induced DNA off-target site identification unified toolkit (BEIDOU) to identify high-confidence base substitution in this paper is available at GitHub (<https://github.com/YangLab/BEIDOU>).

References

45. Wang, L. et al. Protocol for examining and eliminating base editor-induced genome-wide and transcriptome-wide off-target mutations. *Protocol Exchange* <https://doi.org/10.21203/rs.3.pex-1433/v1> (2021).
46. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
47. McKenna, N. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
48. 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).
49. Kim, S. et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat. Methods* **15**, 591–594 (2018).

Acknowledgements

We thank the staff at the Molecular and Cell Biology Core Facility, School of Life Science and Technology, ShanghaiTech University for providing experimental service. This work was supported by grants 2018YFA0801401 (to J.C. and H.Y.), 2019YFA0802804 (to L.Y. and B.Y.), 2018YFC1004602 (to J.C.), 2019YFA0802801 (to H.Y.) from MoST, 31925011 (to L.Y.), 91940306 (to L.Y.), 32070170 (to B.Y.), 32071442 (to H.Y.), 31871345 (to H.Y.), 31822016 (to J.C.), 81872305 (to J.C.), 31801073 (to W.X.) and 31972936 (to Y.Z.) from NSFC, Medical Science Advancement Program (to Basic Medical Sciences) of Wuhan University TFJC2018004 (to H.Y.), Applied Basic Frontier Program of Wuhan City (to H.Y.), the Youth Innovation Promotion Association from CAS (to W.X.), Fundamental Research Funds for the Central Universities (to H.Y. and Y.Z.), Hubei Health Commission Young Investigator Award (to H.Y.), National Postdoctoral Program for Innovative Talents (BX20190256 to H.Z.) and China Postdoctoral Science Foundation (2019M662704 to H.Z.).

Author contributions

J.C., L.Y., H.Y. and B.Y. conceived, designed and supervised the project. L.W., H.Z., R.G. and H.Q. performed most of the experiments with the help of L.Z., X.W., X.L., C.L., J. Wu and Q.C. on cell culture and plasmid construction. J.Wei prepared libraries for deep sequencing and W.X. and Y.-N.L. performed bioinformatics analyses, supervised by L.Y.; H.M., X.H., C.C and Y.Z. provided support with techniques. J.C., L.Y., H.Y. and B.Y. wrote the paper with input from the other authors. J.C. managed the project.

Competing interests

J.C., B.Y., L.Y., X.H. and L.W. have filed patent applications (PCT/CN2019/074577, PCT/CN2020/074218) relating to the published work through ShanghaiTech University.

Additional information

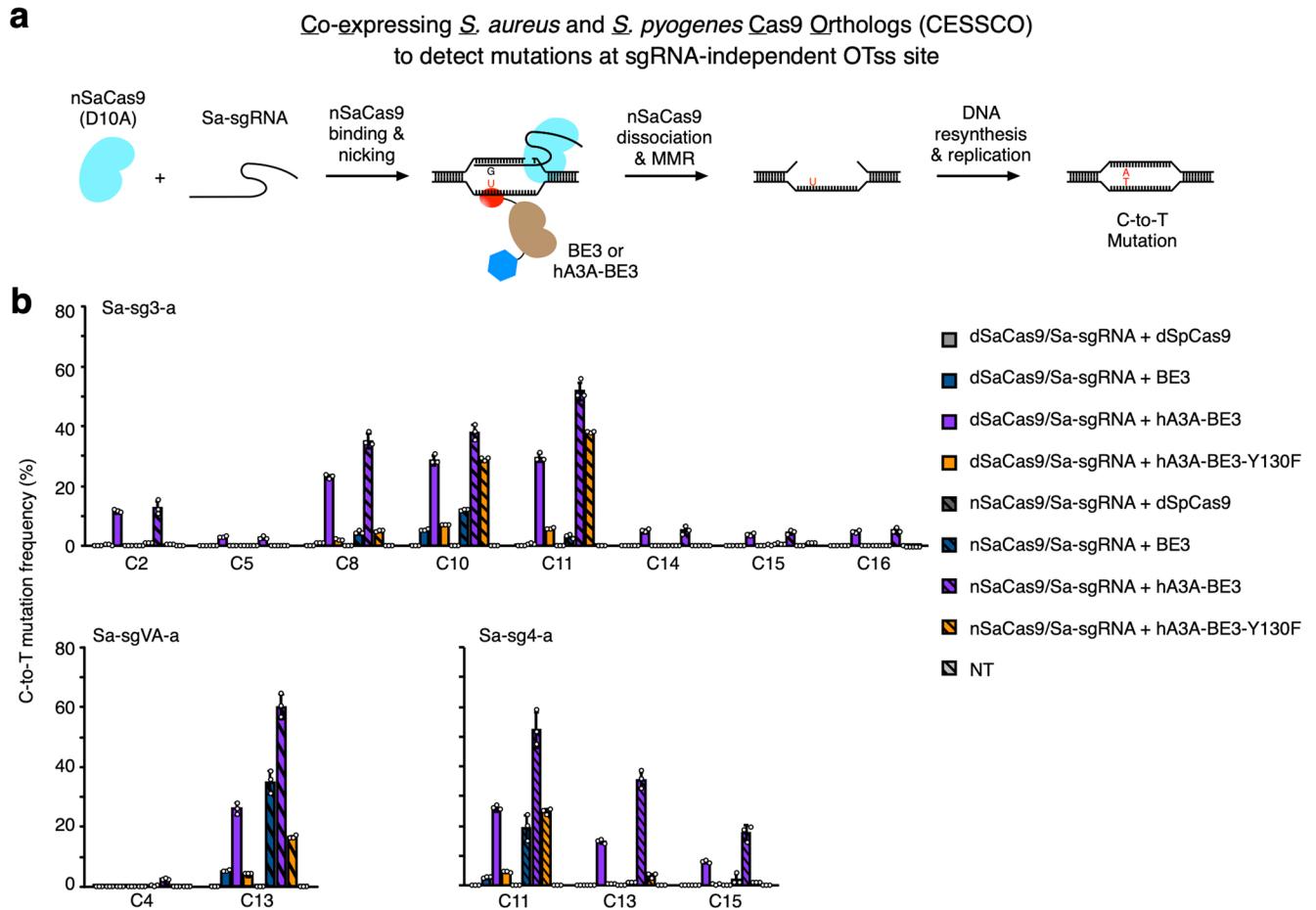
Extended data is available for this paper at <https://doi.org/10.1038/s41556-021-00671-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-021-00671-4>.

