nature chemical biology

Article

https://doi.org/10.1038/s41589-024-01595-4

Engineering APOBEC3A deaminase for highly accurate and efficient base editing

Received: 16 May 2023

Accepted: 29 February 2024

Published online: 29 March 2024

Check for updates

Lei Yang[1,7], Yanan Huo[1,7], Man Wang[1,7], Dan Zhang[1,7], Tianai Zhang[1], Hao Wu[2], Xichen Rao[2], Haowei Meng[2], Shuming Yin[1], Jiale Mei[1], Dexin Zhang[3], Xi Chen[4], Jia Lv[1], Meizhen Liu[1], Yiyun Cheng[1], Yuting Guan[1], Bo Feng[5], Gaojie Song[1], Chengqi Yi[2], Mingyao Liu[1,4], Fanyi Zeng[6] ✉, Liren Wang[1] ✉ & Dali Li[1] ✉

Cytosine base editors (CBEs) are effective tools for introducing C-to-T base conversions, but their clinical applications are limited by off-target and bystander effects. Through structure-guided engineering of human APOBEC3A (A3A) deaminase, we developed highly accurate A3A-CBE (haA3A-CBE) variants that efficiently generate C-to-T conversion with a narrow editing window and near-background level of DNA and RNA off-target activity, irrespective of methylation status and sequence context. The engineered deaminase domains are compatible with PAM-relaxed SpCas9-NG variant, enabling accurate correction of pathogenic mutations in homopolymeric cytosine sites through flexible positioning of the single-guide RNAs. Dual adeno-associated virus delivery of one haA3A-CBE variant to a mouse model of tyrosinemia induced up to 58.1% editing in liver tissues with minimal bystander editing, which was further reduced through single dose of lipid nanoparticle-based messenger RNA delivery of haA3A-CBEs. These results highlight the tremendous promise of haA3A-CBEs for precise genome editing to treat human diseases.

Base editing is a revolutionary technology that enables efficient conversion of one base pair to another without the need for double-stranded DNA breaks or donor templates, in both dividing and nondividing cells[1]. There are two primary types of base editor: adenine base editors (ABEs) and cytosine base editors (CBEs), which induce highly efficient A-to-G and C-to-T conversions, respectively. While ABEs typically catalyze base conversions with very high product purity (>99.9%)[2], CBEs generate uracil as a result of cytosine deamination, which in turn stimulates uracil DNA glycosylase to generate abasic sites, activating cellular base excision repair pathway to convert cytosines to thymidines and other nucleotides[3]. To enhance the product purity of CBEs, uracil glycosylase inhibitor (UGI) is used to fuse with nCas9 and various cytidine deaminase domains to develop CBEs with high C•G-to-T•A product purity[3]. Additionally, several groups have also developed C-to-G base editors that efficiently generate C•G-to-G•C transversions using uracil DNA glycosylase to substitute for UGI in CBEs or solely using evolved cytidine deaminase variants in CBE architectures[4,5]. These base editors have been widely used in genetic engineering in animals[6,7], plants[8,9] and have shown promising potential in gene therapy[10–12]. However, several studies have reported that the deaminase domain of CBEs induced Cas9-independent DNA[13,14] and RNA[15,16] off-target mutations, which may hinder broad applications of CBEs for precise editing, especially

[1]Shanghai 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. [2]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. [3]Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [4]BRL Medicine Inc., Shanghai, China. [5]School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China. [6]Department of Histo-Embryology, Genetics and Developmental Biology, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [7]These authors contributed equally: Lei Yang, Yanan Huo, Man Wang, Dan Zhang. ✉e-mail: fzeng@vip.163.com; lrwang@bio.ecnu.edu.cn; dlli@bio.ecnu.edu.cn

in clinical settings[17]. Additionally, bystander effects induced by undesired base conversions at the target loci may also pose challenges for precise editing.

Several strategies have been implemented to improve the accuracy of base editors, including narrowing editing window[18] or recruiting deaminase variants with defined sequence contexts[19,20]. Truncation of deaminase CDA1 or shortening the linker between nCas9 and deaminase domain can moderately narrow the editing window[18]. Furthermore, engineering of some deaminases that preferentially deaminate cytosine in specific sequence contexts has led to the development of eA3A-CBE (ref. 19) and A3G-CBE (ref. 20), which catalyze C-to-T conversion in TC and CC motifs, respectively. Especially for eA3A-CBE, it showed greatly reduced bystander edits, high efficiency and minimal off-target effects due to restricted motif preference. Protein engineering of rAPOBEC1 deaminase has yielded CBE variants such as YE1 (ref. 21), FE1 (ref. 22) and R33A (ref. 15) that demonstrate a narrow editing window and dramatically reduced off-target effects. The transformer base editors have minimal deaminase-induced off-target effects[23], but this system is complicated for application and bystander editing effects remain unimproved.

rAPOBEC1-based CBEs are inefficient in a methylated genomic loci or GC contexts (GpC dinucleotides) due to the inherent feature of the original rAPOBEC1 deaminase[3]. By contrast, A3A-CBEs are highly efficient regardless of the methylation status and sequence context of the target cytosines[24]. However, they exhibit a wider editing window and much higher off-target effects compared to other CBE variants[25]. To overcome these issues, we aimed to engineer A3A deaminase to obtain efficient CBEs with high accuracy and a narrow editing window, regardless of methylation status of the target loci. Through structure-guided molecular engineering, we have developed three potent haA3A-CBEs that efficiently catalyze C-to-T conversion within a condensed editing window in various genomic loci, including highly methylated regions or GC context, with minimal Cas9-independent DNA and RNA off-target editing. Furthermore, we have demonstrated that the engineered haA3A-CBEs can precisely correct human pathogenic single-nucleotide variants (SNVs) in cell lines, especially in homopolymeric cytosine sites. haA3A-CBEs induced efficient and selective correction of a desired mutation among two adjacent cytosines in a hereditary tyrosinemia type 1 (HT1) mouse model through viral vector or lipid nanoparticle (LNP)–mRNA mediated delivery, suggesting their promising potential for precise editing to treat human diseases.

Results

Structure-guided molecular engineering of A3A

To improve the precision of A3A-CBE without the requirement of specific sequence context, we aimed to engineer the A3A deaminase based on its structure. The structure suggests that a deep groove formed by loop 1, 3, 5 and 7 plays an important role in accommodating single-stranded DNA (ssDNA) substrate[26,27] (Fig. 1a). Several residues in these loops interact with either the bases or the backbone of the substrate. For instance, the H29 residue in loop 1 forms hydrogen bonds to the backbone phosphates of the target C (dC_0) and the nucleotide at the −1 and +1 positions. Additionally, N57 from loop 3 and S99 from loop 5 form a hydrogen bond to O3′ atom and the amino group of dC_0, respectively, while Y130 from loop 7 forms a π–π interaction with the pyrimidine ring and a hydrogen bond with the 5′ phosphate of dC_0 (Fig. 1a). Beyond that, the W98, Y130, D131 and Y132 residues influence the −1 preference[26,27]. Guided by this structural indication, we selected 13 residues from these four loops and generated 36 mutants bearing single amino acid substitutions with changed side chain size, polarity or hydrophilic–hydrophobic properties (Fig. 1b).

Our aim was to select mutants that exhibited good on-target efficiency and dramatically reduced Cas9-independent off-target editing, which could induce unpredictable safety issues for clinical applications. To achieve this, we improved the R-loop assay[25] using SaCas9 nickase (nSaCas9) instead of catalytically dead SaCas9 (dSaCas9), which dramatically enhanced the sensitivity for evaluation of Cas9-independent DNA off-target editing (Extended Data Fig. 1a–g). We evaluated the efficiency and accuracy of the 36 mutants using the 'BE4max' architecture[28] at the on-target ABL1 site 1 and the off-target Sa site 5, which contained multiple cytosines in various sequence contexts (Fig. 1b and Extended Data Fig. 1a). High-throughput sequencing (HTS) data showed that intact A3A efficiently edited all six cytosines from the C_3–C_{11} position in all four sequence contexts at the ABL1 site 1 and induced numerous off-target edits in Sa site 5, while YE1 exhibited a much narrower editing window and background level of off-target effect, which was consistent with previous studies[25]. Some variants (T31D, W98A, Y130A, D131P, D133N, D133P) exhibited dramatically reduced off-target effects but retained or even slightly higher on-target activity compared to the intact A3A (Fig. 1b and Extended Data Fig. 1h). Notably, the Y130A mutant, which was previously reported as inactive in deaminating a 5′-TC-containing ssDNA substrate[26], showed preferential editing on a single cytosine within this homopolymeric cytosine site, and the D131P mutant showed efficient cytosine editing within all sequence contexts, suggesting they have different editing pattern from YE1, eA3A and other variants mentioned above (Fig. 1b). We also observed that the variants containing Y130F, D131Y or Y132D, which were previously reported with a narrowed editing window and reduced RNA off-target activity[16,24], showed moderate reduction of Cas9-independent off-target effect (Fig. 1b). These data suggest that engineering of A3A deaminase has the potential to narrow the editing window of A3A-BE4max and reduce its off-target activity.

A3A variants reduced Cas9-independent off-target activity

To further evaluate the editing performance of Y130A and D131P mutants, we conducted on-target analysis at ten endogenous sites and R-loop assays at six orthogonal SaCas9 sites. We analyzed the HTS data and observed that both Y130A and D131P variants displayed high on-target editing efficiency and substantially reduced Cas9-independent off-target editing compared to intact A3A (Fig. 1c,d and Supplementary Fig. 1a,b). Specifically, the Y130A variant demonstrated an on-target editing efficiency ranging from 62.3 to 87.7% (average of 71.5%) on the most efficiently edited cytosine of each site with a relatively narrow editing window (C_4–C_9) (Fig. 1c,e). On the other hand, the off-target editing efficiency of Y130A was generally below 1%, with the exception of AC_6 at Sa site 2 and CC_{11} at Sa site 5 (Fig. 1d). The D131P variant, while exhibiting comparable or higher on-target editing efficiency (average of 81.9%, ranging from 74.4 to 87.8%) compared to intact A3A (average of 76.8%, ranging from 58.2 to 82.0%), had a narrower editing window (Fig. 1c,e). It induced low levels of off-target editing, which were higher than those observed for YE1, eA3A and Y130A variants, as determined by improved R-loop assays (Fig. 1d). The two A3A mutants as well as eA3A showed low levels of indel formation, except at the PPP1R12C site 7 (Supplementary Fig. 1c). Based on these data, we found that the Y130A variant exhibited high efficiency with substantially reduced off-target editing activity and a narrow editing window.

Development of haA3A-CBE by further engineering of A3A

The data obtained from the Y130A variant suggest that Y130 residue is critical for A3A activity. Thus, variants of all other amino acid substitutions on this residue were generated (Fig. 2a). Moreover, additional variants based on Y130A that combined individual point mutations (A71V, I96V and I96T mutations identified in the study of eA3A-CBE with decreased bystander editing[19], and D131P, D131F, D131R and D131Y mutations identified above that reduced Cas9-independent off-target activity while retaining high editing efficiency) were also generated (Fig. 2a). After testing these variants at two on-target sites, we found that Y130G, Y130P, I96T/Y130A (TA) and I96V/Y130A (VA) variants had high editing efficiencies at CC_6 of EMX1 site 1 and AC_7 of PD1 site 1, with

https://doi.org/10.1038/s41589-024-01595-4

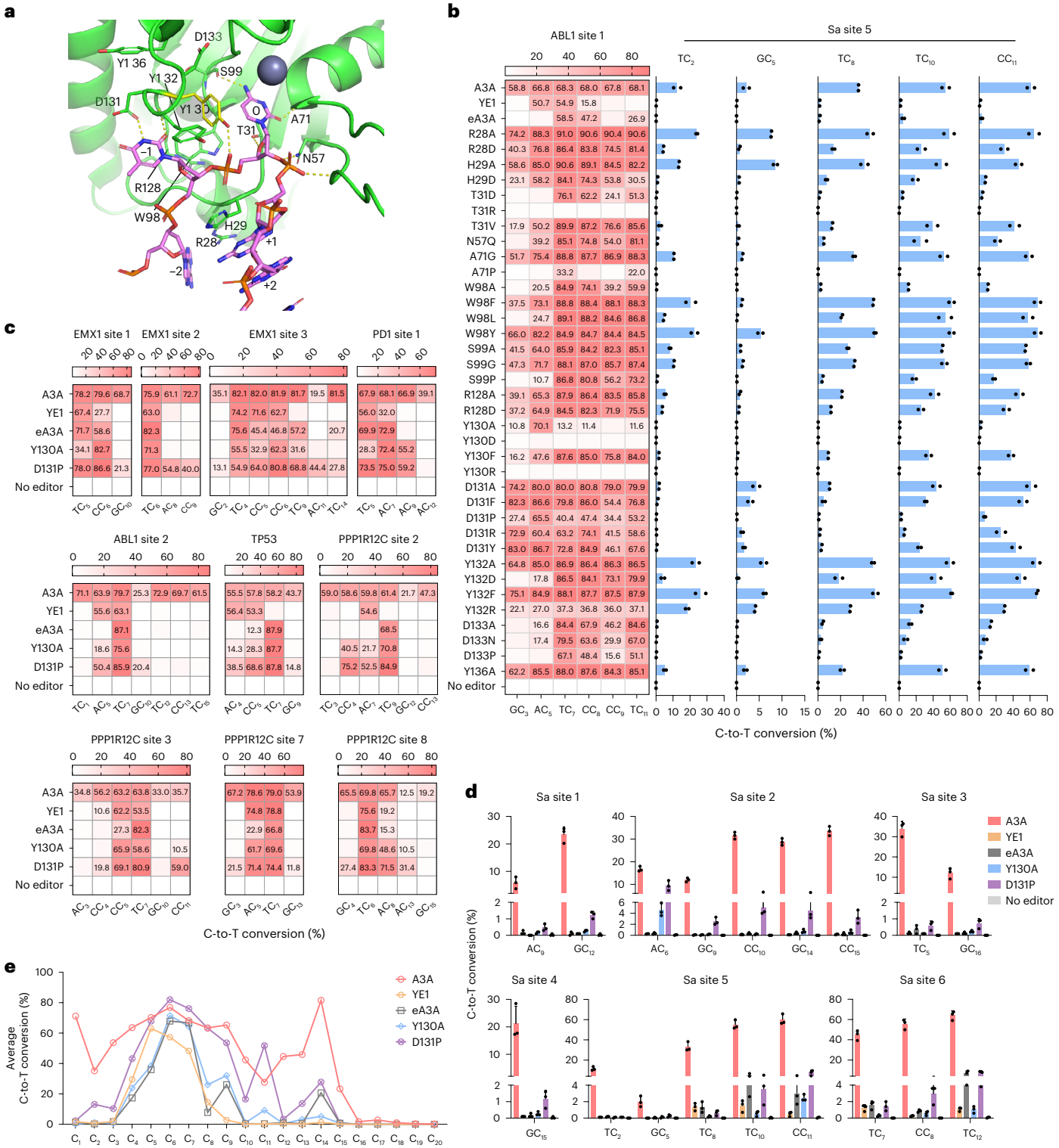


Fig. 1 | Structure-based engineering of A3A. a, Human A3A (PDB code 5sww) in complex with ssDNA was used as a template to elucidate the rationale for mutations. Amino acids spatially contacting or adjacent to the substrate DNA are labeled. The editing base is labeled 0 and the bases before and after it are labeled −1 and −2, and +1 and +2, respectively. Hydrogen bonds are shown as yellow dashed lines. **b**, Heatmap showing on-target editing frequencies for A3A-BE4max variants at ABL1 site 1. Histograms showing Cas9-independent off-target editing frequencies for A3A-BE4max variants at Sa site 5 from two independent biological replicates. **c**, Heatmaps showing editing frequencies for A3A-BE4max variants at ten genomic loci. **d**, Cas9-independent off-target editing frequencies for A3A-BE4max variants at six SaCas9 loci. **e**, Average editing frequencies for A3A-BE4max variants at each protospacer position from ten endogenous loci shown in **c**. The editing window is defined as the protospacer positions for which average editing efficiency is ≥30% of the average editing at the maximally edited position. In heatmaps, editing efficiencies shown represent the mean of two (**b**) or three (**c**) biologically independent samples, editing frequencies higher than 10% are labeled in cells. For **d**, dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates.

