nature cell biology

TECHNICAL REPORT

https://doi.org/10.1038/s41556-020-0518-8

Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain

Xiaohui Zhang[1,5], Liang Chen[1,5], Biyun Zhu[1,5], Liren Wang[1], Caiyu Chen[1], Mengjia Hong[1], Yifan Huang[1], Huiying Li[1], Honghui Han[2], Bailian Cai[3], Weishi Yu[1,4], Shuming Yin[1], Lei Yang[1], Zuozhen Yang[4], Meizhen Liu[1], Ying Zhang[1], Zhiyong Mao[3], Yuxuan Wu[1], Mingyao Liu[1] and Dali Li[1]✉

Cytidine base editors are powerful genetic tools that catalyse cytidine to thymidine conversion at specific genomic loci, and further improvement of the editing range and efficiency is critical for their broader applications. Through insertion of a non-sequence-specific single-stranded DNA-binding domain from Rad51 protein between Cas9 nickase and the deaminases, serial hyper cytidine base editors were generated with substantially increased activity and an expanded editing window towards the protospacer adjacent motif in both cell lines and mouse embryos. Additionally, hyeA3A-BE4max selectively catalysed cytidine conversion in TC motifs with a broader editing range and much higher activity (up to 257-fold) compared with eA3A-BE4max. Moreover, hyeA3A-BE4max specifically generated a C-to-T conversion without inducing bystander mutations in the haemoglobin gamma gene promoter to mimic a naturally occurring genetic variant for amelioration of β-haemoglobinopathy, suggesting the therapeutic potential of the improved base editors.

As most known genetic diseases are caused by point mutations[1], nucleotide substitution is a critical technology for generating desired mutations for disease modelling and more importantly for therapeutic correction of pathogenic variants[2,3]. Although the CRISPR–Cas9 system can efficiently generate double-stranded breaks that stimulate the homology-directed repair pathway in the presence of donor templates, precise genome editing through Cas9-stimulated homology-directed repair is inefficient, especially in non-dividing somatic cells[4,5]. Several groups have shown that fusion of cytidine or adenosine deaminases with DNA recognition modules, such as Cas9 nickase (Cas9n) or Cpf1, efficiently generates site-specific C·G-to-T·A or A·T-to-G·C conversions[2,3,6,7].

Through fusion of cytidine deaminases, such as rAPOBEC1 (ref. [3]), hAID[8], hAPOBEC3A[9–11] or PmCDA1 (refs. [12–14]) at the amino (N) or carboxy (C) terminus of Cas9n (Cas9 D10A), cytidine base editors (CBEs) were developed to efficiently generate C-to-T transitions. Their efficiency was further enhanced by introducing a uracil glycosylase inhibitor domain either through direct fusion to the C terminus or separately delivered with CBEs[13,15]. Similarly, an adenosine base editor was developed through conjugation of Cas9n with TadA mutants, derived from an *Escherichia coli* transfer RNA adenosine deaminase ecTadA, to deaminate adenosine in DNA backbones to generate A·T-to-G·C substitutions[2]. The base editors directly catalyse deamination to achieve nucleotide conversion without generating double-stranded breaks, thus minimizing the formation of undesired mutations such as indels, large deletions, translocations or DNA rearrangements[7]. These features make them promising instruments in disease modelling and the improvement of production traits in crops and livestock and, more importantly, for gene therapy[7]. However, the base editors generate base conversions in a relatively narrow window, which limits their targeting scope.