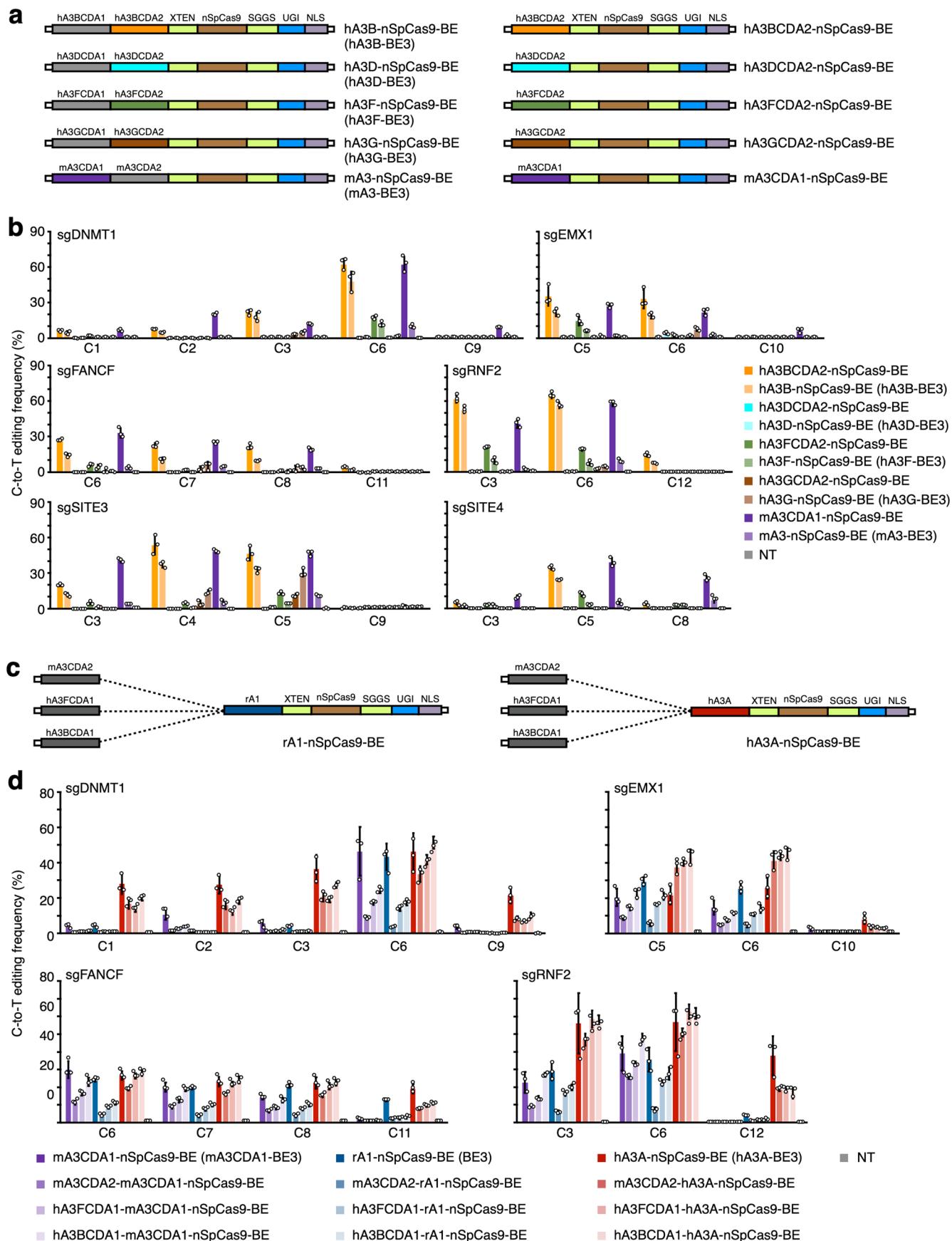
Correspondence and requests for materials should be addressed to B.Y., H.Y., L.Y. or J.C.

Peer review information *Nature Cell Biology* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