https://doi.org/10.1038/s41589-024-01595-4

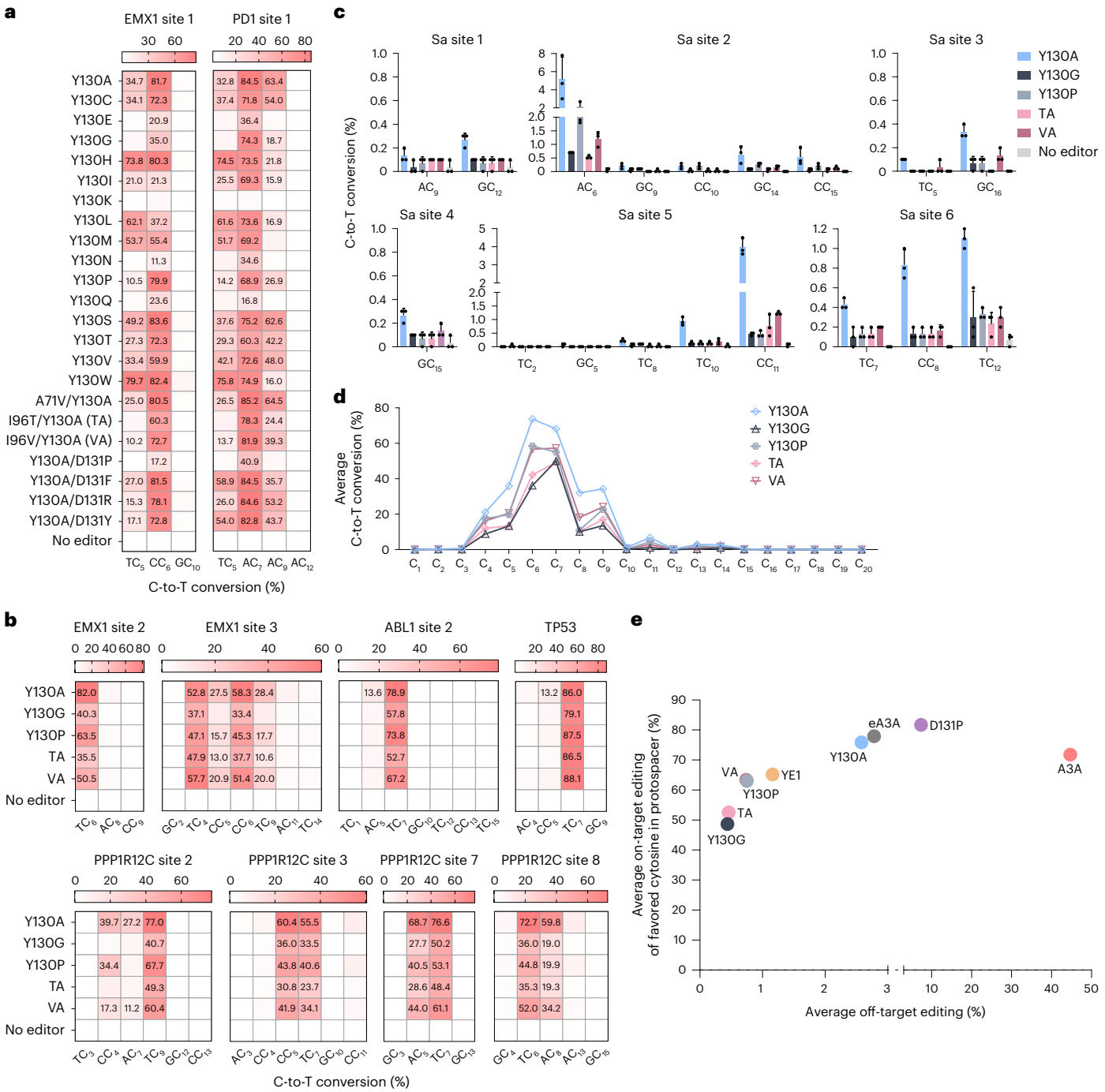


Fig. 2 | Further engineering of A3A to develop haA3A-CBE. a, Heatmaps showing editing frequencies for A3A-BE4max variants at two genomic loci. **b**, Heatmaps showing editing frequencies for A3A-BE4max variants at eight genomic loci. **c**, Cas9-independent off-target editing frequencies for A3A-BE4max variants at six SaCas9 loci. **d**, Average editing frequencies for A3A-BE4max variants at each protospacer position from ten endogenous sites shown in **a** and **b**. The editing window is defined as the protospacer positions for which average editing efficiency is ≥30% of the average editing at the maximally edited position. **e**, Average on-target editing versus average off-target editing for A3A-BE4max variants. The *y* axis reflects the average editing frequencies of the most highly edited cytosines across ten on-target sites shown in **a** and **b**. The *x* axis reflects the mean cumulative off-target C-to-T editing in orthogonal R-loops assay. On- and off-target editing frequencies of A3A, YE1, eA3A and D131P are from Fig. 1c,d, while those of others are from **a** and **b**. In heatmaps, editing efficiencies shown represent the mean of three biologically independent replicates; editing frequencies higher than 10% are labeled in cells. In **c**, dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates.

obvious reduction of efficiency on adjacent cytosines compared to the Y130A mutant (Fig. 2a).

We further evaluated these four variants in parallel with Y130A at eight additional on-target sites and six orthogonal R-loop sites. Compared to Y130A, the four variants exhibited comparable or moderately decreased on-target efficiency with much lower off-target editing at AC_6 of Sa site 2 (0.7% for Y130G, 2.1% for Y130P, 0.5% for TA and 1.2% for VA versus 5.2% for Y130A) and CC_{11} of Sa site 5 (0.5% for Y130G, 0.5% Y130P, 0.8% for TA and 1.2% for VA versus 4.0% for Y130A), and showed a more restricted editing window (Fig. 2b–d and Supplementary Fig. 2a,b). They also exhibited a low level of indel formation (Supplementary Fig. 2c). A3A-BE4max variants containing the Y130A, Y130G, Y130P, TA

or VA mutation showed on-target activity at an average of 105.6, 67.8, 87.8, 73.1 and 88.3% of intact A3A-BE4max's activity, respectively. They also showed a reduction of Cas9-independent off-target activity at an average of 17.3-, 101.6-, 59.6-, 97.2- and 60.4-fold, respectively (Fig. 2e). Together, Y130A was comparable to eA3A, while VA and Y130P showed comparable on-target but less off-target activity compared to YE1, TA and Y130G induced the lowest off-target editing (Fig. 2e). Among these well-performing variants, VA and Y130P are comparable in terms of accuracy and efficiency, but the Y130P showed slightly higher off-target activity at Sa site 2 (Fig. 2c,e). TA and Y130G are also comparable, while TA showed a marginal disadvantage in editing window (Fig. 2d,e). Therefore, Y130A, VA and Y130G variants were designated as highly accurate A3A-CBEs (haA3A-CBEs), named haA3A-CBE-A, haA3A-CBE-VA and haA3A-CBE-G, respectively.

Efficient editing at methylated or GC context sites

As cytosines in methylated regions and GC context are challenging for rAPOBEC1-based CBEs, we next evaluated haA3A-CBEs on nine genomic sites with varying levels of DNA methylation and five nonmethylated sites with GC motifs[24]. haA3A-CBE-A, haA3A-CBE-VA and haA3A-CBE-G exhibited significantly higher editing efficiency than YE1-BE4max at methylated regions, with an average of 71.7, 66.5 and 51.8%, respectively, and up to 90.2, 88.8 and 82.1%, respectively. By contrast, YE1-BE4max displayed moderate editing efficiency with an average of 27.2% (ranging from 6.7–53.0%) (Fig. 3a,b). Moreover, haA3A-CBEs exhibited much higher editing efficiency at genomic sites containing GC context, with averages of 64.9, 53.4 and 43.2%, respectively, and up to 73.4, 70.8 and 72.0%, respectively. In comparison, YE1-BE4max exhibited an average editing efficiency of only 19.9%, ranging from 13.2 to 41.8% (Fig. 3a,b). eA3A-BE4max showed a higher average activity than YE1-BE4max but lower than haA3A-CBEs at these sites, as there were few favored motifs for eA3A-BE4max (Fig. 3a,b).

To expand the targeting scope of haA3A-CBEs, we fused these deaminase variants with the SpCas9-NG nickase, which recognizes a broadened NG PAM[29]. We found that haA3A-CBE-A-NG, haA3A-CBE-VA-NG and haA3A-CBE-G-NG exhibited high editing efficiency, with an average of 71.8, 70.0 and 53.6%, respectively, and up to 82.9, 84.2 and 81.1%, respectively (Fig. 3c). In addition, indels induced by haA3A-CBEs were typically lower than or equivalent to those induced by A3A-BE4max at most of the sites (Supplementary Fig. 3). Our results indicate that haA3A-CBEs enable efficient cytosine conversion in methylated regions and GC context and are compatible with the PAM-relaxed Cas9 variant, thus expanding their targeting scope.

Further characterization of haA3A-CBEs

To further assess the performance of haA3A-CBEs, we conducted tests on 15 additional endogenous targets and found that they exhibited higher or comparable editing efficiency relative to YE1-BE4max (Extended Data Fig. 2a). Analyzing the editing outcome of a total of 39 sites with NGG PAM revealed that the activity of haA3A-CBE-G was comparable to that of YE1-BE4max, while the activity of haA3A-CBE-A and haA3A-CBE-VA were significantly higher (average 1.7- and 1.5-fold) than that of YE1-BE4max (Fig. 4a). The haA3A-CBE-A and haA3A-CBE-VA worked well in all sequence contexts (with >50% C-to-T conversion rates in average) and showed significantly higher editing efficiency in GC and AC context compared to YE1-BE4max and eA3A-BE4max (Extended Data Fig. 2b). Although the editing windows of haA3A-CBE-A and haA3A-CBE-VA were similar to YE1-BE4max and eA3A-BE4max, the central editing window (defined as the protospacer positions for which average editing efficiency is ≥70% of the average editing at the maximally edited position) of the haA3A-CBEs (C_6–C_7) was narrower than YE1-BE4max (C_5–C_7) and eA3A-BE4max (C_6–C_8) (Fig. 4b and Extended Data Fig. 3a). haA3A-CBE-G showed the most condensed editing window and the highest editing precision (Fig. 4b and Extended Data Fig. 3). High activity and condensed editing window of haA3A-CBEs were also confirmed in HeLa cells (Extended Data Fig. 4), suggesting they could be applied in various cell lines.

To assess the Cas9-dependent off-target activity of haA3A-CBEs, we analyzed 12 potential off-target sites from two targets, including six known off-target sites of EMX1 site 1 identified by GUIDE-seq[30] and six in silico-predicted off-target sites (by https://benchling.com) of ABL1 site 2. Owing to sequence similarity, A3A-BE4max induced extremely high off-target editing (23.6–94.9%, 63.1% on average) across 12 off-target sites. YE1-BE4max and eA3A-BE4max showed much lower off-target editing (3.7 and 5.5% on average) compared to A3A-BE4max, but they induced off-target editing up to 15.6 and 30.8%, respectively. The haA3A-CBEs, especially haA3A-CBE-VA and haA3A-CBE-G, exhibited high on-target activity and minimal levels of off-target editing compared to A3A-BE4max, eA3A-BE4max and YE1-BE4max (Fig. 4c and Supplementary Fig. 4). Moreover, whole-genome sequencing (WGS) of clonal samples derived from single cell revealed that YE1-BE4max generated background level of off-target edits comparable to nCas9, as previously reported[25]. Similar trends were observed in haA3A-CBE-VA and haA3A-CBE-G, indicating their base-line off-target activity (Fig. 4d and Supplementary Fig. 5). Moreover, haA3A-CBEs showed near-background level of RNA off-target editing events, lower than eA3A-BE4max and YE1-BE4max (Fig. 4e). In contrast to A3A-BE4max, haA3A-CBEs did not induce detectable off-target edits at three reported RNA off-targeting hotspots[31](Fig. 4f and Supplementary Fig. 6). We also noticed that the protein expression levels of haA3A-CBE-VA and haA3A-CBE-G were slightly higher than that of other CBE variants (Extended Data Fig. 5). These data collectively demonstrate that haA3A-CBEs are highly efficient and accurate with minimal Cas9-dependent and independent DNA and RNA off-target activity.

Precise correction of human pathogenic SNVs in cells

To assess the therapeutic potential of haA3A-CBEs in correcting genetic diseases, we evaluated their effectiveness in correcting nine pathogenic SNVs, including missense mutations in *TRPV4* (causing metatropic dysplasia)[32], *CFTR* (causing cystic fibrosis)[33], *F8* (causing hemophilia A)[34], *GCK* (causing hyperinsulinemic)[35], *GAA* (causing Pompe disease)[36], *PAH* (causing non-PKU hyperphenylalaninemia)[37], *RDH12* (causing retinal dystrophy)[38], *GJB3* (causing deafness)[39] and a mutation in splice site of *BMPR2* (causing primary pulmonary hypertension)[40] (Extended Data Fig. 6). We transfected haA3A-CBE variants into stable cell lines containing the pathogenic SNVs, and found that from SNV 1 to SNV 6, haA3A-CBE-VA and haA3A-CBE-G exhibited accurate and efficient editing, with an average perfect correction (without bystander editing) rate of 50.4% (40.9–81.3%) and 42.7% (23.7–76.9%), respectively. Compared to YE1-BE4max, haA3A-CBEs exhibited much higher activity at SNVs in GC context, with 4.8-, 3.1- and 1.7-fold corrections at SNV 4; 3.2-, 3.3- and 1.9-fold corrections at SNV 5, respectively. Though eA3A-BE4max exhibited efficient and precise cytosine editing at SNV 1 and SNV 2 within the TC motif, it showed low efficiency at SNVs (SNV 4, 5) within the GC motif, which was also an unfavorable context for YE1-BE4max (Fig. 5a). These data suggest that haA3A-CBEs, especially VA and G variants, are very accurate for correction of mutations, even in some homopolymeric cytosine sites.