The editing window of CBEs (such as BE3, BE4 or BE4max) is typically located 4–8 nucleotides downstream from the 5′ end of the targeted sequence distal to the protospacer adjacent motif (PAM) site[3,13,16]. Base editors could only edit the nucleotides within this window, which is determined by the intrinsic feature of distinct deaminases. Great efforts have been made to either reduce or increase the editing window to generate versatile base editors for various purposes[17,18]. Three BE3s have been developed with the editing window within three nucleotides, but their activity is either comparable or reduced compared with BE3 (ref. [17]). Through structure-based design and modifications of hAPOBEC3A, eA3A-BE3 has been developed to selectively generate C-to-T conversions within the TC motifs but not in a non-cognate bystander motif[10]. In contrast, base editors with increased targeting range are also valuable to generate more genotypes for saturation mutagenesis, disrupt gene function or correct genetic mutations for gene therapy. Through fusion of cytidine and adenosine deaminases with circularly permuted Cas9 variants, a series of CP-CBEmax and CP-ABEmax variants was generated with almost twice the editing window and reduced byproduct formation[18]. However, these innovative modifications did not improve the editing efficiency. Several CBE variants with enhanced base editing activity have been developed through codon optimization, nuclear localization signal modification[16,19] or molecular evolution through phage-assisted continuous evolution of distinct cytidine deaminases[20], but they are inefficient at catalysing cytidines adjacent to the PAM site. Although expanding the editing window has been achieved via recruiting the engineered deaminase through single guide RNA (sgRNA) tethering (CRISPR-X)[21] or protein–protein

[1]Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China. [2]Bioray Laboratories, Shanghai, China. [3]Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China. [4]Cipher Gene, Beijing, China. [5]These authors contributed equally: Xiaohui Zhang, Liang Chen, Biyun Zhu. ✉e-mail: dlli@bio.ecnu.edu.cn

interaction modules (BE-PLUS)[22], the activity of these CBEs remains limited, which hinders broader applications.

Here, we demonstrate that fusion of the single-stranded DNA-binding domain (ssDBD) from Rad51 (a critical protein involved in DNA repair), between the cytidine deaminases and Cas9n, dramatically enhanced the editing efficiency of BE4max, A3A-BE4max and eA3A-BE4max without increasing indel activity, cellular toxicity or RNA- or sgRNA-dependent DNA off-target effects. These engineered CBEs (hyper CBEs (hyCBEs)) exhibited enhanced activity in both cell lines and mouse embryos, demonstrating their broader applicability in disease modelling and gene therapy.

Results

Screen of ssDBDs to increase BE4max activity. Since cytidine deaminase mainly catalyses C-to-T conversion on the single-stranded DNA (ssDNA) substrate generated by Cas9n, we hypothesized that fusion of a non-sequence-specific ssDBD might increase the binding affinity and editing activity (Fig. 1a). To test our hypothesis, ten ssDBDs belonging to four structural topologies from different human proteins, including RPA70, RPA32, BRCA2, the HNRNPK KH domain, the PUF60 RNA recognition motif and the DNA-binding domain (DBD) from Rad51 (Supplementary Information)[23], were cloned and fused to the N terminus of BE4max, and the C-to-T conversion activities were determined at two endogenous sites in HEK293T cells (Fig. 1b,c). High-throughput sequencing (HTS) demonstrated that two ssDBDs—Rad51DBD and RPA70-C—enhanced the activity of BE4max (Fig. 1c). Given the better performance, Rad51DBD was selected and either inserted between APOBEC1 and Cas9n (Rad51DBDm-BE4max) or fused to the C terminus of BE4max to estimate the optimal position for ssDBD fusion. Interestingly, the base editing efficiency was dramatically increased when Rad51DBD was inserted in the middle. Thus, Rad51DBDm-BE4max is named hyper BE4max (hyBE4max) hereafter (Fig. 1c). Further examination of the total ten targets showed that the activity of hyBE4max in the conventional editing window (C4–C8) was higher than that of BE4max (average: ~15.4–69.6 versus 10.2–47.7%). Also, it exhibited much higher activity at positions (C9–C15) proximal to the PAM site (average: 15.35–44.7% for hyBE4max and 4.7–22% for BE4max), and up to an 18-fold increase was observed at C11 on the TIM3-sg1 site (Fig. 1d,e). These data suggest that the major editing window of hyBE4max is extended from C4–C8 to C4–C12 (Fig. 1d,e), similar to that of CP-CBEmax (C4–C11)[18]. As CBEmax and CP-CBEmax have similar editing activity[18], our data suggest that hyBE4max has higher activity than CP-CBEmax. However, hyBE4max was not efficient at editing cytidine in a **GC** context (C11 in CDK10-sg1 and C13 in HPRT1-sg6) (Fig. 1d), similar to BE4max[20]. Meanwhile, hyBE4max retained a very low indel rate similar to BE4max (Fig. 1f).