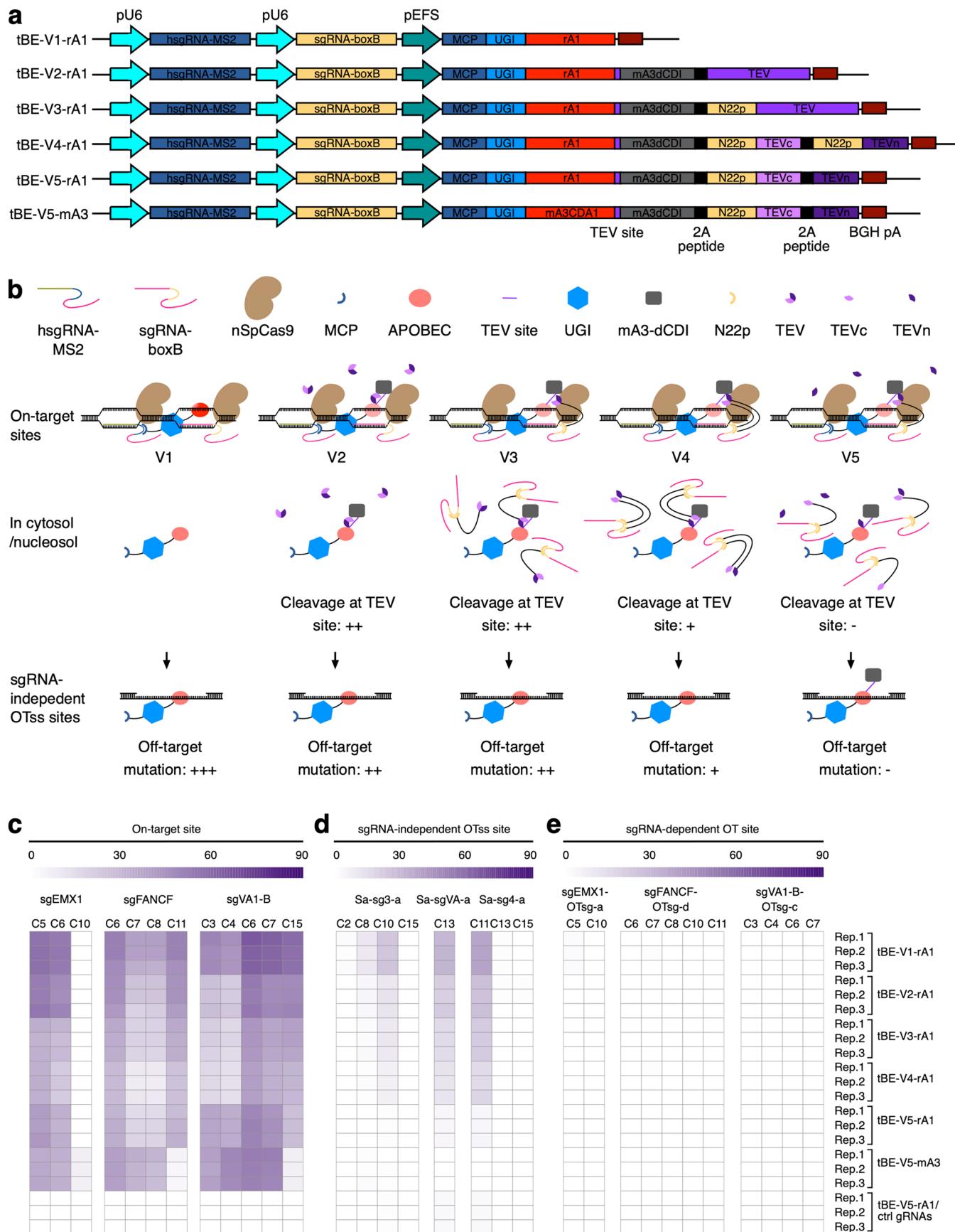


Extended Data Fig. 1 | Development and application of CESSCO to detect the sgRNA-independent mutations induced by BE3 at OTss sites. **a**, Schematic diagrams illustrate the co-expressing *S. aureus* and *S. pyogenes* Cas9 orthologs (CESSCO) method, which expresses two Cas9 orthologs, to detect sgRNA-independent mutations induced by BEs at OTss sites. **b**, C-to-T mutation frequencies induced by dSpCas9, BE3, hA3A-BE3 and hA3A-BE3-Y130F at the indicated OTss sites triggered by Sa-sgRNA/dSaCas9 or Sa-sgRNA/nSaCas9 pairs. The data of C-to-T mutation frequencies induced by hA3A-BE3-Y130F at the indicated OTss sites triggered by Sa-sgRNA/nSaCas9 pairs are same as the ones shown in Fig. 2b. Data are presented as mean ± s.d. from three independent experiments. NT, non-transfected control. Numerical source data for **b** are provided.



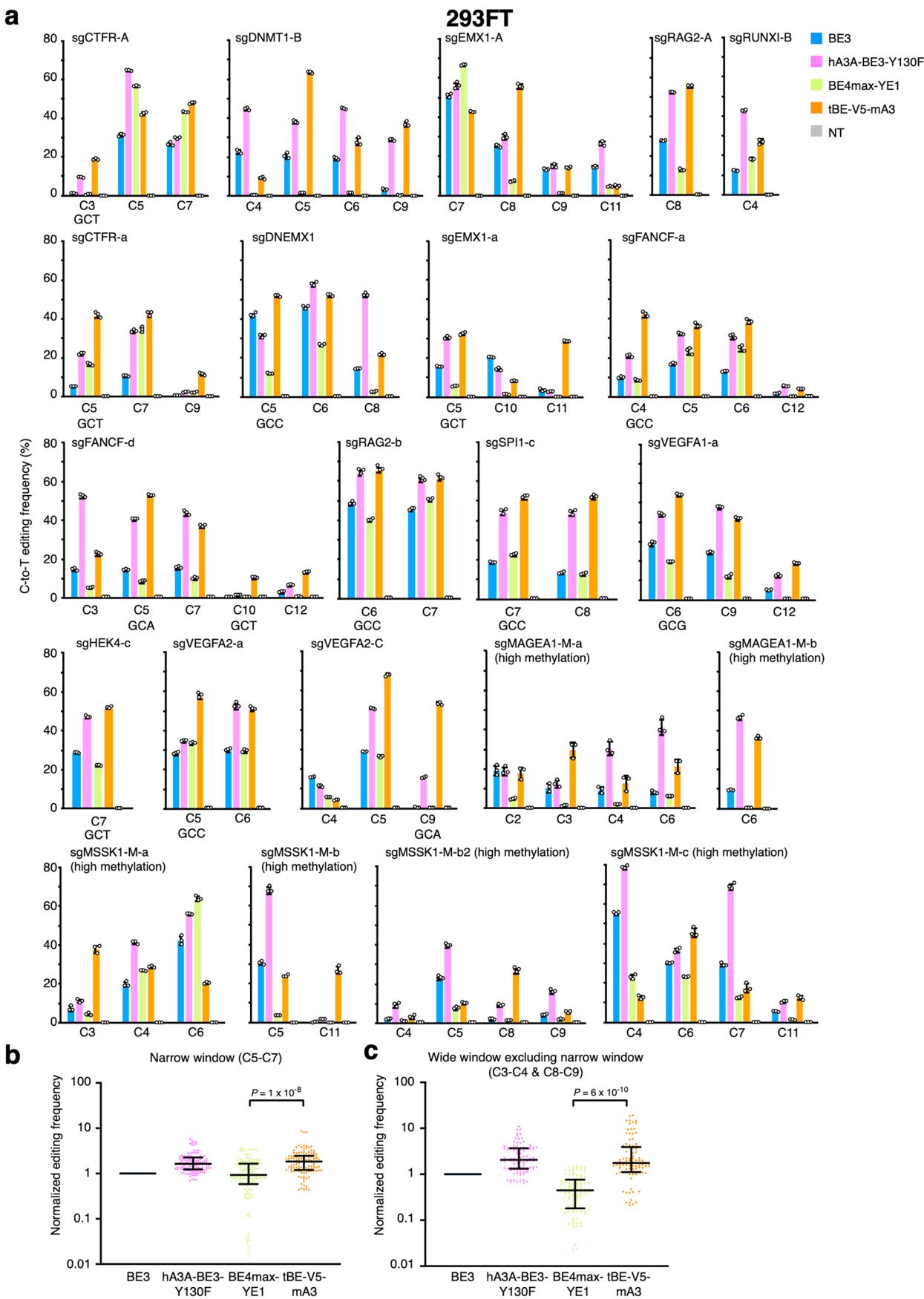
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Characterization of dCDI domains. **a**, Schematic diagrams illustrate base editors constructed by fusing the indicated CDA domains to nSpCas9 and uracil DNA glycosylase inhibitor (UGI). The regulatory CDA domains are in grey shadow and the active CDA domains are in colors. NLS, nuclear localization sequence; XTEN and SGGS, linker peptides. **b**, C-to-T editing frequencies induced by the indicated BEs at six genomic loci. **c**, Schematic diagrams illustrate the fusion of different dCDI domains to the N-terminus of BE3 and hA3A-BE3. **d**, C-to-T editing frequencies induced by the indicated BEs at four genomic loci. Data in **b** and **d** are presented as mean \pm s.d. from three independent experiments. Numerical source data for **b** and **d** are provided.