While haA3A-CBEs generated a higher ratio of perfect edits than YE1-BE4max, the correction ratio at SNVs 7, 8 and 9 were low due to severe bystander edits (Fig. 5b). To minimize undesired edits, we designed new single-guide RNAs (sgRNAs) with NG PAMs to shift bystander cytosines outside the editing window. We found that all CBE-NG variants with the newly designed sgRNAs dramatically reduced bystander edits, resulting in much higher ratios of perfect corrections compared to using NGG PAMs (Fig. 5b). Specifically, the average correction rates of the three SNVs by haA3A-CBE-A-NG, haA3A-CBE-VA-NG and haA3A-CBE-G-NG were 53.0, 44.5 and 26.7% on average, which were 4.7-, 2.2- and 1.3-fold higher than that of using NGG PAM, respectively. YE1-BE4max-NG and eA3A-BE4max-NG also exhibited reduced

https://doi.org/10.1038/s41589-024-01595-4

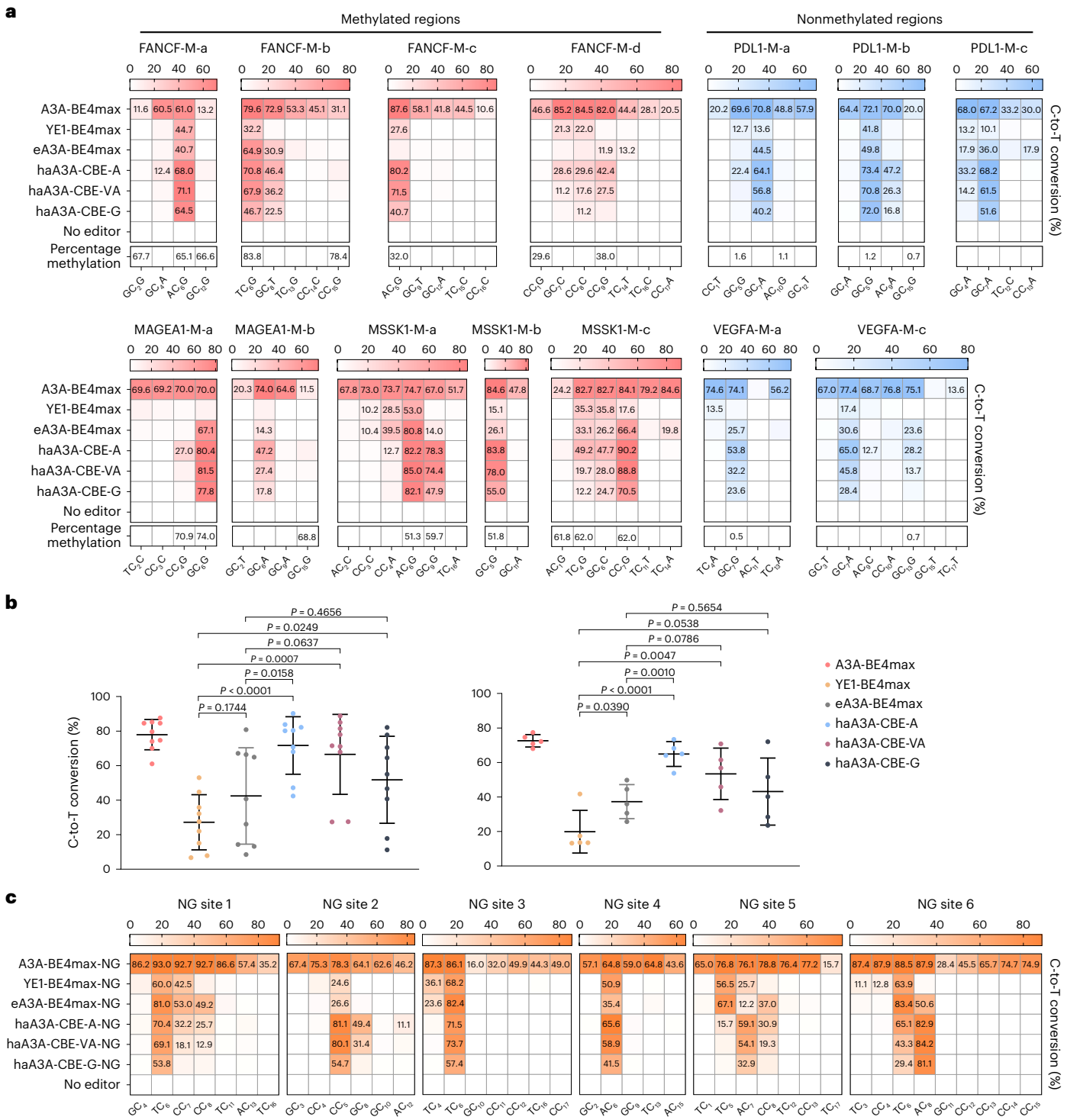


Fig. 3 | Efficient base editing in methylated regions and GC contexts by haA3A-CBEs. a, Editing frequencies for CBE variants at nine genomic loci with varying levels of DNA methylation (red heatmaps) and five nonmethylated genomic loci containing GC context (blue heatmaps). The DNA methylation levels of cytosines in CpG dinucleotides are labeled at the bottom of each target. **b**, Statistical analysis of C-to-T editing frequencies in genomic loci with varying levels of DNA methylation (left) and nonmethylated regions containing GC context (right). **c**, Editing frequencies yielded by SpCas9-NG-based CBE variants at six sites with NG PAMs (non-NGG). In heatmaps, editing efficiencies shown represent the mean of three biologically independent replicates, editing frequencies higher than 10% are labeled in cells. In **b**, each dot represents the average editing frequency of the most highly edited cytosine in each site calculated from three independent experiments, and bars represent mean ± s.d. ($n = 9$ (left) and 5 (right) genomic loci). *P* values were calculated using the two-tailed Student's *t*-test.

bystander editing, but they showed lower activity for correcting SNV 8 and SNV 9 in a GC context compared to haA3A-CBE-A-NG and haA3A-CBE-VA-NG (Fig. 5b). Taken together, haA3A-CBEs demonstrated high efficiency in correcting genetic variants while inducing low levels of indels (Supplementary Fig. 7). Furthermore, the engineered deaminase variants in haA3A-CBEs are compatible with PAM-relaxed Cas9, enabling highly precise correction of pathogenic SNVs with reduced bystander editing (Fig. 5b,c).

https://doi.org/10.1038/s41589-024-01595-4

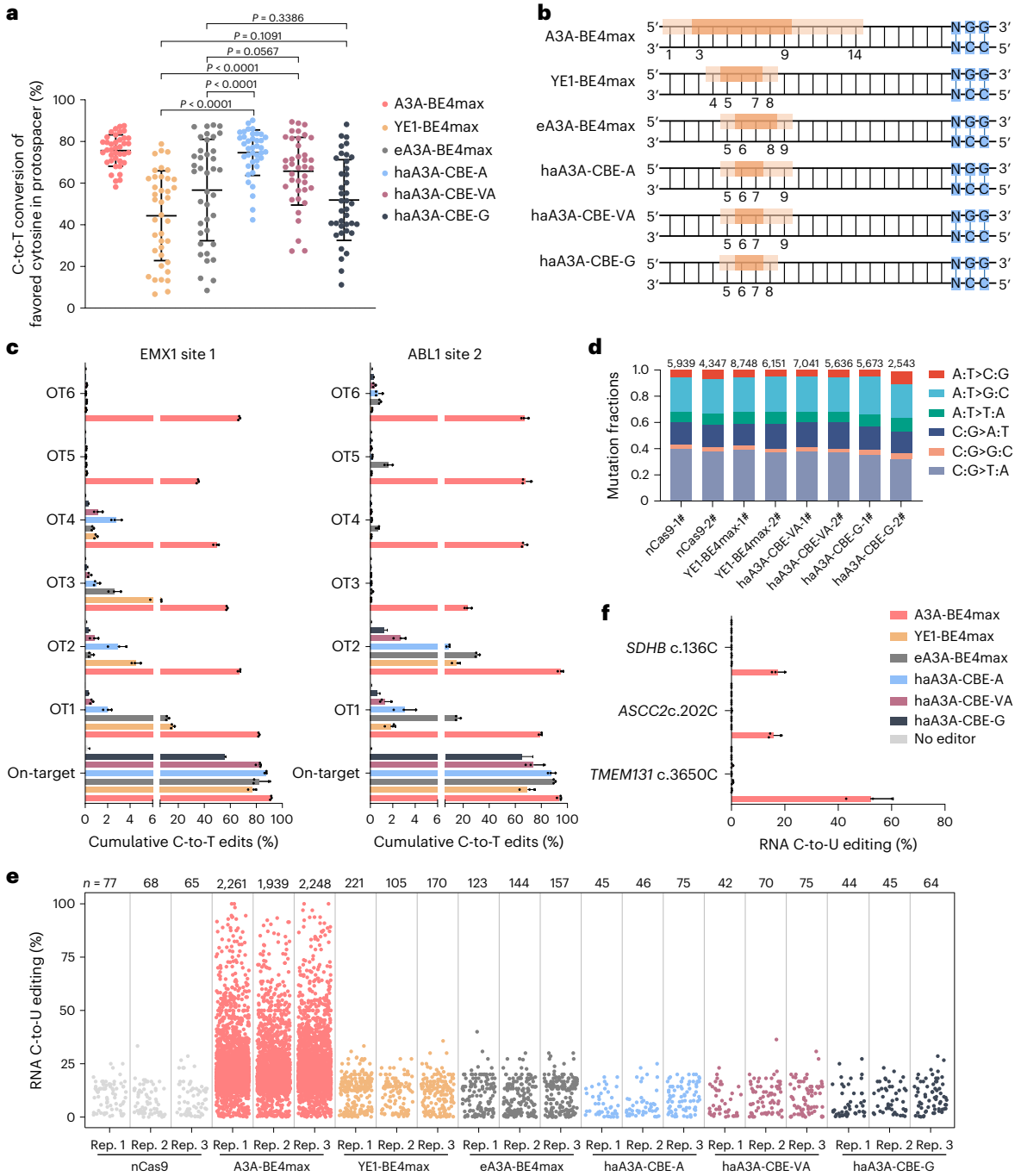


Fig. 4 | Further characterization of haA3A-CBEs. a, Statistical analysis of C-to-T editing frequencies at genomic loci with NGG PAM shown in Figs. 1c, 2a,b and 3a and Supplementary Fig. 5a. Each dot represents the average editing frequency of the most highly edited cytosine in each site calculated from three independent experiments, and bars represent mean ± s.d. ($n = 39$ tested sites). *P* values were calculated using the two-tailed Student's *t*-test. **b**, Cytosine base editing activity windows for CBE variants, light orange indicates editing window and dark orange denotes central editing window. Base positions are numbered relative to the PAM-distal end of the guide RNA. **c**, Cas9-dependent DNA on- and off-target analysis for CBE variants at indicated targets. **d**, Fractions of each type of mutation in clonal population of cells. Numbers above the column indicate the total number of SNVs. **e**, Jitter plots from RNA-seq experiments in HEK293T cells showing RNA cytosines modified by expression of CBE variants or nCas9(D10A). The *y* axis represents the efficiencies of C-to-U RNA editing. *n* = total number of modified cytosines observed. Rep., replicate. **f**, RNA C-to-U editing of three A3A hotspots in HEK293T cells transfected with CBE variants. In **c** and **f**, dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates.

Precise correction of the point mutation in HT1 mice

To assess the in vivo efficacy of haA3A-CBEs, we used the *Fah*$^{NS/NS}$ HT1 mouse model that contains an A-to-G substitution at the ATG start codon of the *Fah*, which impairs protein translation[41]. The correction of this mutation requires a C•G-to-T•A transition on the antisense strand at the G_1 position (Extended Data Fig. 7a). However, our previous study has demonstrated that BE4max generates a high ratio of simultaneous G-to-A conversions at G_1 and G_3 positions, resulting in bystander edits and preventing the successful correction of the start codon[41]. Given the condensed editing window of haA3A-CBEs, we hypothesized that they could precisely induce the desired G_1•C_1-to-A_1•T_1 conversion to restore protein translation in a more efficient and accurate manner.

https://doi.org/10.1038/s41589-024-01595-4

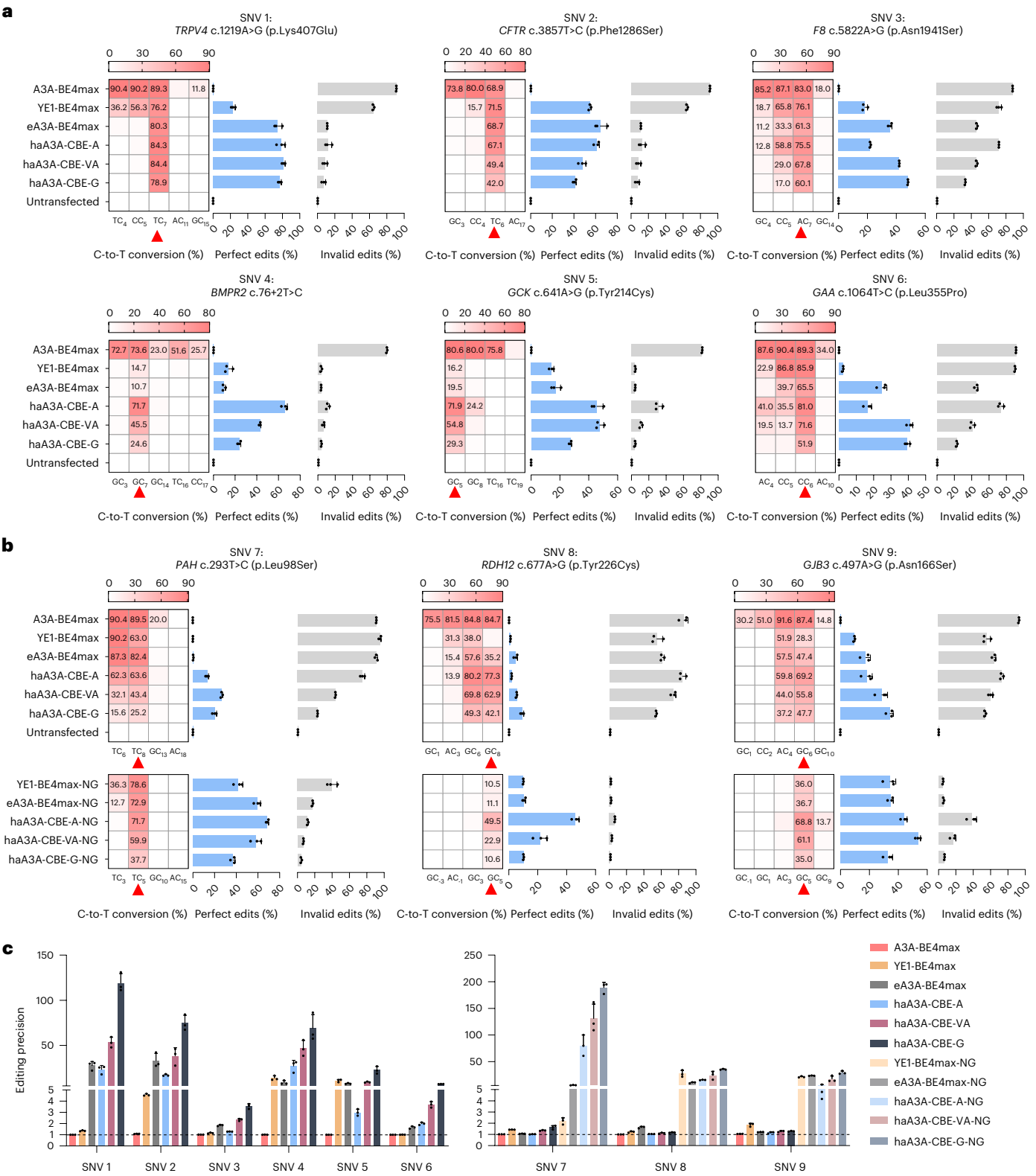


Fig. 5 | Precise editing of pathogenic SNVs by haA3A-CBEs in HEK293T cell lines. a,**b**, Efficient correction of pathogenic SNVs by CBE variants using NGG PAMs (**a**) or NG PAMs (**b**). In heatmaps, editing efficiencies shown represent the mean of three biologically independent replicates, editing frequencies higher than 10% are labeled in cells. Red triangles indicate the pathogenic SNVs. Histograms represent the frequencies of perfect edits (desired C-to-T editing at pathogenic SNV without any other editing) and invalid edits (other types of editing, including multiple Cs editing, C-to-G/A editing or combinations of them) of SNVs by CBE variants. **c**, Editing precision (defined as the ratio of the editing efficiency of the most highly edited base to that of the second highly edited base in each target site) for CBE variants in editing nine SNVs. Dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates.

We designed two sgRNAs (sg1 and sg2) to evaluate the performance of various CBE variants in a human embryonic kidney 293T (HEK293T) cell line stably transfected with the target sequence containing the *Fah*$^{NS/NS}$ allele, as previously described[41] (Extended Data Fig. 7a). HTS data showed that A3A-BE4max had a high cytosine conversion rate, but the desired single G_1•C_1-to-A_1•T_1 conversion was only 0.11%, primarily due to off-target bystander editing for both sgRNAs. The rAPOBEC1-based CBE variants (YE1-, YE2-, YEE-, EE-, R33A- and R33A/K34A-BE4max), which have a narrow editing window, showed higher accuracy than A3A-BE4max, resulting in 2.1–22.9% precise correction at the C_1·G_1 position (with 19.0- to 207.8-fold increases compared to A3A-BE4max) at the sg1 target. haA3A-CBE-VA and haA3A-CBE-G developed in this study demonstrated much higher precise correction rates (36.9 and 50.5%) than other CBE variants at the C_1·G_1 position at the sg1 target, with 335.7- and 458.7-fold increases compared to A3A-BE4max, respectively (Extended Data Fig. 7b). They also induced lower levels of byproducts and indels compared to A3A-BE4max (Supplementary Fig. 8). As the sequence lengths of the CBEs exceeded the packaging capacity of adeno-associated viral (AAV) vectors, we split the haA3A-CBE-G using an intein peptide, as previously described[41] (Extended Data Fig. 7c). Split haA3A-CBE-G efficiently induced the desired C-to-T conversions with comparable efficiency to intact haA3A-CBE-G. Similar trends were observed at the sg2 site transfected with the CBE variants, but the editing efficiency was lower than sg1 (Extended Data Fig. 7b). Therefore, sg1 and split haA3A-CBE-G were selected for in vivo studies.