Fusion of Rad51DBD dramatically increases the activity and editing window of A3A-BE4max. A3A-BE3 was generated through fusion of Cas9n with APOBEC3A, which is a very active cytidine deaminase variant in human cells and plants[9–11]. HyA3A-BE4max was generated and its efficiency was compared with A3A-BE4max[20] at ten endogenous sites in HEK293T cells (Fig. 2a). A3A-BE4max was very efficient at positions C3–C11, with an activity ranging from 22.5–44.7%, but the activity decreased from C12 towards the PAM site with an efficiency from 2.7–27.1% (Fig. 2a,b). HyA3A-BE4max exhibited a substantial (1.2- to 2-fold) increase in activity from C3–C11 (14.5–70.3 versus 22.5–44.7%) and a more dramatic 3.1- to 4.1-fold elevation at positions C12–C17 compared with A3A-BE4max (~6.6–62.4 versus 2.7–27.1%) on average (Fig. 2a,b). The major editing window of hyA3A-BE4max was extended to C3–C15 and hyA3A-BE4max was even active at C17 and C21 on some targets such as EGFR-sg5 (Fig. 2a). Similar indel rates between A3A-BE4max and hyA3A-BE4max were observed in the above ten targets (Fig. 2c). A previous study reported that A3A-BE3 exhibited better performance on methylated cytidines than other cytidine deaminases[9]. Investigation of the reported methylated targets[9] showed that the activity of hyA3A-BE4max was higher at the positions proximal to the PAM site than A3A-BE4max (Fig. 2d). These data show that ssDBD fusion also increased A3A-BE4max activity, suggesting that the fusion strategy might be a versatile method to enhance the performance of CBE variants. Considering that hyA3A-BE4max has a broader editing window and higher efficiency than previously reported tools such as CRISPR-X[21] and targeted activation-induced cytidine deaminase-mediated mutagenesis[8], it may be used to generate genome diversification for protein evolution or drug resistance mutation screens.

Efficient base editing of hyA3A-BE4max in mouse embryos. To test the activity of hyA3A-BE4max in vivo, hyA3A-BE4max messenger RNA (mRNA) and sgRNA were injected into mouse embryos to create a premature stop codon in the *Dystrophin* gene to generate a Duchenne muscular dystrophy (DMD) model. The desired mutation converts CAA into a TAA stop codon at position C10 in the target site (Fig. 3a). HTS of all the F0 mice receiving A3A-BE4max or hyA3A-BE4max with DMD-sg3 injection revealed that 91% (10/11 for A3A-BE4max) and 100% (10/10 for hyA3A-BE4max) of the F0 mice carried at least one C-to-T mutation at the target site (Fig. 3b and Extended Data Fig. 1a,b), suggesting that these two types of A3A-BE4max were very efficient. However, six of ten F0 mice (60%) harboured the desired C-to-T (C10 in DMD-sg3) homozygous nonsense mutation after hyA3A-BE4max injection, but no homozygous mice were obtained in the A3A-BE4max group (Supplementary Table 1). Moreover, a significantly higher editing efficiency to create the stop codon was detected in hyA3A-BE4max-treated pups (Fig. 3c, $P=0.0017$). The disrupted expression of *Dystrophin* in the homozygous founder BD03 was confirmed by immunostaining (Fig. 3d). Next, we examined off-target effects of hyA3A-BE4max through HTS of 15 off-target sites predicted by Cas-OFFinder in three homozygous founders. No off-target mutations were detected, indicating its accuracy in animal embryos (Extended Data Fig. 2a). Moreover, the mutations were efficiently transmitted to the F1 generation (Extended Data Fig. 2b). These results showed that hyA3A-BE4max is more efficient than A3A-BE4max in vivo, especially for cytidines proximal to the PAM, suggesting it is a powerful tool for disease model generation.