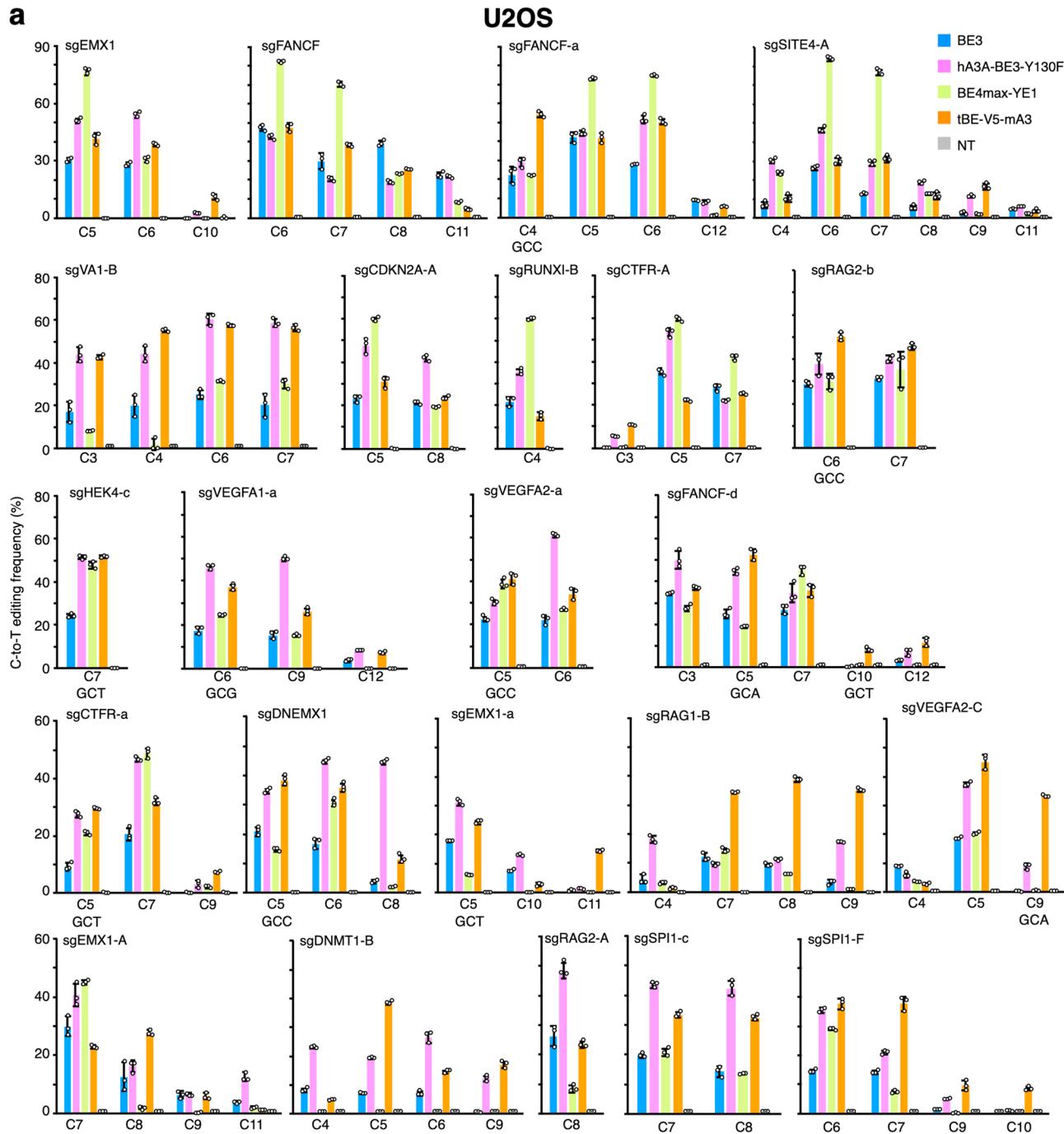
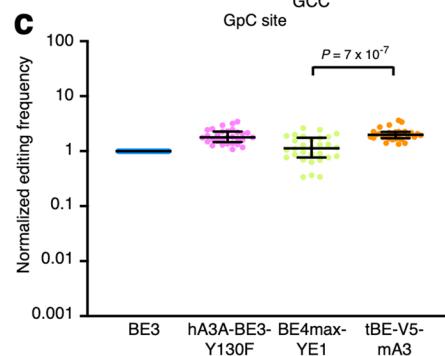
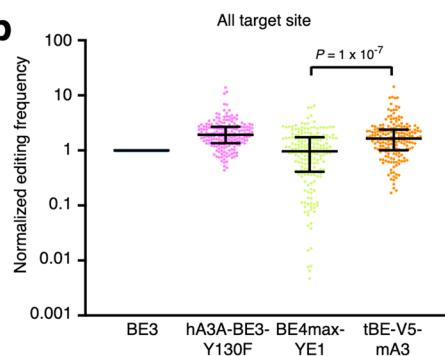
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Versions of tBE and their performance in cells. **a**, Schematic diagrams illustrate the construction and development of various versions of tBEs by using different strategies to cleave mA3dCDI off. **b**, The interaction of molecular components in different versions of tBEs. Due to free diffusion, the dCDI domain could be cleaved off from APOBEC through a two-component interaction of the TEV site and a free TEV protease (V2), a N22p-fused TEV protease (V3) or a TEV protease reconstituted by an sgRNA-boxB (V4) in cytosol/nucleosol, and then the resulted MCP-UGI-APOBEC fusion protein triggers mutations at sgRNA-independent OTss sites. In the version 5 (V5) of tBE, the dCDI is unlikely to be cleaved off from APOBEC in cytosol/nucleosol as the chance for the free diffusion induced three-component interaction of TEV site, TEVn and N22p-TEVc is very low. **c-e**, Comparison of the editing or mutation frequencies induced by different versions of tBEs at on-target sites (**c**), sgRNA-independent OTss sites (**d**) and sgRNA-dependent OT sites (**e**). Numerical source data for **c-e** are provided.