Next, to achieve liver-specific base editing, we replaced the cytomegalovirus promoter with a strong liver-specific promoter Lp1 (refs. 42,43) and packaged the split haA3A-CBE into AAV2 serotype 8 particles (referred to as AAV8-haA3A-CBE-G-N and AAV8-haA3A-CBE-G-C-sg1). Two groups of 6 week-old *Fah*$^{NS/NS}$ mice were treated with a tail vein injection of 1×10^{12} vg of each AAV vector. To evaluate the initial editing efficiency, the AAV-treated mice were given 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC) water to support the survival of all hepatocytes and prevent the expansion of corrected cells[44]. Liver tissues were collected from one group at 2 weeks postinjection and from another group by biopsy at 4 weeks postinjection. The mice were then supplied with normal water for another 5 weeks (Extended Data Fig. 7d). HTS analysis revealed that the editing efficiency gradually increased from 40.7% on average at 2 weeks to 54.5% on average at 4 weeks at the targeted C_1 position (Fig. 6a). The presence of NTBC in the water during the first 4 weeks ensured that the editing efficiency represented the actual frequency without the expansion of corrected hepatocytes. These results indicate that haA3A-CBE-G is highly efficient in vivo. However, the frequency of bystander edits unexpectedly increased at 9 weeks postinjection of AAV (Fig. 6a,b), suggesting that prolonged expression of the CBE by AAV increases bystander edits, resulting in a slight decrease in the ratio of perfect corrections (Fig. 6b). Histological examination of liver sections revealed the abundance of FAH-positive hepatocytes at various time points (Fig. 6c), which was further confirmed by the evaluation of the mRNA and protein levels of the *Fah* gene (Fig. 6d,e). These results indicate that haA3A-CBE-G efficiently corrected the mutation and restored *Fah* gene expression. Restored *Fah* expression enabled the survival and weight gain of *Fah*$^{NS/NS}$ mice after the withdrawal of NTBC, while the untreated controls lost 20% of their body weight and were euthanized within 3 weeks (Fig. 6f). The serum levels of aspine aminotransferaerase (AST), alanine aminotransferase (ALT) and total bilirubin (TBIL) were significantly decreased in treated mice compared to untreated controls on NTBC withdrawal, which induced severe liver damage in untreated mice as demonstrated by liver histology (Fig. 6g,h). HTS analysis of ten potential off-target loci by computational prediction revealed no off-target effects in liver tissues of all treated mice (Extended Data Fig. 7e). Moreover, no apparent editing events were observed in nonliver tissues (Extended Data Fig. 7f). To test whether sustained expression of CBEs delivered by AAV would induce tumorigenic potential, we performed transcriptome profiling of haA3A-CBE treated or untreated *Fah*$^{NS/NS}$ mouse liver tissues. In total, 273 differentially regulated genes were identified (Extended Data Fig. 8), and gene ontology analysis showed that these genes were not involved in oncogenic pathway but were mainly associated with virus infection, most likely induced by AAV vectors (Fig. 6i).

As prolonged expression of CBEs increased bystander edits (Fig. 6a,b), we thought that using LNP to transiently deliver haA3A-CBE mRNA and sgRNA would help to reduce bystander editing. Compared to plasmid delivery, mRNA delivery improved editing precision of haA3A-CBEs in reporter cells containing *Fah*$^{NS/NS}$ mutation (Extended Data Fig. 9). A single injection of LNP-encapsulated haA3A-CBEs mRNA and sgRNA with a dose of 3 mg kg^{-1} resulted in a mean editing efficiency of 19.6% for haA3A-CBE-G and 33.0% for haA3A-CBE-VA at the targeted C_1 position 1 week after injection, with mild bystander editing at C_3 and C_{-2} (Fig. 6j). Notably, LNP-mediated delivery of haA3A-CBE-VA mRNA led to 30.1% of desired edits 1 week after injection, which was comparable to or higher than that of haA3A-CBE-G delivered via AAV at various time points (Fig. 6b,k). Additionally, although the desired edits increased 5 weeks after NTBC withdrawal (6 weeks postinjection), the byproducts remained unchanged (Fig. 6j,k), suggesting that LNP-mRNA delivery could further reduce bystander editing of haA3A-CBEs, showing promising potential for gene therapeutics.

Discussion

Base editing technology has great potential for correcting pathogenic SNVs and treating human diseases, but off-target and bystander effects have been major challenges to its broader application. Using structure-guided molecular evolution of human A3A deaminase, we have generated haA3A-CBEs, which demonstrate high efficiency, a narrow editing window and reduced DNA and RNA off-target effects in cell lines and adult animals.

We tested 58 A3A variants on 13 residues adjacent to the catalytic pocket and found that several Y130 mutants, including Y130A, Y130G and Y130P, exhibited high activity with reduced Cas9-independent off-target effects. It is surprising that the Y130A mutant, along with other mutants (such as T31D, A71P, W98A, S99A, S99G, S99P and D133N) that showed no deaminase activity in vitro[26,45], exhibited varying levels of base editing activity after being fused with nCas9 (Fig. 1b). This phenomenon was also observed previously in the A3A-N57A mutant, which, when fused with nCas9, showed TC motif preference[19], but it alone showed no deaminase activity in vitro[45]. As these residues play key roles in accommodating the substrate, the removal of their side chains would compromise the binding activity of deaminase to ssDNA backbone or contacts with target cytosines, leading to the abolition of deaminase activity. The fusion of these A3A mutants with nCas9 would enhance the accessibility of the deaminase to the substrate, offering high on-target activity while reducing Cas9-independent DNA and RNA off-target editing. Additionally, the reduced catalyzing activity might prevent A3A mutants from catalyzing successive rounds of deamination before being displaced from ssDNA, thus, reducing the possibility of editing multiple Cs, leading to a narrowed editing window. It is worth noting that haA3A-CBEs with the Y130A or Y130G mutation showed improved activity on GC and AC motif compared to eA3A-BE4max (Extended Data Fig. 2b). We propose that substituting the Y130 residue with short side chain residues conferred this loosening selectivity, since the bulky Y130 might preclude purine base with larger size from fitting in the −1 position[27].

The haA3A-CBEs inherit the distinct features of A3A, such as high efficiency in methylated cytosines and GC context, which are challenging sites for the representative CBE variant YE1 engineered from rAPOBEC1 deaminase. The evoAPOBEC1-BE4max and

https://doi.org/10.1038/s41589-024-01595-4

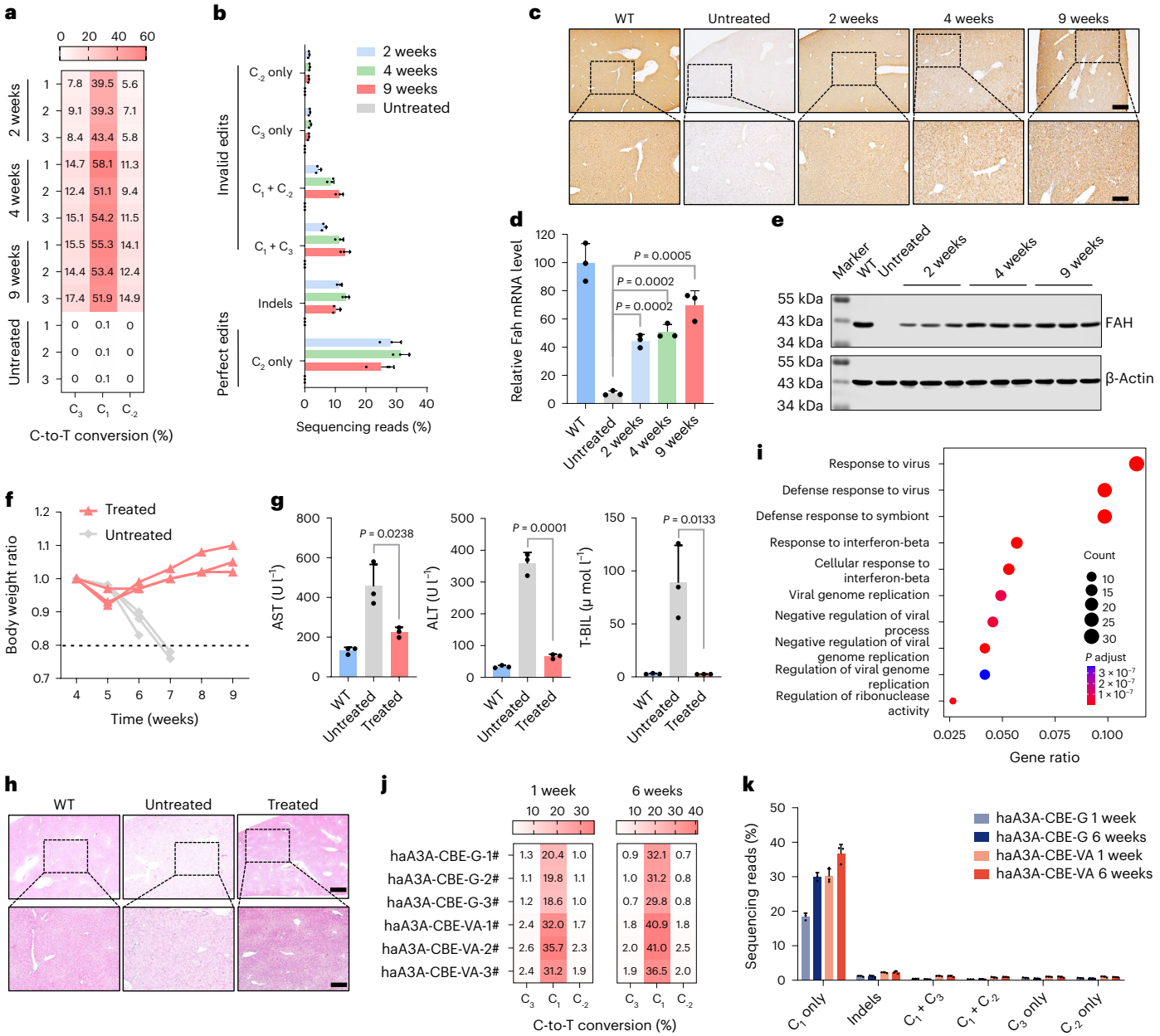


Fig. 6 | Efficient and precise C-to-T base editing in *Fah*$^{NS/NS}$ mice. a, Heatmaps represent the in vivo C-to-T editing frequencies of *Fah*$^{NS/NS}$ allele in each mouse after AAV injection. **b**, Frequencies of perfect edits, invalid edits and indels of the *Fah*$^{NS/NS}$ allele in mice after AAV injection. **c**, FAH IHC staining of liver sections from wild-type (WT), untreated and injected mice. Scale bars, 500 μm (top) and 200 μm (bottom). **d**,**e**, Relative *Fah* mRNA expression (**d**) and western blot analysis of FAH expression (**e**) in liver tissue from wild-type, untreated and injected mice. **f**, Body weight curves of mice after NTBC withdrawal. Body weight was normalized to the weight on the day when NTBC was withdrawn (4 weeks postinjection). **g**, Serum AST, ALT and TBIL levels in peripheral blood from wild-type, untreated and treated mice 9 weeks after injection. **h**, H&E staining of liver tissue sections from wild-type, untreated and treated mice at 9 weeks postinjection. Scale bars, 500 μm (top) and 200 μm (bottom). **i**, Top gene ontology terms associated with the differentially expressed genes in the comparison of untreated and treated *Fah*$^{NS/NS}$ mice. *P* values were obtained using DESeq2 (Methods). **j**, C-to-T editing frequencies of *Fah*$^{NS/NS}$ allele in each mouse 1 or 6 weeks after LNP administration. **k**, Frequencies of edited *Fah*$^{NS/NS}$ allele 1 or 6 weeks after LNP administration. In **d** and **g**, dots represent individual values, and bars represent mean ± s.d. of three mice. *P* values were calculated using the two-tailed Student's *t*-test.

evoFERNY-BE4max have been developed to efficiently edit cytosines in all sequence contexts[46]. However, their off-target editing and bystander effects remain unimproved. Recently, we and others have developed TadA-derived CBEs, which showed robust on-target activity with reduced off-target activity and a narrowed editing window[47–49]. Nonetheless, their efficiency in methylated cytosines remains to be elucidated. While eA3A-BE4max excels in editing TC motif targets, haA3A-CBEs demonstrate accurate editing in wider sequence contexts especially outperforming in methylated genomic loci, GC and AC contexts (Extended Data Fig. 2b). We believe they complement each other, each possessing unique strengths for specific application scenarios.

Genome editing is a 'hit-and-run' process, long-term expression of genome editors would increase byproducts such as off-target mutations and bystander edits. As LNP delivered base editor mRNA lasted no more than 3 days in mouse liver[17], we demonstrated that transient expression of haA3A-CBE mRNA reduced bystander editing in *Fah*$^{NS/NS}$ mice (Fig. 6j,k). As several studies showed that AAV has integration

potential and is immunogenic[50], mRNA delivery would be a safer and superior strategy for base editing.

In the present study, we demonstrated the remarkable efficiency and precision of haA3A-CBEs in correcting pathogenic SNVs in various sequence contexts, both in cell lines and an animal model of human disease. Using PAM-relaxed SpCas9-NG coupled with flexible sgRNA designing further improved the accuracy of the haA3A-CBEs for correcting human pathogenic SNVs (Fig. 5b,c). Notably, a substantial proportion of disease-associated T-to-C mutations lie in the CpG dinucleotide, which are typically methylated, or in the context of GC motif[20,24]. haA3A-CBEs are particularly well-suited for correcting these types of mutation, making them promising candidates for treating human diseases and expanding the CBE toolkit for various applications.

Online content

Any methods, additional references, Nature Portfolio 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/s41589-024-01595-4.

References

1. 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).
2. 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).
3. 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).
4. Zhao, D. D. et al. New base editors change C to A in bacteria and C to G in mammalian cells. *Nat. Biotechnol.* **39**, 35–40 (2021).
5. Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).
6. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
7. 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).
8. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).
9. Shimatani, Z. et al. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).
10. Villiger, L. et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat. Med.* **24**, 1519–1525 (2018).
11. Rossidis, A. C. et al. In utero CRISPR-mediated therapeutic editing of metabolic genes. *Nat. Med.* **24**, 1513–1518 (2018).
12. Li, G. L. et al. Gene editing and its applications in biomedicine. *Sci. China Life Sci.* **65**, 660–700 (2022).
13. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
14. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
15. Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
16. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
17. Rothgangl, T. et al. In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. *Nat. Biotechnol.* **39**, 949–957 (2021).
18. Tan, J. J., Zhang, F., Karcher, D. & Bock, R. Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nat. Commun.* **10**, 439 (2019).
19. Gehrke, J. M. et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
20. 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).
21. 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).
22. 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).
23. Wang, L. et al. Eliminating base-editor-induced genome-wide and transcriptome-wide off-target mutations. *Nat. Cell Biol.* **23**, 552–563 (2021).
24. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
25. 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).
26. Shi, K. et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **24**, 131–139 (2017).
27. Kouno, T. et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.* **8**, 15024 (2017).
28. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
29. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
30. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
31. Sharma, S. et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat. Commun.* **6**, 6881 (2015).
32. Andreucci, E. et al. TRPV4 related skeletal dysplasias: a phenotypic spectrum highlighted byclinical, radiographic, and molecular studies in 21 new families. *Orphanet J. Rare Dis.* **6**, 37 (2011).
33. Dorval, I. et al. Analysis of 160 CF chromosomes: detection of a novel mutation in exon 20. *Hum. Genet* **91**, 254–256 (1993).
34. Higuchi, M. et al. Molecular characterization of severe hemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the Factor VIII gene. *Proc. Natl Acad. Sci. USA* **88**, 7405–7409 (1991).
35. Cuesta-Munoz, A. L. et al. Severe persistent hyperinsulinemic hypoglycemia due to a de novo glucokinase mutation. *Diabetes* **53**, 2164–2168 (2004).
36. Hermans, M. M. et al. Twenty-two novel mutations in the lysosomal alpha-glucosidase gene (GAA) underscore the genotype-phenotype correlation in glycogen storage disease type II. *Hum. Mutat.* **23**, 47–56 (2004).
37. Guldberg, P. et al. A novel missense mutation in the phenylalanine hydroxylase gene of a homozygous Pakistani patient with non-PKU hyperphenylalaninemia. *Hum. Mol. Genet* **2**, 1061–1062 (1993).