Fusion of Rad51DBD to eA3A-BE4max substantially promotes its efficiency and editing range with minimized bystander activity. HyA3A-BE4max is super active within a broader editing window, but accurate correction of point mutations is essential in circumstances where bystander mutations cannot be tolerated, such as in human gene therapy[10]. eA3A-BE3 containing a mutation (N57G) in hAPOBEC3A preferentially catalyses C-to-T conversion in TC motifs according to a TCR > TCY > VCN hierarchy, and provided up to 264-fold higher editing of cognate motifs than bystander motifs[10]. Therefore, we introduced the N57G mutation into hyA3A-BE4max to generate hyeA3A-BE4max and examined its editing activity at 11 endogenous sites in HEK293T cells. Both eA3A-BE4max and hyeA3A-BE4max exhibited a high preference for TC motifs with greatly reduced activity at bystander cytidines compared with A3A-BE4max (Fig. 4a and Extended Data Fig. 3a). Among the targets that were efficiently edited (>20%) by eA3A-BE4max within the major editing window (C4–C9), hyeA3A-BE4max showed a 1.4- to 2.8-fold increase, with the editing efficiency ranging from 46.6–78.4%. HyeA3A-BE4max had a more dramatic activity elevation at the cytidines outside the editing window (C10–C15) proximal to the PAM site compared with eA3A-BE4max. On average, the activity of hyeA3A-BE4max (efficiency: 7.6–79.1%) increased 1.7- to 15.2-fold compared with eA3A-BE4max (efficiency: 0.5–25.7%) (Fig. 4a b).