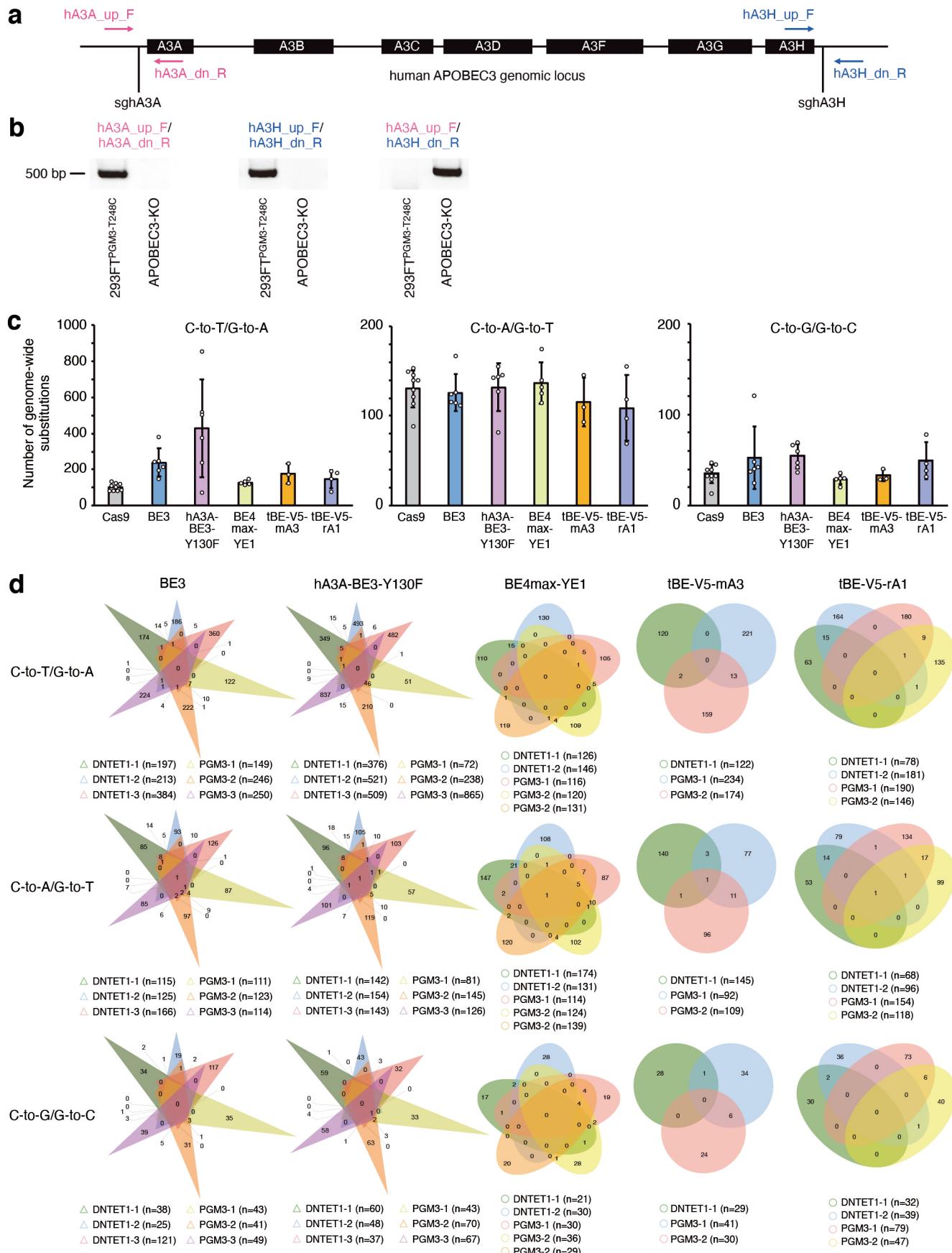
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Comparing editing efficiencies of tBE-V5-mA3 with other BEs in 293FT cells. **a**, Editing frequencies induced by indicated BEs at 22 genomic loci, including 6 on-target sites with natively high DNA methylation levels and 10 on-target sites containing GpC dinucleotides. Data are presented as mean \pm s.d. from three independent experiments. **b**, Statistical analysis of normalized editing frequencies within the narrow window shown in **a**, Fig. 2a,d. $n=123$ edited cytosines at 27 on-target sites from three independent experiments. **c**, Statistical analysis of normalized editing frequencies within the wide window excluding the narrow window shown in **a**, Figs. 2a,d. $n=96$ edited cytosines at 22 on-target sites from three independent experiments. **b-c**, Setting the editing frequencies induced by the BE3 as 100%. P value, two-tailed Student's t test. The median and interquartile range (IQR) are shown. Numerical source data for **a-c** are provided.