38. Janecke, A. R. et al. Mutations in RDH12 encoding a photoreceptor cell retinol dehydrogenase cause childhood-onset severe retinal dystrophy. *Nat. Genet.* **36**, 850–854 (2004).
39. Liu, X. Z. et al. Digenic inheritance of non-syndromic deafness caused by mutations at the gap junction proteins Cx26 and Cx31. *Hum. Genet* **125**, 53–62 (2009).
40. Machado, R. D. et al. Pulmonary arterial hypertension: a current perspective on established and emerging molecular genetic defects. *Hum. Mutat.* **36**, 1113–1127 (2015).
41. Yang, L. et al. Amelioration of an inherited metabolic liver disease through creation of a de novo start codon by cytidine base editing. *Mol. Ther.* **28**, 1673–1683 (2020).
42. Yin, S. et al. Enhanced genome editing to ameliorate a genetic metabolic liver disease through co-delivery of adeno-associated virus receptor. *Sci. China Life Sci.* **65**, 718–730 (2022).
43. Nathwani, A. C. et al. Self-complementary adeno-associated virus vectors containing a novel liver-specific human Factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. *Blood* **107**, 2653–2661 (2006).
44. Song, C. Q. et al. Adenine base editing in an adult mouse model of tyrosinaemia. *Nat. Biomed. Eng.* **4**, 125–130 (2020).
45. Bulliard, Y. et al. Structure-function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.* **85**, 1765–1776 (2011).
46. Thuronyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
47. Chen, L. et al. Re-engineering the adenine deaminase TadA-8e for efficient and specific CRISPR-based cytosine base editing. *Nat. Biotechnol.* **41**, 663–672 (2023).
48. Neugebauer, M. E. et al. Evolution of an adenine base editor into a small, efficient cytosine base editor with low off-target activity. *Nat. Biotechnol.* **41**, 673–685 (2023).
49. Lam, D. K. et al. Improved cytosine base editors generated from TadA variants. *Nat. Biotechnol.* **41**, 686–697 (2023).
50. Greig, J. A. et al. Integrated vector genomes may contribute to long-term expression in primate liver after AAV administration. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-01974-7 (2023).

Methods

Plasmid construction

BE4max (pCMV_BE4max_3xHA) was a gift from D. Liu (Addgene plasmid; 112096). Human codon-optimized APOBEC3A, N- and C-terminal intein sequences were synthesized by GENEWIZ. The A3A-BE4max was constructed by substituting APOBEC3A for rAPOBEC1 in BE4max. The YE1-BE4max was constructed by introducing W90Y and R126E into rAPOBEC1 in BE4max through PCR. The eA3A-BE4max and haA3A-CBEs were constructed by introducing specific mutations into APOBEC3A in A3A-BE4max using PCR. DNA fragments were amplified using KOD-Plus-Neo DNA polymerase (Toyobo, KOD-401). Serial CBE plasmids were constructed using ClonExpress MultiS One Step Cloning Kit (Vazyme). Amino acid sequences of CBE variants are listed in the Supplementary Sequences. For sgRNA plasmid construction, oligonucleotides were denatured at 95 °C for 5 min followed by slow cooling to room temperature. The annealed oligonucleotides were ligated into BbsI-linearized U6-sgRNA-EF1α-EGFP (enhanced green fluorescent protein) plasmid for sgRNA expression. Lentiviral vector plasmids for generating stable HEK293T cell lines were constructed by inserting ~200 base pairs (bp) of gene sequences containing pathogenic SNVs from the ClinVar database (ClinVar (nih.gov)) or the mouse *Fah*$^{NS/NS}$ allele and a DsRed expression sequence driven by the EF-1a promoter into the lentiCRISPR v2 (52961) plasmid purchased from Addgene.

Cell culture, cell transfection and genomic DNA preparation

HEK293T (American Type Culture Collection, CRL-3216) cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), 100 U ml^{-1} penicillin and 100 mg ml^{-1} streptomycin (basal media) at 37 °C and 5% CO_2. Cells were seeded onto 24-well plates (Corning Life Sciences) and transfected with a total of 1,000 ng of plasmids at 80% confluency using polyethylenimine (PEI) (Polysciences) following the manufacturer's recommended protocol. For on-target editing experiments, 400 ng of CBE-expressing plasmid and 300 ng of guide RNA-expressing plasmid were cotransfected into HEK293T cells. For orthogonal R-loop assays to measure off-target editing, 400 ng of CBE-expressing plasmid, 300 ng of SpCas9 guide RNA-expressing plasmid and 300 ng of the plasmid expressing both SaCas9 and SaCas9 guide RNA were cotransfected into HEK293T cells. For split haA3A-CBE-G transfection, 310 ng of haA3A-CBE-G-N encoding plasmid, 320 ng of haA3A-CBE-G-C encoding plasmid and 300 ng of sgRNA-expressing plasmid were cotransfected into HEK293T cells. An empty pcDNA 3.1(+) vector from Invitrogen was used to maintain the total quantity of transfected plasmid at 1,000 ng. Three days after transfection, green fluorescent protein (GFP)-positive cells were sorted on FACSAria III (BD Biosciences) and genomic DNA was extracted using a TIANamp Genomic Kit (Tiangen Biotech, DP304-03) according to the manufacturer's instructions.

HTS and data analysis

To examine the on- and off-target base editing efficiency, KOD-Plus-Neo DNA polymerase and site-specific primers containing adapter sequences (forward 5′-GGAGTGAGTACGGTGTGC-3′; reverse 5′-GAGTTGGATGCTGGATGG-3′) at the 5′ end were used to amplify genomic regions of interest from genomic DNA (roughly 100–150 ng). The above products were then subjected to a second round of PCR using primers containing unique barcode sequences. The resulting PCR products were mixed as HTS libraries and then sequenced on an Illumina HiSeq platform. HTS data were analyzed by BE-Analyzer[51] to output the C-to-T conversion and indel rates. The tested endogenous genomic loci, SaCas9 loci and Cas9-dependent off-target sites, along with the PCR primers used for amplifying them are listed in Supplementary Tables 1–3.

Bisulfite sequencing

Genomic DNA was isolated from HEK293T cells using TIANamp Genomic Kit and was then subjected to bisulfite conversion using the EZ DNA Methylation Kit (Zymo Research, D5001) according to the manufacturer's instructions. Subsequently, the sites of interest were amplified via PCR using TaKaRa Taq Hot Start Version (Takara, R007A), with the bisulfite-converted DNA serving as the template. The PCR products were subjected to a second round of PCR for HTS. The HTS data were analyzed using BE-Analyzer[51]. As the bisulfite treatment converts nonmethylated cytosine residues to uracil, while the 5-methylcytosine residues remain unaffected, the percentage of reads with cytosine stands for the methylation level of a certain cytosine residue. Primers used for bisulfite sequencing are listed in Supplementary Table 4.

Lentiviral vector production and the creation of stable cell lines

Here, 300 ng of psPAX2, 300 ng of pMD2.G and 300 ng of lentiviral vector plasmid were cotransfected into HEK293T cells seeded in a 24-well plate using PEI. After 8 h of transfection, the medium was replaced with fresh culture medium. After 48 h of transfection, a 0.45 μm syringe filter (Pall) was used to filter the supernatant containing lentiviral particles. The filtered lentivirus-containing supernatant was added to a 12-well plate cultured with HEK293T cells. After 48 h of transduction, cells were split into a 24-well plate and cultured in culture medium supplemented with puromycin (2 μg ml^{-1}). Following 7 days of puromycin selection, cells with stable integration were used for transfection to test the correction rate by CBE variants.

AAV virus production

AAV virus was produced as previously described[41]. HEK293T cells were maintained in DMEM supplemented with 10% FBS in 15 cm dishes (Corning Life Sciences). When cells reached 90–100% confluence, they were split 1:2.5 into 15 cm dishes 16–20 h before PEI transfection. Each plate of HEK293T cells were transfected with 10 μg of AAV8 capsid plasmid, 10 μg of AAV helper plasmid and 10 μg of AAV expression vector at the confluency of approximately 80%. After 8 h of incubation of the transfection component, the medium was replaced with fresh DMEM supplemented with 10% FBS. At 60–72 h posttransfection, cells were scraped from the plates using a rubber cell scraper (Corning Life Sciences) and pipetted into 50 m centrifuge tubes for centrifugation at 1,575*g*. Cell pellets were resuspended in 500 μl lysis buffer (150 mM NaCl, 20 mM Tris (pH 8.0)) per plate and subjected to three consecutive cycles of freezing and thawing between a dry ice and ethanol bath and a 37 °C water bath for complete lysis. Adding $MgCl_2$ and Benzonase (Merck) into the cell lysates to final concentrations of 1 mM and 25 U ml^{-1}, respectively. After 30 min of incubation at 37 °C, the cell lysates were centrifuged at 3,889*g*, 4 °C for 20 min. The virus-containing supernatants were transferred into Quick-Seal tubes (Beckman) for iodixanol gradient ultracentrifugation at 264,429*g*, 16 °C for 2 h in a Beckman Ti70 rotator. Then, approximately 4 ml per tube of the 40% layer containing virus was collected using a 5 ml needle and mixed with 1× PBS to reach a volume of 15 ml. The diluted liquid was centrifuged in a 50 kDa Amicon Ultra Centrifugal Filters (Merk) to remove the iodixanol and concentrate the virus. Finally, quantitative PCR (qPCR) was performed to determine the viral titer of the concentrated virus. Primers used for qPCR are listed in Supplementary Table 5.

Animal studies

All experiments involving mice complied with 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. The C57BL/6J mice were purchased as adults from Shanghai Laboratory Animal Center. The *Fah*$^{NS/NS}$ mice were obtained as previously described[41], and bred and maintained in East China Normal University Center for Animal Research. All mice were housed in a specific pathogen-free facility on 12 h light–dark cycles with a temperature between 20 and 22 °C and humidity of 40–60%.

All mice had ad libitum access to food and water in standard cages. *Fah*$^{NS/NS}$ mice were maintained on NTBC water (10 mg l^{-1}) to prevent liver damage. Six-week-old *Fah*$^{NS/NS}$ mice were tail vein injected with AAV8 virus at a dose of 2×10^{12} vg total (1×10^{12} vg per half) per mouse. One group of mice was euthanized by carbon dioxide asphyxiation at 2 weeks postinjection, and another group received biopsies at 4 weeks postinjection for analyses. After biopsy, NTBC was withdrawn and mice were euthanized at 9 weeks after injection for analyses. Untreated mice with over 20% of body weight loss were humanely euthanized after NTBC withdrawal. Serum was obtained by centrifuging blood collected via retro-orbital puncture from treated or untreated *Fah*$^{NS/NS}$ mice, as well as wild-type C57BL/6J mice at 13,800*g* for 10 min at 4 °C. AST, ALT and TBIL levels of mouse serum were measured by the Adicon Clinical Laboratory. Both male and female *Fah*$^{NS/NS}$ mice were used since the phenotype is highly identical in both sexes. Eight-week-old wild-type C57BL/6J mice were used to determine *Fah* gene expression levels and serum AST, ALT and TBIL levels.

Western blot, IHC and H&E staining

The western blot assays, immunohistochemistry (IHC) and hematoxylin and eosin (H&E) staining were performed as previously described[41]. Liver tissues from treated or untreated *Fah*$^{NS/NS}$ mice, as well as wild-type C57BL/6J mice, were collected after euthanizing and subjected to homogenization. HEK293T cells were seeded onto 24-well plates and transfected with 400 ng of CBE-expressing plasmid and 300 ng of guide RNA-expressing plasmid using PEI. Cells were collected 48 h posttransfection. Liver tissue homogenates or HEK293T cells were suspended in radioimmunoprecipitation assay buffer with proteinase and phosphatase inhibitors (Calbiochem) on ice for 30 min. Then, lysates were subjected to centrifugation (13,800*g* for 10 min at 4 °C), and total protein in the supernatant was quantified using a Pierce protein bicinchoninic acid assay kit (Thermo Fisher Scientific). Next, 30 µg of total protein lysate of each sample was loaded on SDS–PAGE gel for electrophoresis, and then transferred to nitrocellulose membranes (Millipore). Subsequently, the membranes were blocked with 5% BSA for 1 h at room temperature and then incubated with anti-FAH antibody (AbboMax, 1:2,000), anti-CRISPR–Cas9 antibody (Abcam, 1:2,000) or anti-β-actin antibody (Sigma, 1:5,000) overnight. For IHC and H&E staining, mouse livers were fixed in 4% paraformaldehyde at 4 °C overnight and then subjected to dehydration in ascending concentrations of ethanol and xylene. The dehydrated tissues were embedded in paraffin and cut into sections (thickness, 5 µm). IHC staining was performed to detect FAH-positive hepatocytes using the anti-FAH antibody (AbboMax, 1:2,000). H&E staining was performed by staining sections with H&E to detect the histopathological changes.

PCR with reverse transcription and qPCR

Liver tissue homogenates form mice were suspended in RNAiso Plus (Takara) for total RNA extraction according to standard protocols. Subsequently, total RNA was used to generate complementary DNA (cDNA) using a HiScript II Q RT SuperMix kit (Vazyme). A3A hotspots of interest were amplified from cDNA (100–150 ng) via PCR using KOD-Plus-Neo DNA polymerase and site-specific primers listed in Supplementary Table 6. For qPCR reactions performed on QuantStudio3 (Applied Biosystems), 100–150 ng of cDNA and Hieff qPCR SYBR Green master mix (Yeasen Biotechnology) were used. qPCR data normalization was conducted using *Actb*. Primers used for qPCR are listed in Supplementary Table 7.

RNA-seq experiments

HEK293T cells seeded on 10 cm dishes were transfected with 17 µg of CBE-P2A-EGFP variants expressing plasmids or negative control (nCas9-P2A-EGFP expressing plasmids) at 80% confluency. Two days after transfection, cell sorting was performed on FACSAria III (BD Biosciences), and around 500,000 cells (top 15% GFP signal) were collected. Then, total RNA was extracted from the collected cells using RNAiso Plus according to standard protocols. RNA-sequencing library preparation and sequencing was performed as previously described[52].

RNA-seq variant calling and quality control

RNA-seq data analysis was performed by following previous researches[15,53]. In the quality-control step, clean data (clean reads) destined for downstream analyses were acquired by eliminating reads containing adapter and trimming low-quality bases using Trimmomatic (v.0.39) from the raw FASTQ data. Subsequently, an index of the reference genome (hg38) was generated, and the alignment of paired-end clean reads to the reference genome was conducted using STAR (v.2.7.10a). Single-nucleotide polymorphism calling was carried out using GATK software (v.4.2.6.1), and Picard software (v.2.27.2) was used to determine empirical editing efficiencies on PCR-deduplicated aligned reads. The refined BAM files underwent split reads that spanned splice junctions, local realignment, base recalibration and variant calling with SplitNCigarReads, IndelRealigner, BaseRecalibrator and HaplotypeCaller tools, respectively, provided by GATK. Base quality recalibration used the known variants in dbSNP v.138 (https://www.ncbi.nlm.nih.gov/snp/?term=txid9606[Organism:exp]). Downstream analyses concentrated exclusively on SNVs on canonical chromosomes (1–22, X, Y and M) from all called variants. We executed bam-readcount (v.0.8.0) on the .BAM file obtained from the final output of the GATK pipeline to evaluate the per-base nucleotide abundances per variant. Variant loci from base editor overexpression samples underwent filtration to remove sites lacking high-confidence reference genotype calls from the control samples. For a specific SNV in a control sample, the read coverage should exceed the 90th percentile of the read coverage across all SNVs in the corresponding base editor overexpression sample. These loci had to exhibit a consensus of at least 99% of reads carrying the reference allele across all the control samples. For RNA edits in nCas9 controls, only loci with ten or more reads and with over 0% of reads containing alternate alleles were retained. They also needed to demonstrate a consensus of at least 99% of reads containing the reference allele across all other control samples. C-to-U base edits encompassed C-to-U edits identified on the positive strand and G-to-A edits originating from the negative strand.

Differential gene expression analysis

The quality assessment of Illumina RNA-seq data was conducted using the FastQC tool (v.0.11.8). Trimmomatic (v.0.39) was applied for read trimming, and RseQC (v.4.0.0) was used to ascertain strandness. The alignment of reads to the mm10 reference genome was conducted using HiSAT2 (v.2.2.1) with default settings and -rna-strandness R. Subsequently, featureCounts (v.2.0.4) was used to counter the aligned reads. Differential gene expression analysis was performed using the R package DESeq2 (v.1.38.1), considering a fold-change greater than 2 and adjusted *P* value less than 0.05.