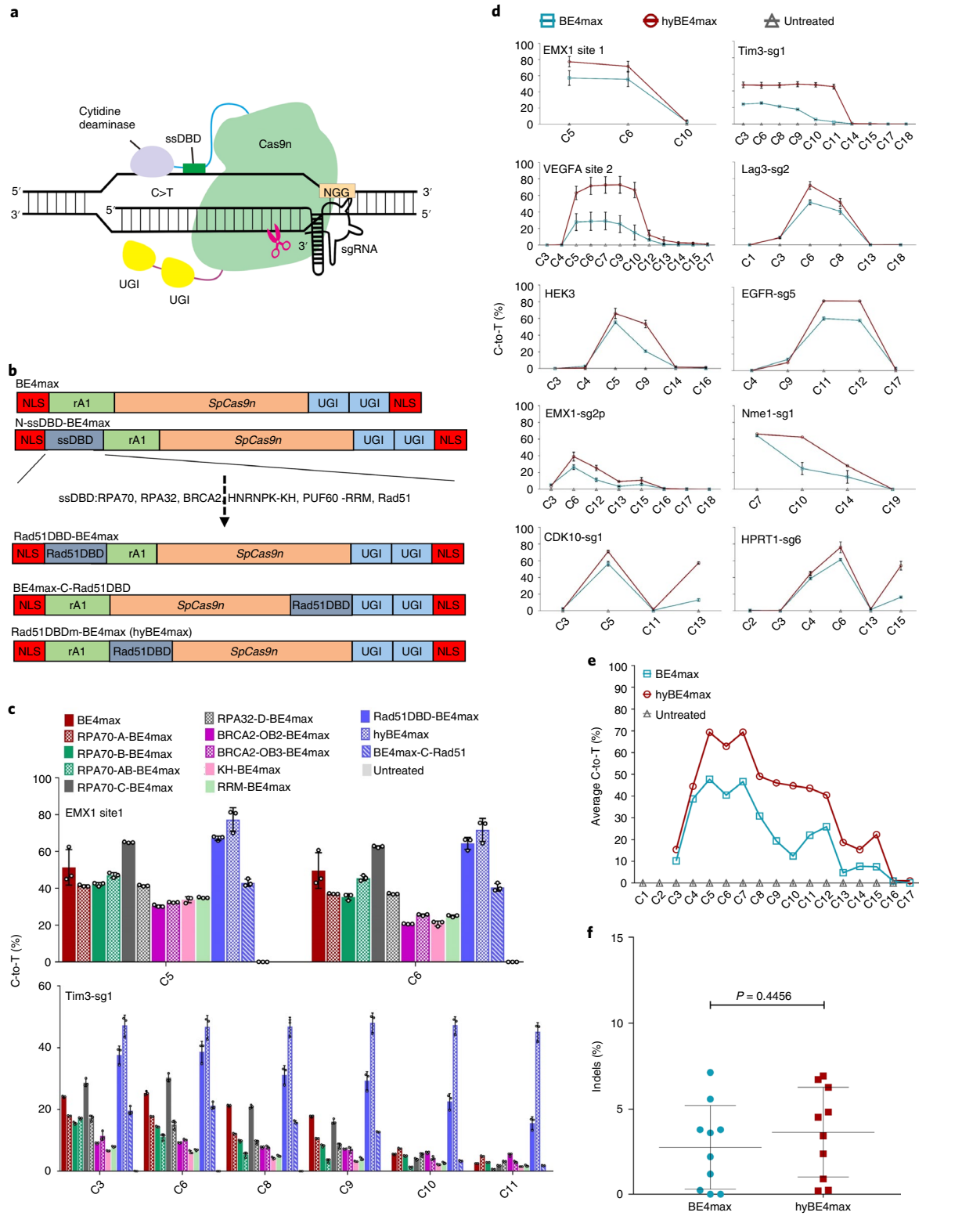


Fig. 1 | Design and optimization of BE4max through fusion of an ssDBD. a, Schematic of the ssDBD fusion strategy to increase CBE activity. UGI, uracil glycosylase inhibitor. **b**, Schematics of the constructs with different ssDBDs fused to BE4max. rA1, rat Apobec1; NLS, nuclear localization signal. **c**, Base editing efficiency of BE4max fused with variant ssDBDs. HEK293T cells were transfected with the indicated constructs, and editing efficiency was determined by HTS. Data are means ± s.d. ($n = 3$ independent experiments). **d**, Comparison of the base editing efficiency of BE4max and hyBE4max at ten endogenous genomic loci in HEK293T cells. Data are means ± s.d. ($n = 3$ independent experiments). **e**, Average C-to-T editing efficiency at different positions in the same ten targets shown in **d**, edited by BE4max or hyBE4max. Data represent means from three independent experiments. **f**, Frequency of indel formation by BE4max and hyBE4max at the same ten targets as shown in **d**. Each data point represents the average indel frequency at each target site, as calculated from three independent experiments. The error bars and *P* value were derived from these ten data points. Data are means ± s.d. The *P* value was determined by two-tailed Student's *t*-test. Statistical source data are provided with the paper.