a**b**

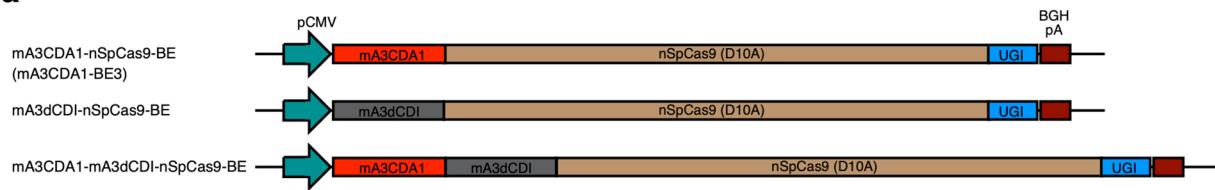
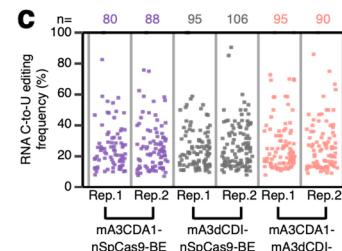
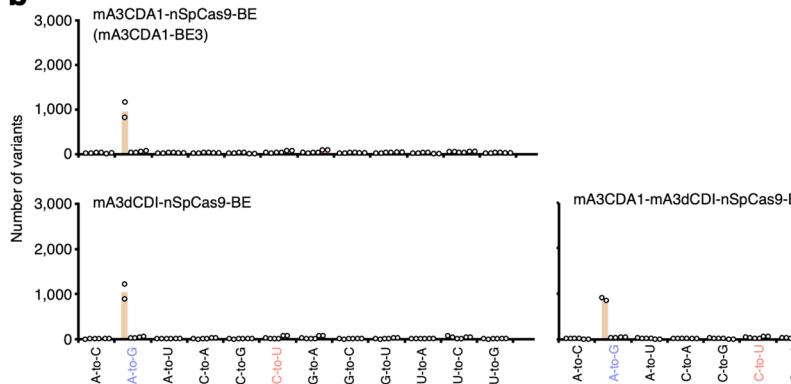
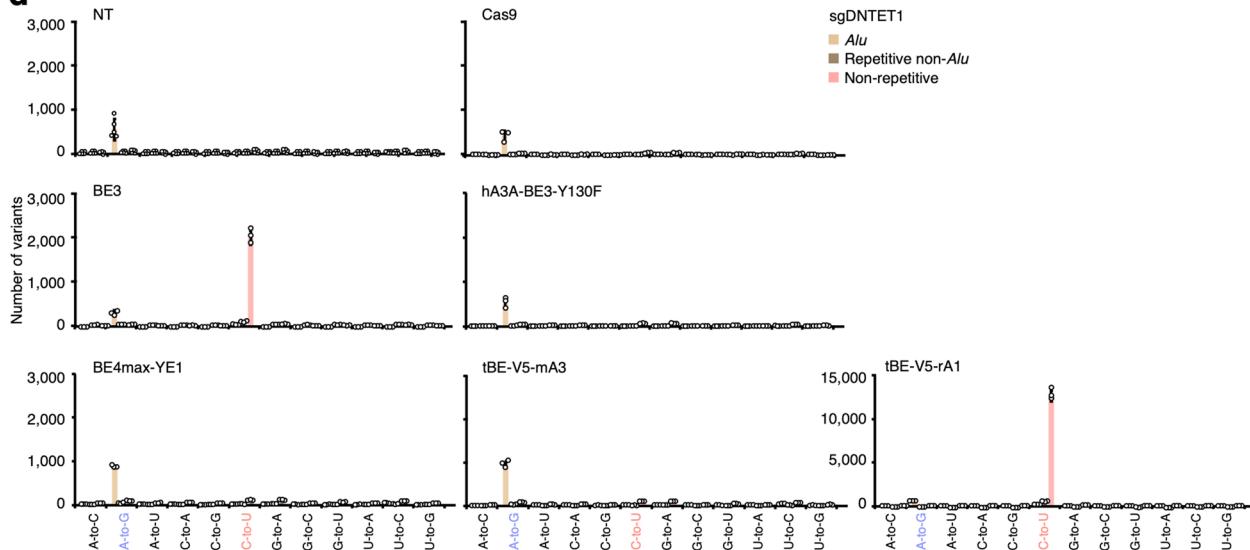
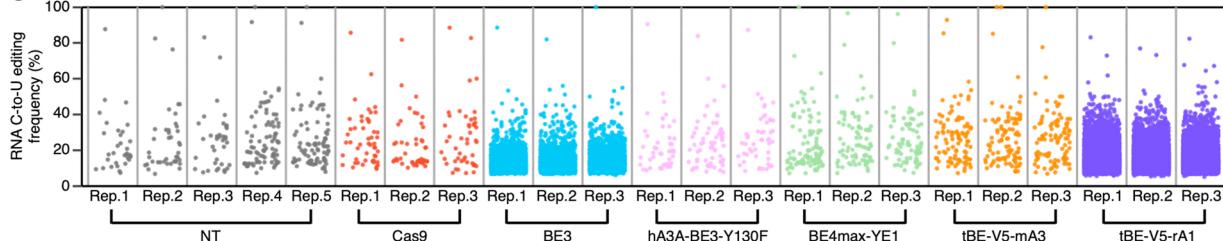
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Comparing editing efficiencies of tBE-V5-mA3 with other BEs in U2OS cells. **a**, Editing frequencies induced by indicated BEs at 23 genomic loci. Data are presented as mean \pm s.d. from three independent experiments. **b**, Statistical analysis of normalized editing frequencies at all 23 on-target sites shown in **a**. n=189 edited cytosines at 23 on-target sites from three independent experiments. **c**, Statistical analysis of normalized editing frequencies at 10 on-target sites containing GpC dinucleotides shown in **a**. n=30 edited cytosines at 10 on-target sites from three independent experiments. **b-c**, Setting the editing frequencies induced by the BE3 as 100%. P value, two-tailed Student's t test. The median and interquartile range (IQR) are shown. Numerical source data for **a-c** are provided.