RNA synthesis and transfection

Capped and polyadenylated CBE mRNAs, incorporating N^1-methyl pseudo-U, were produced through in vitro transcription using linearized plasmid DNA templates and RiboMAX Large Scale RNA Production Systems (Promega, P1300), and were purified using Ammonium Acetate Precipitation. The transcripts' concentrations were determined by measuring the light absorbance at 260 nm (Nanodrop). To avoid the drawback of possible imperfect transcription of CBE at carboxy-terminal containing UGI, a UGI mRNA was independently synthesized. Chemically synthesized sgRNA with 2′-O-methyl and phosphorothioate modifications at the first three 5′ and 3′ terminal RNA residues was procured from a commercial supplier (Genscript, Supplementary Sequences). HEK293T cells were seeded onto 24-well plates and transfected at 80% confluency with 400 ng of CBE mRNA,

50 ng of sgRNA, 100 ng of UGI mRNA and 100 ng of EGFP mRNA using Lipofectamine 2000 (Invitrogen, 11668030), following the manufacturer's recommended protocol. Then 60 h after transfection, GFP-positive cells were sorted on FACSAria III (BD Biosciences) and genomic DNA was extracted.

WGS and data analysis

A plasmid expressing both CBE-P2A-EGFP and ABL1 site 2-targeting guide RNA was transfected into HEK293T cells seeded on a 24-well plate. Two days after transfection, around 50,000 cells (top 15% GFP signal) were collected by cell sorting and then diluted for clonal expansion in 96-well plates. Single cell-derived clones were again expanded in 24-well plates. After 14–16 days, cells were collected and genomic DNA was extracted. Library preparation and sequencing was performed at GENEWIZ on an Illumina NovaSeq 6000. The analysis of WGS data was performed by following previous research[54]. In brief, BWA (v.0.7.17-r1188), Samtools (v.1.14), Picardtools (gatk, v.4.2.6.1) and Varscan (v.2.4.6) were used to analyze the sequencing data. BWA (v.0.7.17-r1188) was used to align the fastq files to the hg38 human genome. Next, alignments were merged and sorted by Samtools (v.1.14), and duplicate reads were marked by Picardtools (gatk, v.4.2.6.1). Variant calling was conducted using Varscan (v.2.4.6) somatic, where the reads from each clone were compared to the reads of the parental clone. Somatic mutations meeting the criteria of $P < 0.001$, 'variant allele frequency in normal' <5%, 'variant allele frequency in tumor' >33.33% and 'reads supporting variant in tumor' >6 were deemed valid calls.

LNP encapsulation

ALC0315, DSPC, cholesterol and ALC0159 were mixed in 100% ethanol at a molar ratio of 46.3:42.7:9.4:1.6. The CBE mRNA, UGI mRNA and sgRNA were dissolved in 50 mM acetate buffer (pH 4.5) at a weight ratio of 4:1:4. The LNP was formed by microfluidic mixing of the lipid and mRNA solution at a flow ratio of 1:3 (ethanol:aqueous phase). The obtained mixture was diluted with isovolumetric PBS, then the remaining buffer was exchanged into PBS via ultrafiltration.

Statistics and reproducibility

Statistical analyses were conducted using GraphPad Prism v.8 (GraphPad Software). Data are presented as mean ± s.d. A minimum of $n = 3$ biologically independent experiments or three biologically independent samples were included for statistical evaluation. Unpaired two-tailed Student's t-tests were used to determine the P values. $P < 0.05$ was considered significant. The IHC and HE staining experiments were independently repeated twice with similar results.

Reporting summary

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

Data availability

HTS data have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under the accession codes PRJNA938162, PRJNA938284, PRJNA938523, PRJNA938572, PRJNA938580 and PRJNA1032216. RNA-seq data and HTS data of A3A hotspots have been deposited in the SRA database under the accession code PRJNA938578. WGS data have been deposited in the SRA database under the accession code PRJNA1042830. Data for mouse liver treated with CBE have already been posted in the SRA database under the accession code PRJNA937584. The hg38 reference genome (*Homo sapiens* genome assembly GRCh38.p14, NCBI, NLM (nih.gov)) was used for alignment in the analysis of RNA-seq and WGS data from HEK293T cells. The mm10 reference genome (Mus musculus genome assembly GRCm38.p6, NCBI, NLM (nih.gov)) was used for the analysis of RNA-seq data from mice. ClinVar database (ClinVar (nih.gov)) was used to identify pathogenic SNVs that can be correct by cytosine base editing. The published structure of A3A (PDB ID 5SWW) can be accessed at the RCSB Protein Data Bank (PDB) 5SWW: Crystal Structure of Human APOBEC3A complexed with ssDNA. There are no restrictions on data availability. Source data are provided with this paper.

References

51. Hwang, G. H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinf.* **19**, 542 (2018).
52. Chen, L. et al. Engineering a precise adenine base editor with minimal bystander editing. *Nat. Chem. Biol.* **19**, 101–110 (2023).
53. 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).
54. McGrath, E. et al. Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nat. Commun.* **10**, 5353 (2019).

Acknowledgements

We are grateful to the East China Normal University Public Platform for Innovation (011). We thank Y. Zhang from the Flow Cytometry Core Facility of School of Life Sciences in the East China Normal University. This work was partially supported by grants from National Key R&D Program of China (grants 2022YFC3400203 to Y.G., 2023YFC3403400 to D.L., 2023YFE0209200 to D.L., 2019YFA0110802 to D.L., 2019YFA0802800 to Mingyao Liu, 2019YFA0110902 to C.Y. and 2019YFA0801402 to F.Z.), the National Natural Science Foundation of China (grants 32025023 to D.L., 32230064 to D.L., 31971366 to L.W., 82230002 to Mingyao Liu, 82100773 to Y.G., 82271890 to F.Z., 21825701 to C.Y., 91953201 to C.Y. and 92153303 to C.Y.), grants from the Shanghai Municipal Commission for Science and Technology (grant 21JC1402200 to D.L.), the Innovation program of Shanghai Municipal Education Commission (grant 2019-01-07-00-05-E00054 to D.L.), the China Postdoctoral Science Foundation (grant 2022M721161 to L.Y.) and Innovative Research Team of High-Level Local Universities in Shanghai (grant SHSMU-ZDCX20212200 to D.L. and F.Z.).

Author contributions

L.Y. and D.L. conceived the rational engineering of A3A. L.Y., Y.H., M.W. performed the experiments of plasmid construction, cell culture, cell transfection, genomic DNA preparation, AAV virus production, lentiviral vector production and the creation of stable cell lines. L.Y., Y.H., M.W. and X.R. performed the experiments of PCR, with reverse transcription and qPCR and prepared the HTS libraries. L.Y., Y.H., M.W., T.Z. and Y.G. performed the western blot, IHC and HE staining experiments. L.Y., Y.H., M.W., T.Z., S.Y., Dexin Zhang and J.M. performed the experiments of cell sorting. L.Y. and T.Z. performed the experiments of bisulfite sequencing. L.Y., Y.H., M.W. and Meizhen Liu performed the animal experiments. L.Y., Y.H., M.W., Dan Zhang, L.W. and D.L. analyzed the HTS data. G.S. and L.Y. analyzed the structure of A3A. Dan Zhang, H.W. and H.M. analyzed the RNA-seq and WGS data. X.C. and J.L. performed the experiments for production of CBE mRNAs and LNP encapsulation, respectively. D.L., L.Y., L.W., F.Z., Mingyao Liu, C.Y., B.F., Y.C. and G.S. designed the experiments and wrote the manuscript with the input from all the authors. D.L. supervised the research.

Competing interests

A patent application (application number CN202210791520.3) based on the A3A mutants such as Y130A, VA and Y130G reported in this study has been submitted, not yet authorized. The patent applicant is East China Normal University, and D.L., L.Y., Y.H. and M.W. are the inventors. The remaining authors declare no competing interests.

Additional information

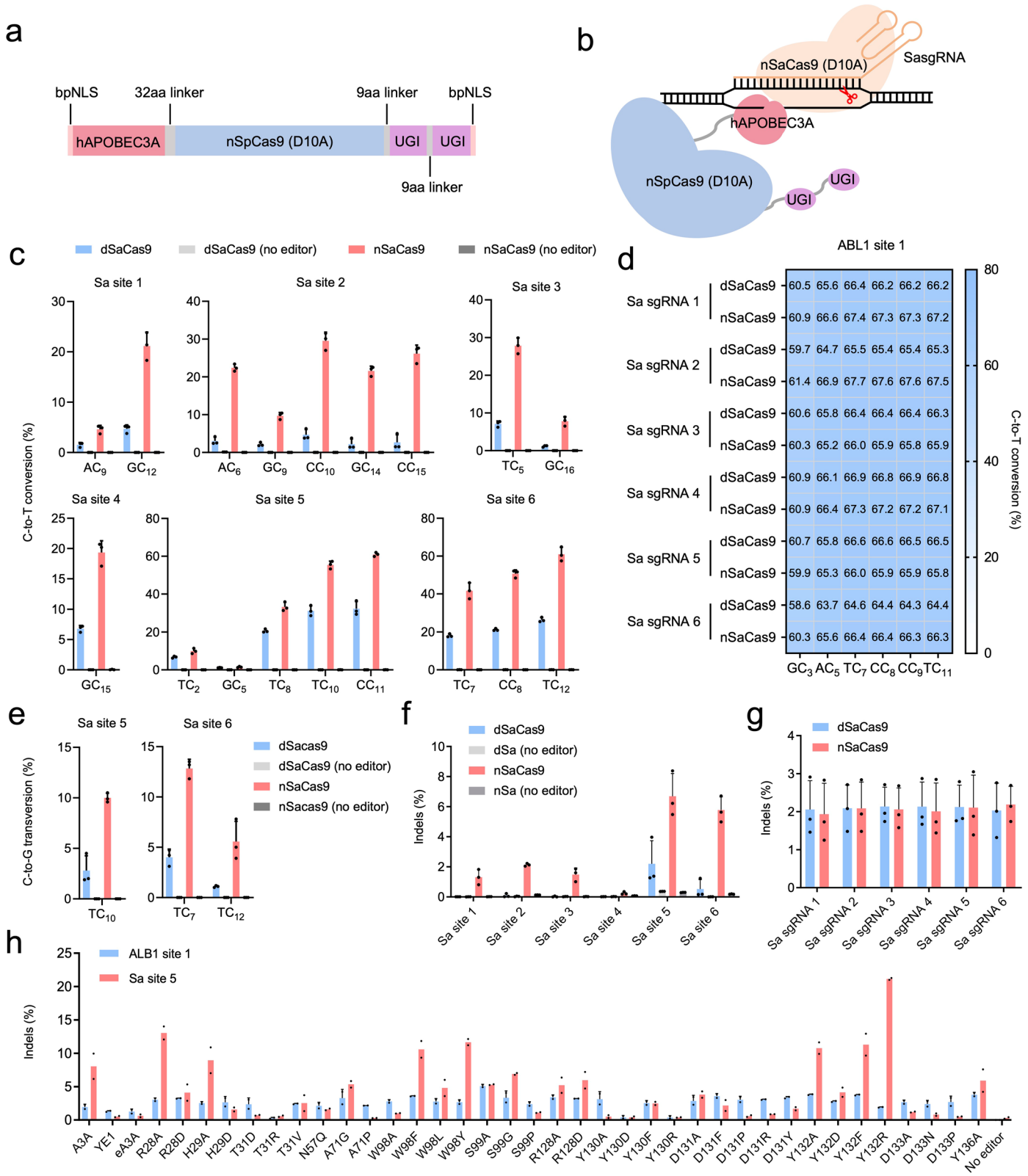
Extended data is available for this paper at https://doi.org/10.1038/s41589-024-01595-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-024-01595-4.

Correspondence and requests for materials should be addressed to Fanyi Zeng, Liren Wang or Dali Li.

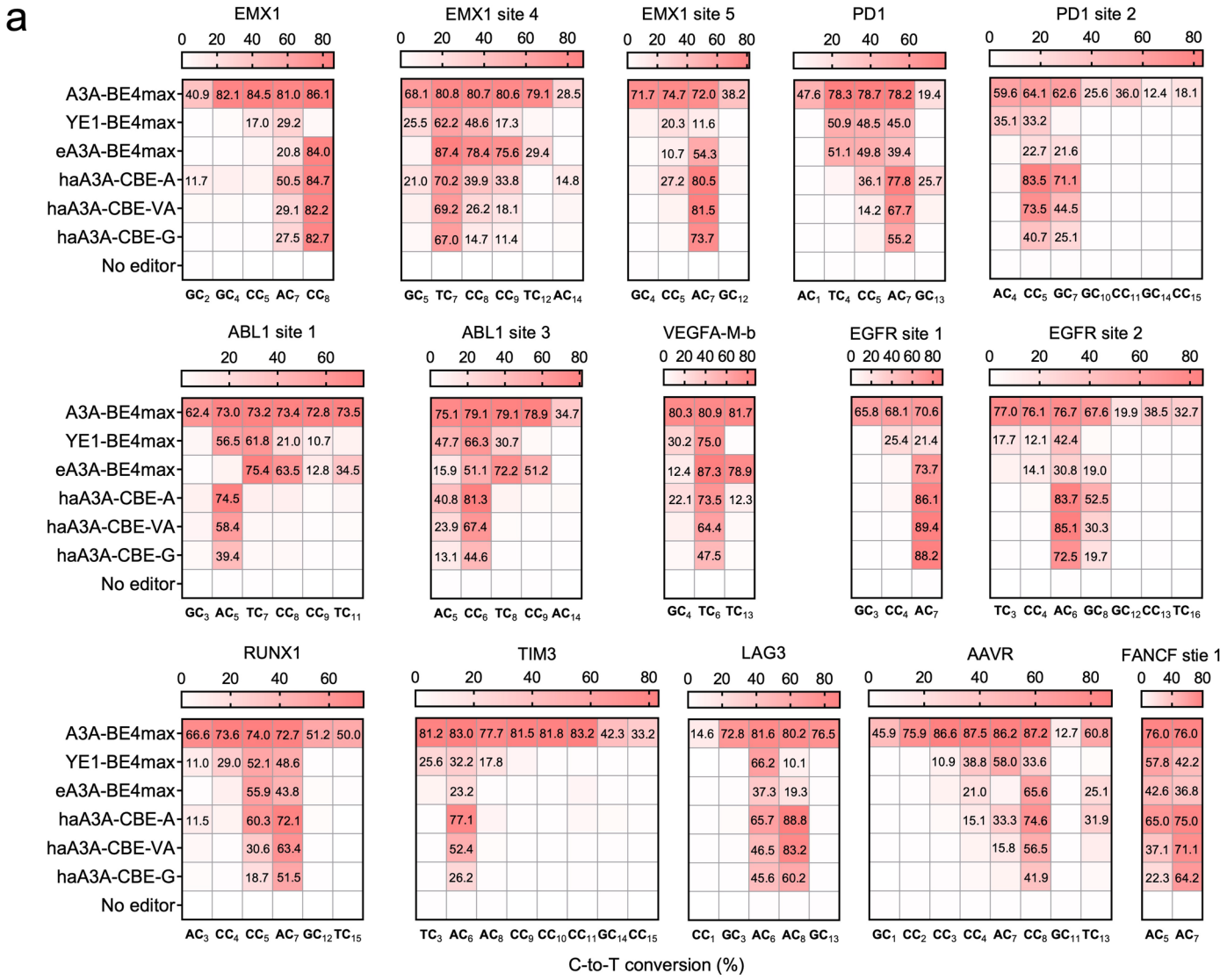
Peer review information *Nature Chemical Biology* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

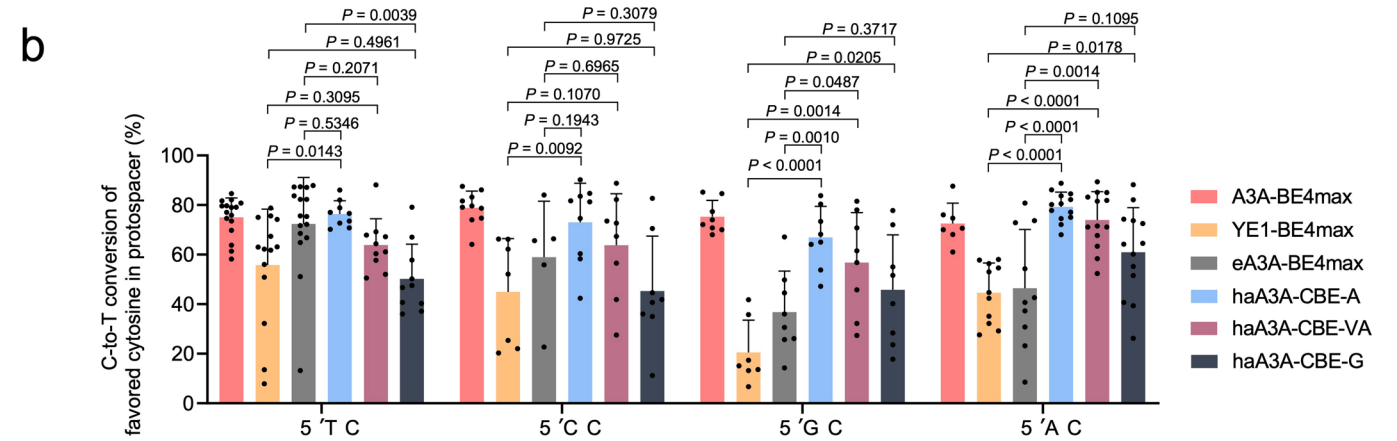
https://doi.org/10.1038/s41589-024-01595-4



Extended Data Fig. 1 | Optimizing human cell orthogonal R-loop assay to evaluate the Cas9-independent off-target activity of SpCas9-based CBEs in HEK293T cells. a, Architecture of A3A-BE4max for all CBE constructs used in this study. **b**, Schematic of Cas9-independent deamination of cytosines within nSaCas9(D10A)-induced R-loops by SpCas9-based CBEs. **c**, Cas9-independent off-target editing frequencies for A3A-BE4max at six SaCas9 loci using dSaCas9 or nSaCas9. **d**, On-target editing frequencies for A3A-BE4max at ABL1 site 1 when either dSaCas9 or nSaCas9 were used to target six SaCas9 loci. **e**, Cas9-independent off-target C-to-G editing frequencies for A3A-BE4max at SaCas9 site 5 and SaCas9 site 6 using dSaCas9 or nSaCas9. **f**, Frequencies of indels induced by A3A-BE4max's Cas9-independent off-target deamination at six SaCas9 loci using dSaCas9 or nSaCas9. **g**, Frequencies of indels induced by A3A-BE4max at ABL1 site 1 using dSaCas9 or nSaCas9. **h**, Frequencies of indels induced by A3A-BE4max variants at ABL1 site 1 and Sa site 5 in HEK293T cells. Dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates.

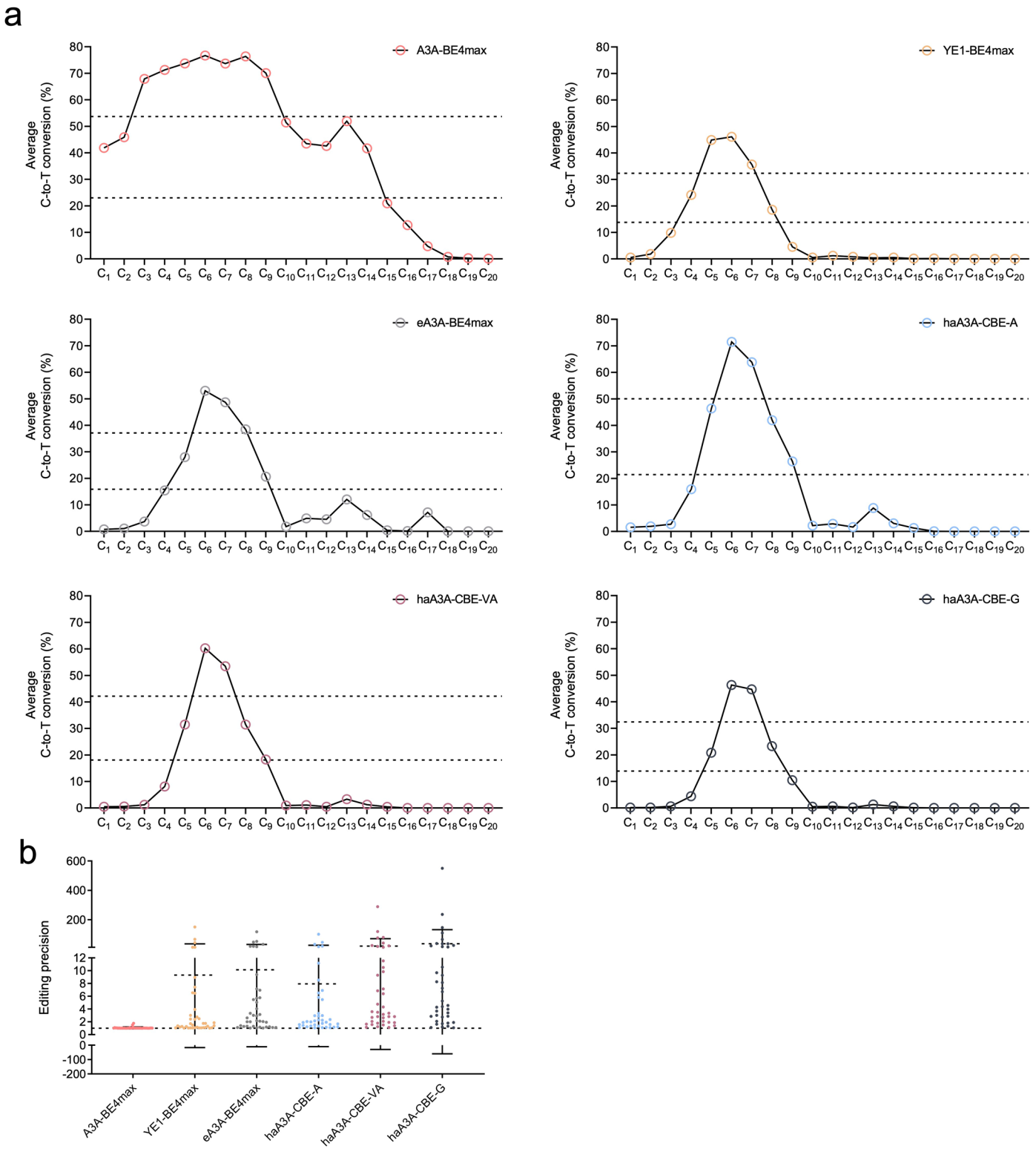
https://doi.org/10.1038/s41589-024-01595-4





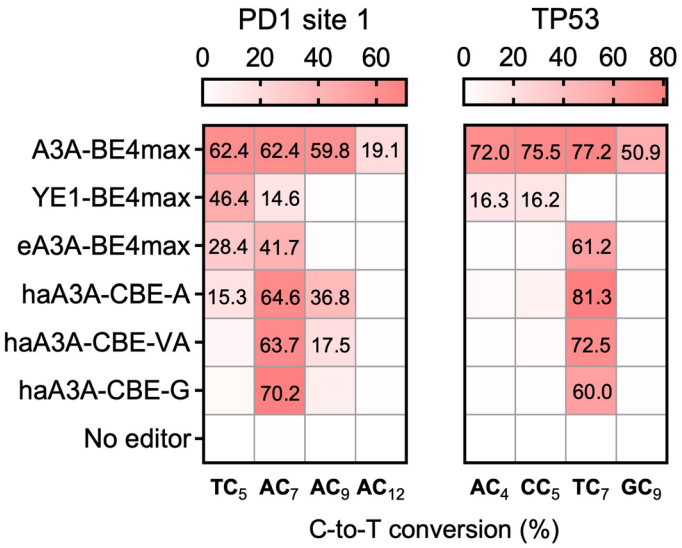
Extended Data Fig. 2 | Further characterize the on-target activity of haA3A-CBEs. a, Editing frequencies for haA3A-CBEs at additional 15 genomic sites in HEK293T cells from three independent biological replicates, editing frequencies higher than 10% are labeled in cells. **b**, Performance of CBE variants in editing favored cytosines in TC context (n = 16, 14, 16, 9, 10, and 10 cytosines for A3A-BE4max, YE1 –BE4max, eA3A-BE4max, haA3A-CBE-A, haA3A-CBE-VA, and haA3A-CBE-G, respectively.), CC context (n = 10, 7, 5, 9, 8, and 8 cytosines for A3A-BE4max, YE1 –BE4max, eA3A-BE4max, haA3A-CBE-A, haA3A-CBE-VA, and haA3A-CBE-G, respectively.), GC context (n = 8, 7, 8, 8, 8, and 8 cytosines for A3A-BE4max, YE1 –BE4max, eA3A-BE4max, haA3A-CBE-A, haA3A-CBE-VA, and haA3A-CBE-G, respectively.), and AC context (n = 7, 11, 10, 13, 13, and 13 cytosines for A3A-BE4max, YE1 –BE4max, eA3A-BE4max, haA3A-CBE-A, haA3A-CBE-VA, and haA3A-CBE-G, respectively.) from 39 sites with NGG PAM shown in **a**, Fig. 1c, Fig. 2a,b and Fig. 3a. Each dot represents the average editing frequency of the most highly edited cytosine in each site calculated from three independent experiments, and bars represent mean ± s.d. *P* values were calculated using the two-tailed Student's t-test.

https://doi.org/10.1038/s41589-024-01595-4



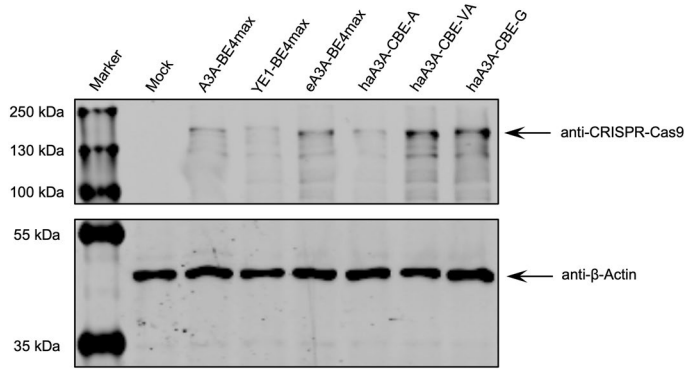
Extended Data Fig. 3 | Editing window plots (a) and editing precision (b) for CBE variants. In **a**, the lower and upper dotted lines in each plot denote the editing window cutoff and central editing window cutoff, respectively. The editing window and central editing window are defined as the protospacer positions for which the average editing efficiency is ≥30% and ≥70%, respectively, of the average editing at the maximally edited position. In **b**, each dot indicates the editing precision for each CBE variant at each tested site shown in Fig. 1c, Fig. 2a,b, Fig. 3a, and Extended Data Fig. 2a (n = 39 tested sites). Editing precision is defined as the ratio of the editing efficiency of the most highly edited base to that of the second highly edited base in each target site. Bars represent mean ± s.d.

https://doi.org/10.1038/s41589-024-01595-4



Extended Data Fig. 4 | Editing activities of haA3A-CBEs in Hela cells. In heatmaps, editing efficiencies shown represent the mean of three biologically independent replicates, editing frequencies higher than 10% are labeled in cells.

https://doi.org/10.1038/s41589-024-01595-4



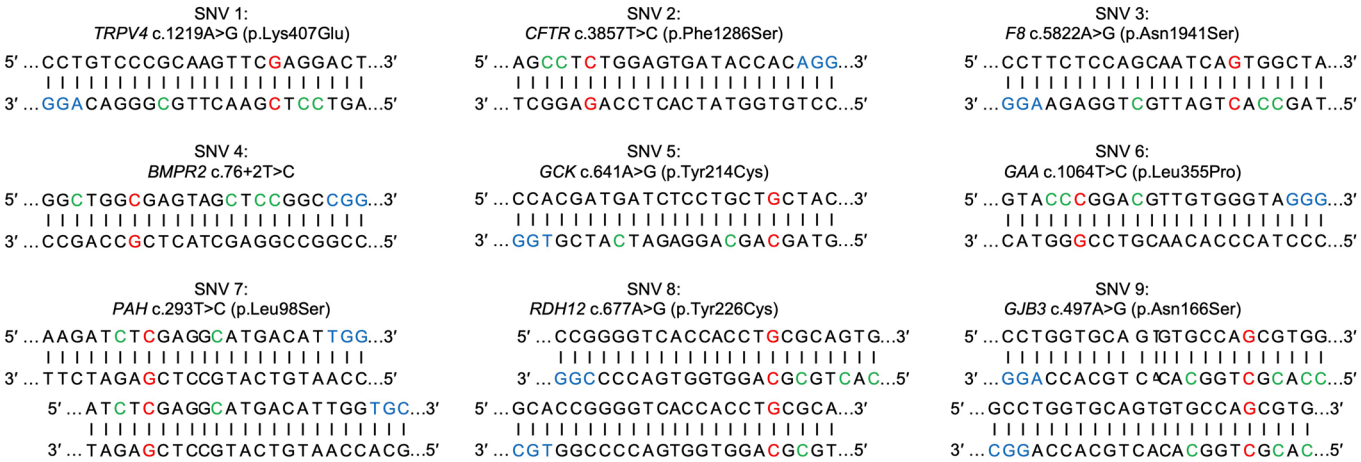
Extended Data Fig. 5 | Protein expression levels of CBE variants. Western blot showing the protein expression of the indicated CBE variants in HEK293T cells.

https://doi.org/10.1038/s41589-024-01595-4

a

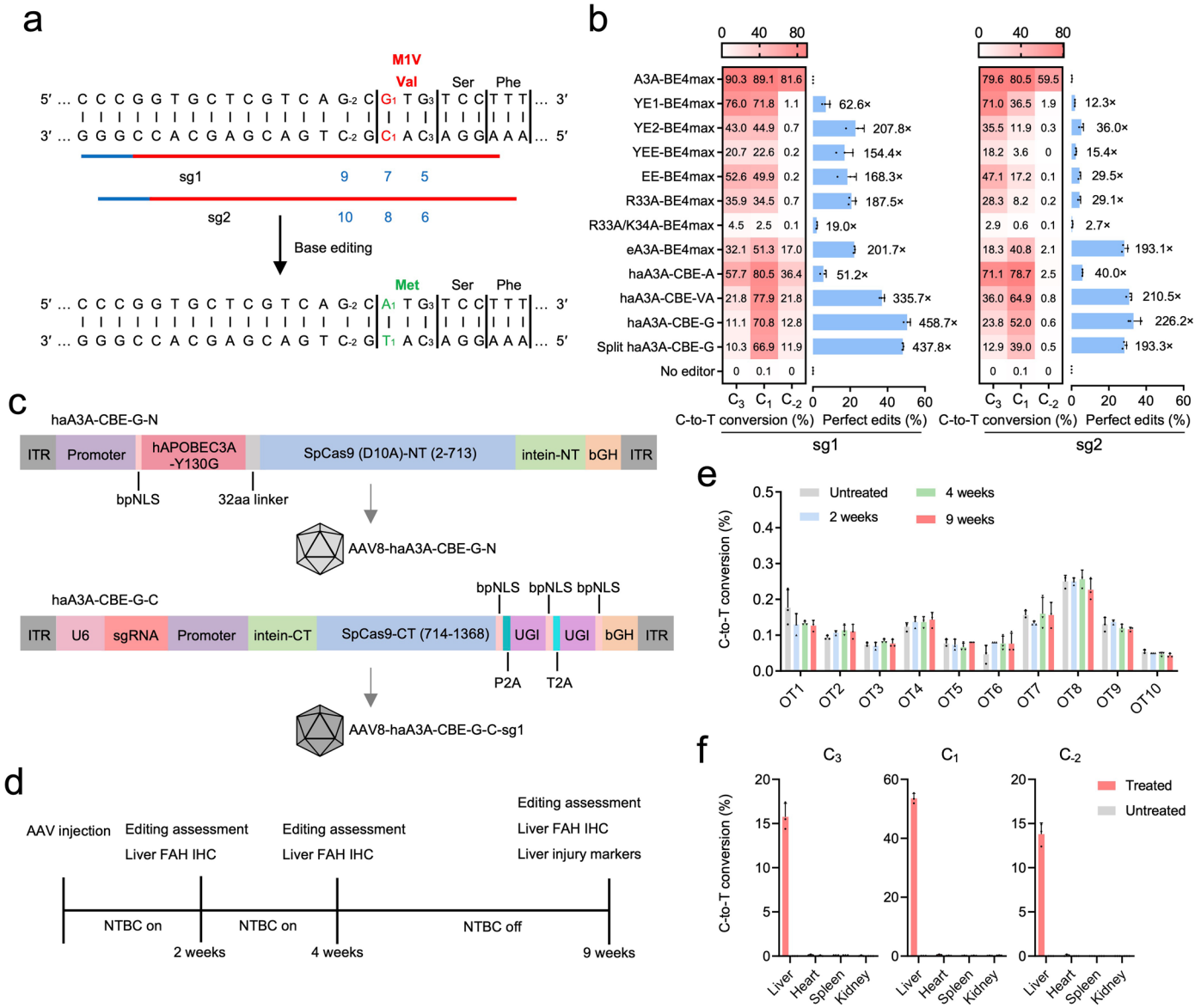
No.	Gene	Genotype	Variation ID	Associated genetic disease
SNV 1	*TRPV4*	c.1219A>G(p.Lys407Glu)	126463	Metatropic dysplasia
SNV 2	*CFTR*	c.3857T>C(p.Phe1286Ser)	7184	Cystic fibrosis
SNV 3	*F8*	c.5822A>G(p.Asn1941Ser)	10294	Hemophilia A
SNV 4	*BMPR2*	c.76+2T>C	425689	Primary pulmonary hypertension
SNV 5	*GCK*	c.641A>G(p.Tyr214Cys)	16144	Severe persistent hyperinsulinemic hypoglycemia
SNV 6	*GAA*	c.1064T>C(p.Leu355Pro)	284093	Glycogen storage disease type II (GSDII, Pompe disease)
SNV 7	*PAH*	c.293T>C(p.Leu98Ser)	627	Mild non-PKU hyperphenylalaninemia
SNV 8	*RDH12*	c.677A>G(p.Tyr226Cys)	2046	Childhood-onset severe retinal dystrophy
SNV 9	*GJB3*	c.497A>G(p.Asn166Ser)	6492	Deafness, digenic, GJB2/GJB3