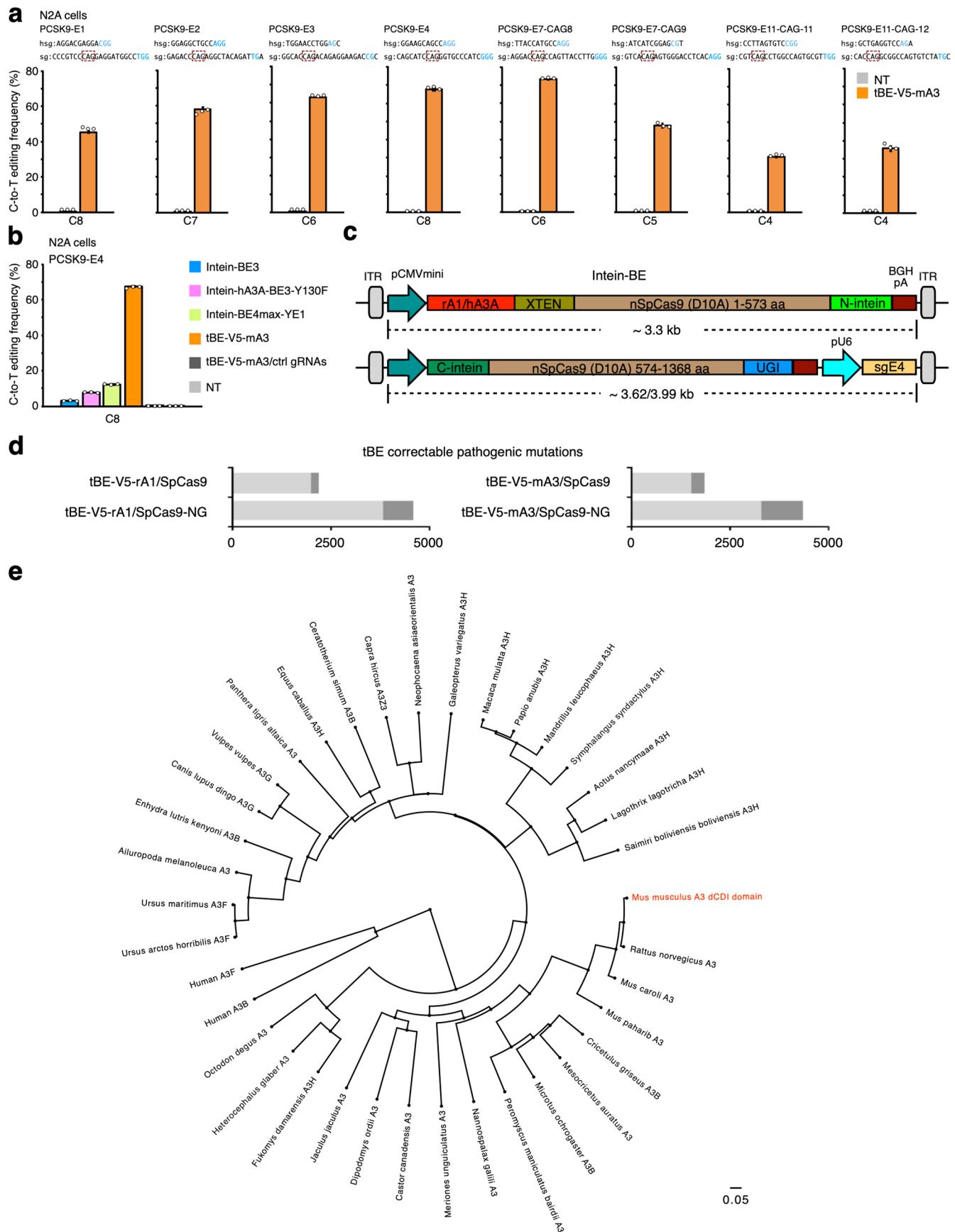


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Determining genome-wide OT mutations induced by various BEs. **a**, Schematic diagrams illustrate the position of sgRNAs for the knockout of human APOBEC3 cluster and primers for PCR detection. **b**, The knockout of APOBEC3 cluster was confirmed by genomic DNA PCR. Agarose gel data are representative of three independent experiments. **c**, The numbers of genome-wide base substitutions on C or G in the single cell clones treated with Cas9 or indicated BEs. Data are presented as mean \pm s.d. from n=9 (Cas9), n=6 (BE3, hA3A-BE3-Y130F), n=5 (BE4max-YE1), n=3 (tBE-V5-mA3) or n=4 (tBE-V5-rA1) single cell colonies. **d**, Rare overlap was found for the genome-wide base substitutions among different clones treated with indicated BEs. Uncropped gels for **b** and numerical source data for **c** and **d** are provided.

a**b****d****e**

Extended Data Fig. 7 | Determining transcriptome-wide OT mutations induced by various BEs. **a**, Schematic diagrams illustrate the construction of mA3CDA1-nSpCas9-BE (mA3CDA1-BE3), mA3dCDI-nSpCas9-BE and mA3CDA1-mA3dCDI-nSpCas9-BE. **b**, Histograms show the numbers of all twelve types of RNA editing in different genome regions in the cells treated with mA3CDA1-nSpCas9-BE, mA3dCDI-nSpCas9-BE and mA3CDA1-mA3dCDI-nSpCas9-BE. Data are presented as mean from two independent experiments. **c**, Manhattan plot of RNA OT editing (C-to-U) frequency shown in **b**. **d**, Histograms show the numbers of all twelve types of RNA editing in different genomic regions in the cells treated with indicated BEs. Data are presented as mean \pm s.d. from three (five for NT) independent experiments. **e**, Manhattan plot of RNA OT editing (C-to-U) frequency shown in **d**. Numerical source data for **b** and **d** are provided.



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

Extended Data Fig. 8 | Enriching tBE system by the combination of different Cas9 orthologs and dCDIs. **a**, Editing efficiencies induced by tBE-V5-mA3 and *Pcsk9*-targeting sgRNAs pairs at eight premature stop codon sites in six different exons of *Pcsk9* in N2A cells. **b**, Comparison of editing efficiencies of indicated BEs before dual-AAV8 delivery in N2A cells. **c**, Schematic diagrams illustrate the dual-AAV system for intein-BEs delivery. **d**, Numbers of tBE correctable pathogenic mutations when combining with Cas9 or Cas9-NG. The gray portions show the number of preferentially correctable pathogenic mutations. A preferentially correctable pathogenic mutation is the only edited cytosine in the editing window of indicated tBE. **e**, The phylogenetic tree of mA3dCDI domain. Data in **a** and **b** are presented as mean \pm s.d. from three independent experiments. Numerical source data for **a,b,d** are provided.