b



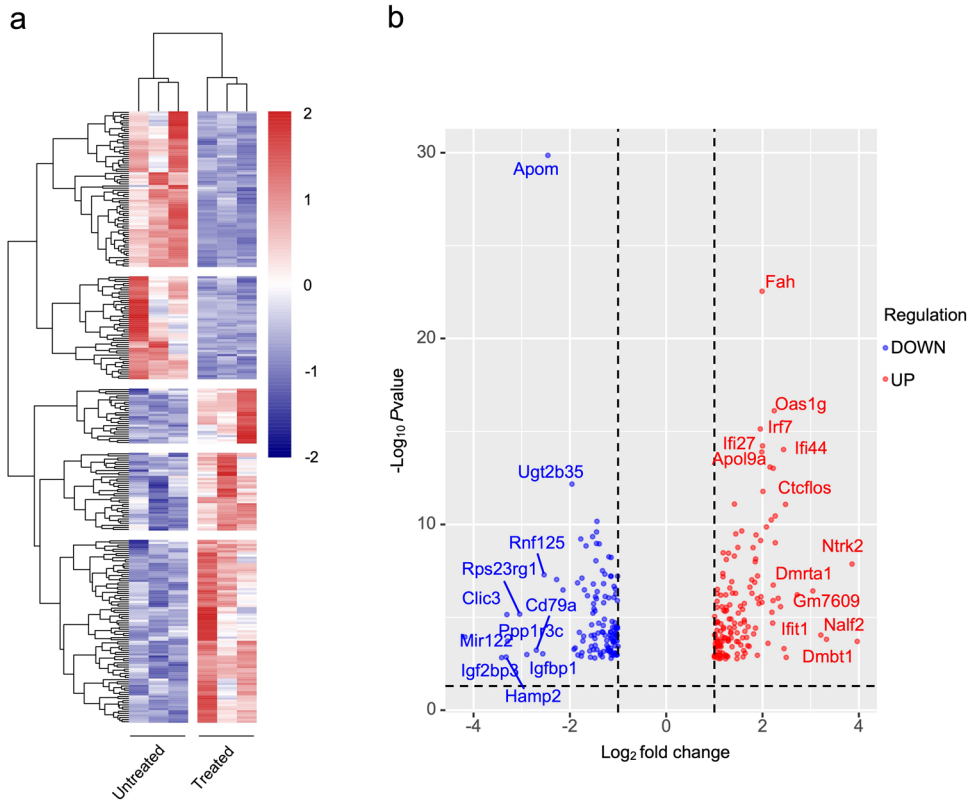
Extended Data Fig. 6 | Correction of human pathogenic SNVs. a, Information of nine human pathogenic SNVs caused by T-to-C or A-to-G mutations from ClinVar database. **b**, Sequences represent the protospacers and PAMs (blue), pathogenic SNVs are labeled in red, and bystander Cs edited by A3A-BE4amx are labeled in green. For SNV 7, SNV8 and SNV9, upper sequences represent the protospacers with NGG PAMs, and lower sequences represent the protospacers with NG PAMs.

https://doi.org/10.1038/s41589-024-01595-4



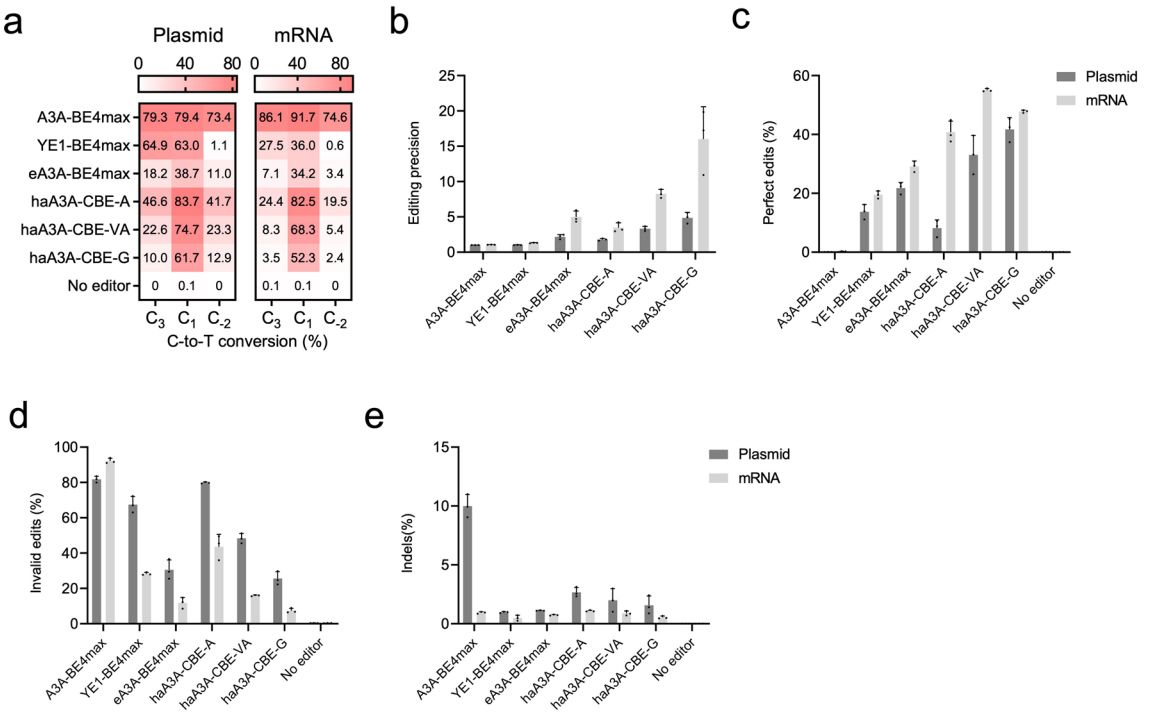
Extended Data Fig. 7 | Efficient and precise correction of the disease-causing mutation in *Fah*$^{NS/NS}$ mice. a, Diagram of two sgRNAs (sg1 and sg2) designed to target the exon 1 of *Fah*$^{NS/NS}$ allele. PAMs are indicated by blue lines and protospacers by red lines. Blue numbers denote the positions of the individual bases within the protospacers, while the subscript numbers denote the base positions relative to the first base of the CDS. The disease-causing mutation1A>G (p.M1 > V) is shown in red. The C_1-to-T_1 (shown in green) conversion in the non-coding strand leads to the desired G_1-to-A_1 (shown in green) conversion in the coding strand, which restores the start codon for methionine (shown in green). **b**, The editing frequencies and correction rates of the *Fah*$^{NS/NS}$ allele by CBE variants in HEK293T reporter cells. Heatmaps represent the average editing frequencies of individual cytosines from three biologically independent replicates. Histograms represent the frequencies of perfect edits (desired C_1-to-T_1 edits without any other edits) of the *Fah*$^{NS/NS}$ allele by CBE variants. The numbers beside bars display fold changes for accurate CBE variants in creating desired G_1-to-A_1 (without bystander edits) edits compared with A3A-BE4max. **c**, Schematic view of intein-mediated split haA3A-CBE-G AAV constructs (haA3A-CBE-G-N and haA3A-CBE-G-C). ITR, inverted terminal repeat; NLS, nuclear localization signal; bGH, bovine growth hormone poly (A) signal; U6, RNA polymerase III promoter for human U6 snRNA; P2A and T2A, 2 A peptide from porcine teschovirus-1 polyprotein and *Thosea asigna* virus capsid protein, respectively. **d**, The experimental scheme for in vivo base editing. **e**, Editing frequencies of computationally predicted off-target loci of sg1. **f**, The editing frequencies of the *Fah*$^{NS/NS}$ allele in different tissues from untreated and treated mice 9 weeks postinjection. Dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates (**b**) or three mice (**e** and **f**).

https://doi.org/10.1038/s41589-024-01595-4



Extended Data Fig. 8 | Transcriptomic changes following AAV delivery of haA3A-CBE-G in *Fah*$^{NS/NS}$ mice. a, Heatmap of 273 differentially expressed genes amongst untreated and treated *Fah*$^{NS/NS}$ mice. **b**, Volcano plot showing fold-change and p-value of genes up-regulated (red) and down-regulated (blue) in treated compared to untreated *Fah*$^{NS/NS}$ mice. *P* values were obtained using DESeq2 (Methods).

https://doi.org/10.1038/s41589-024-01595-4



Extended Data Fig. 9 | Editing profiles of CBE variants delivered by plasmid or mRNA in HEK293T reporter cells. a, Editing frequencies for CBE variants delivered by plasmid or mRNA in HEK293T reporter cells. In heatmaps, editing efficiencies shown represent the mean of three biologically independent replicates. **b**, Editing precision of each CBE variant delivered by plasmid or mRNA. Frequencies of perfect edits (**c**), invalid edits (**d**) and indels (**e**) of *Fah*$^{NS/NS}$ allele induced by CBE variants in HEK293T reporter cells. Dots represent individual values, and bars represent mean ± s.d. of three independent biological replicates.

nature portfolio

Corresponding author(s): Dali Li

Last updated by author(s): Feb 2, 2024

Reporting Summary

Nature Portfolio 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 Portfolio 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 amplicons sequencing data were collected and demultiplexed by an illumina HiSeq X ten instrument. RNA-seq and WGS data were collected and demultiplexed by an illumina NovaSeq 6000 instrument. Histology images were acquired using Olympus BX53. qPCR data was collected using QuantStudio3 (Applied Biosystems). For western blot experiment, the membrane was imaged using an Odyssey CLx Imager (LI-COR). FACS gating data were colleted 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/#!) (PMID: 30587106) for base editing and indels efficiencies. Potential DNA off-target sites for indicated base editors were predicated using Benchling (https://benchling.com). RNA-seq data were analyzed using Trimmomatic (version 0.39), STAR (version 2.7.10a), GATK (version 4.2.6.1), FastQC tool (v.0.11.8), RSeQC (v.4.0.0), HiSAT2 (v.2.2.1), featureCounts (v.2.0.4), and DESeq2 (v.1.38.1) software. GraphPad Prism 9.3 was also used to analyze data. BWA (v.0.7.17-r1188), Samtools (v.1.14), Picardtools (gatk, v.4.2.6.1), and Varscan (v2.4.6) were used to analyses the WGS data.

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

HTS data has been deposited in the NCBI Sequence Read Archive (SRA) database under the accession codes PRJNA938162, PRJNA938284, PRJNA938523, PRJNA938572, PRJNA938580, PRJNA1032216. RNA-seq data and HTS data of A3A hotspots have been deposited in the SRA database under the accession code PRJNA938578. WGS data have been deposited in the SRA database under the accession code PRJNA1042830. Data for mouse liver treated with CBE have already been posted in the SRA database under the accession code PRJNA937584. The hg38 reference genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40/) was used for alignment in the analysis of RNA-seq and WGS data form HEK293T cells. The mm10 reference genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001635.26/) was utilized for the analysis of RNA-seq data from mice. ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) was employed to identify pathogenic SNVs that can be correct by cytosine base editing. The published structure of A3A (PDB ID 5SWW) can be accessed here: https://www.rcsb.org/structure/5SWW. There are no restrictions on data availability.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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

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. We and others determined this sample size to be sufficient in mammalian cell gene editing experiments to yield reproducible mean values (PMID: 36357717 and 36357719). For Figure 1b, two replicates were used to screen the candidates with reduced off-target activity, then three replicates were used to further characterize these candidates.
Data exclusions	No data were excluded from the analysis.
Replication	Three independent biological replicates were performed on different days unless stated otherwise. 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
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	A 1:5000 dilution of anti-β-Actin (Sigma, A5441), a 1:2,000 dilution of anti-CRISPR-Cas9 antibody (Abcam, ab189380), and a 1:2,000 dilution of anti-FAH antibody (AbboMax, 602-910) was uesd.
Validation	The anti-β-Actin antibody has been validated by western blot in NIH-3T3 cell lysate (https://www.sigmaaldrich.cn/CN/zh/product/sigma/a5441). The anti-CRISPR-Cas9 antibody has been validated by western blot in previously published studies (PMID: 36357717).The anti-FAH antibody has been validated by western blot in previously published studies (PMID: 32413280).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T cells (purchased from ATCC CRL-3216) and Hela cells (purchased from ATCC CCL-2).
Authentication	Cells were authenticated by the supplier by STR analysis.
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 research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	The C57BL/6J mice were purchased as adults from Shanghai Laboratory Animal Center. The FahNS/NS mice were obtained as previously described (PMID: 32413280), and bred and maintained in East China Normal University Center for Animal Research. All mice were housed in a specific pathogen-free facility on 12-h light/dark cycles with a temperature between 20°C and 22°C and humidity of 40-60%. All mice had ad libitum access to food and water in standard cages. FahNS/NS mice were maintained on NTBC water (10 mg/L) to prevent liver damage. 6-week-old HTI mice were injected with AAV particles. 2, 4 or 9 weeks post injection, HTI mice were euthanized for analyses. The untreated mice were humanely euthanized if over 20% of body weight was lost after NTBC withdrawal. 8-week-old wild type C57BL/6J mice were used to determine Fah gene expression levels and serum AST, ALT and TBIL levels.
Wild animals	The study did not involve wild animals.
Reporting on sex	Both male and female animals were used for homozygous experiments as the phenotypeis highly penetrant for both sexes.
Field-collected samples	The study did not involve Field-collected samples.
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.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

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	Cells were seeded onto 24-well plates or 10-cm dishes and transfected with plasmids at 80% confluence using polyethylenimine following the manufacturer's recommended protocol. Three or two days after transfection, GFP-positive cells were sorted on FACSAria III and genomic DNA or RNA was extracted. Cells were resuspended in PBS and filtered through a 45μm cell strainer cap before sorting. One group of AAV injected HTI mice were euthanized by carbon dioxide asphyxiation at 2 weeks postinjection and another group received biopsies at 4 weeks postinjection for analyses. After biopsy, NTBC was withdrawn and mice were euthanized at 9 weeks after injection for analyses. Liver tissues form AAV injected HTI mice, untreated HTI mice or wild type C57BL/6J mice were subjected to DNA or RNA extraction.
Instrument	FACSAria III (BD Biosciences)
Software	BD FACSDiva Software Diva 8.0.2
Cell population abundance	The cells were subjected to FACS sorting step via GFP signals. The density of collected cells are about 500,000 cells. 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 signal were collected.

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