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	No software was used for data collection.
Data analysis	High-throughput sequencing read quality was evaluated by FastQC (v0.11.5) and then trimmed by Trimmomatic (v0.38) to remove low quality read sequence. BWA-MEM algorithm (0.7.17-r1188) was used to map clean reads of human 293FT cells to the human reference genome (GRCh38/hg38), and clean reads of mouse livers/kidneys tissues to mouse reference genome (CRCm38/mm10), respectively. Duplicate reads were marked by Picard (v2.21.2). Genome-wide de novo variants were identified by BEIDOU (https://github.com/YangLab/BEIDOU , v1.0.0) to call high-confidence base substitution (SNVs) that could be identified by all three different callers, GATK (v4.1.3), LoFreq (v2.1.3.1) and Strelka2 (v2.9.10). RNA editing was analyzed and visualized by RADAR pipeline (https://github.com/YangLab/RADAR , v1.0.0). Substitution calling was performed by SAMtools (v1.9) mpileup function and indel calling was based on the CIGAR values of the SAM/BAM files (Li et al, Nat Biotechnol, 2018). All statistical analyses were performed with R (v3.6.2, http://www.R-project.org). The custome Perl and Shell scripts for calculating frequencies of base subsutition and indels (CFBI) are available at GitHub (https://github.com/YangLab/CFBI , v1.0.0).

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

Deep sequencing data, whole transcriptome sequencing data and whole genome sequencing data can be accessed in Gene Expression Omnibus under the accession codes GSE164837 and GSE164477, in NCBI BioProject under the accession code PRJNA692761 and in National Omics Data Encyclopedia under the accession codes OEP001688, OEP001689 and OEP001690. Source data are provided with this study. All other data supporting the finding of this study are available from the corresponding author on reasonable request.

Human and mouse genome sequences were downloaded from UCSC Genome Browser (<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz>, <https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/mm10.fa.gz>). The single nucleotide variants (SNVs) information for GATK BaseRecalibrator of human was downloaded from NCBI dbSNP database (ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606_b151_GRCh38p7/VCF/GATK/All_20180418.vcf.gz); and SNVs for GATK BaseRecalibrator of mouse was downloaded from EMBL-EBI EVA (ftp://ftp.ebi.ac.uk/pub/databases/eva/rs_releases/release_1/by_species/Mouse_10090/GRCm38.p4/GCA_000001635.6_current_ids.vcf.gz). Human and mouse repeat regions were downloaded from UCSC Genome Browser (<http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz> and <http://hgdownload.soe.ucsc.edu/goldenPath/mm10/database/rmsk.txt.gz>). GATK VariantRecalibrator for human including multiple VCF files (https://console.cloud.google.com/storage/browser/_details/genomics-public-data/resources/broad/hg38/v0/hapmap_3.3.hg38.vcf.gz, https://console.cloud.google.com/storage/browser/_details/genomics-public-data/resources/broad/hg38/v0/1000G_omni2.5.hg38.vcf.gz, https://console.cloud.google.com/storage/browser/_details/genomics-public-data/resources/broad/hg38/v0/1000G_phase1.snp.high_confidence.hg38.vcf.gz, https://ftp.ncbi.nih.gov/snp/organisms/human_9606/VCF/00-All.vcf.gz and https://console.cloud.google.com/storage/browser/_details/genomics-public-data/resources/broad/hg38/v0/Mills_and_1000G_gold_standard.indels.hg38.vcf.gz). GATK VariantRecalibrator for mouse including VCF file (http://ftp.ebi.ac.uk/pub/databases/eva/rs_releases/release_1/by_species/Mouse_10090/GRCm38.p4/GCA_000001635.6_current_ids.vcf.gz and <http://hgdownload.soe.ucsc.edu/goldenPath/mm10/database/rmsk.txt.gz>). Files used to filter out the background variants for human (https://ftp.ncbi.nih.gov/snp/organisms/human_9606/VCF/00-All.vcf.gz and <http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz>). Files used to filter out the background variants for mouse (http://ftp.ebi.ac.uk/pub/databases/eva/rs_releases/release_1/by_species/Mouse_10090/GRCm38.p4/GCA_000001635.6_current_ids.vcf.gz and <http://hgdownload.soe.ucsc.edu/goldenPath/mm10/database/rmsk.txt.gz>).

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. In previous studies using related experiments, the sample size has been determined to be sufficient to ensure reproducibility. All sample size numbers were listed in the corresponding figure legends. Sample sizes for these experiments were chosen based upon field standards and prior knowledge of experimental variation.
Data exclusions	No data were excluded.
Replication	Experiments were performed three times independently unless indicated. The exact replication numbers were stated in figure legends. The experimental findings in all figures were reproduced successfully.
Randomization	Samples were not randomized. The experimental work-flow in this study did not allow/need randomization.
Blinding	The investigators were not blinded to group allocation. Blinding is not relevant to our study because it is not a subjective trial and the results presented here are based on objective description of our technology.

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

Antibodies

Antibodies used

The following antibodies were used for western blot: Anti-alpha-tubulin at 0.5 µg/ml (mouse monoclonal DM1A, Sigma, Cat#T6199), Anti-2A peptide at 0.5 µg/ml (mouse monoclonal 3H4, Sigma, Cat#MABS2005). The following secondary antibody was used: Goat anti-Mouse IgG-HRP conjugate (1:5000, Cat#115-035-146) from Jackson ImmunoResearch.

Validation

The primary antibody Anti-alpha-tubulin against human α-tubulin has been validated for western blot analysis by manufacturer. The primary antibody Anti-2A peptide against T2A peptide from those asigma virus 2A has been validated for western blot analysis by manufacturer.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

293FT (R70007) was from Thermo Fisher Scientific. HEK293T/17 (ATCC® CRL-11268™), U2OS (ATCC® HTB-96) and N2A (ATCC® CCL-131™) cells were from ATCC.

Authentication

No cell line was authenticated.

Mycoplasma contamination

Cell lines have been tested negative for mycoplasma contamination by PCR methods.

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell line was used.

Animals and other organisms

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

Laboratory animals

C57BL/6, female, 8 weeks old. Mice were given free access to food and water, and were maintained under a 12/12 h light–dark cycle with controlled temperature (20–25°C) and humidity (50 ± 10%).

Wild animals

No wild animals were used in the study.

Field-collected samples

No field-collected samples were used in the study.

Ethics oversight

All mouse studies were approved by the Animal Care and Ethical Committee at Wuhan University.

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