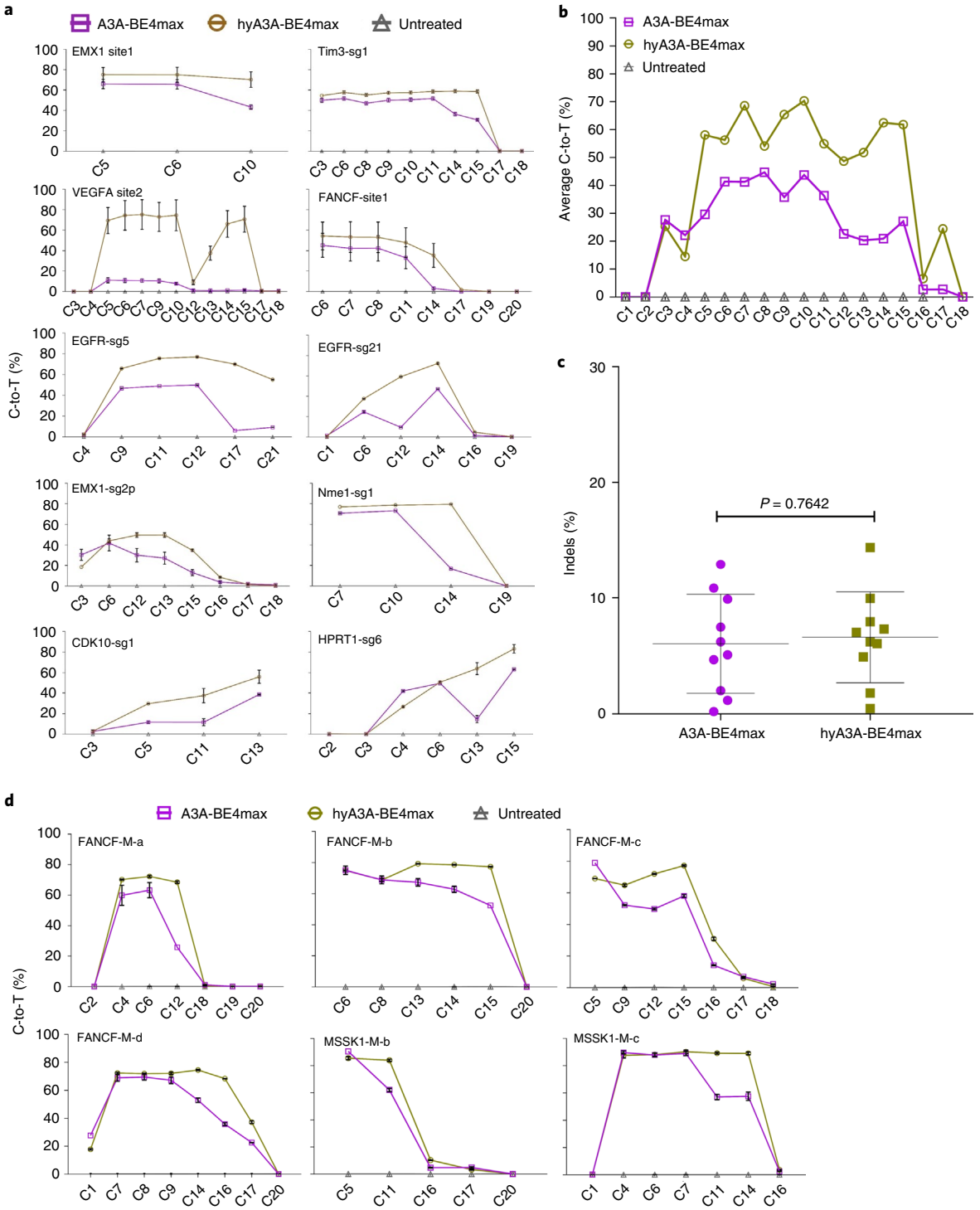


Fig. 2 | Characterization of hyA3A-BE4max in HEK293T cells. a, Comparison of the base editing efficiency of A3A-BE4max and hyA3A-BE4max at ten endogenous genomic loci in HEK293T cells. Data are means ± s.d. ($n = 3$ independent experiments). **b**, Average C-to-T base editing efficiency at different positions in the same ten targets shown in **a**, edited by A3A-BE4max or hyA3A-BE4max. Data represent means from three independent experiments. **c**, Frequency of indel formation by A3A-BE4max and hyA3A-BE4max at the same ten targets in **a**. Each data point represents the average indel frequency at each target site, as calculated from three independent experiments. The error bars and *P* value were derived from these ten data points. Data are means ± s.d. The *P* value was determined by two-tailed Student's *t*-test. **d**, Base editing efficiency of A3A-BE4max and hyA3A-BE4max at six endogenous loci with high levels of DNA methylation in HEK293T cells. Values and error bars reflect the means and s.d. of three independent experiments. Statistical source data are provided with the paper.

Unexpectedly, hyeA3A-BE4max was able to edit some loci that were difficult to mutate by eA3A-BE4max (Fig. 4a, bottom), with an increase up to 257-fold (C10 in EGFR-sg26). In the target site positions C11–C15, hyeA3A-BE4max was very active on the cytidines in the TCR (A/G) motif (41–83%) but not in other motifs such as TCC and TCT (Fig. 4a). These data suggest that the base editing

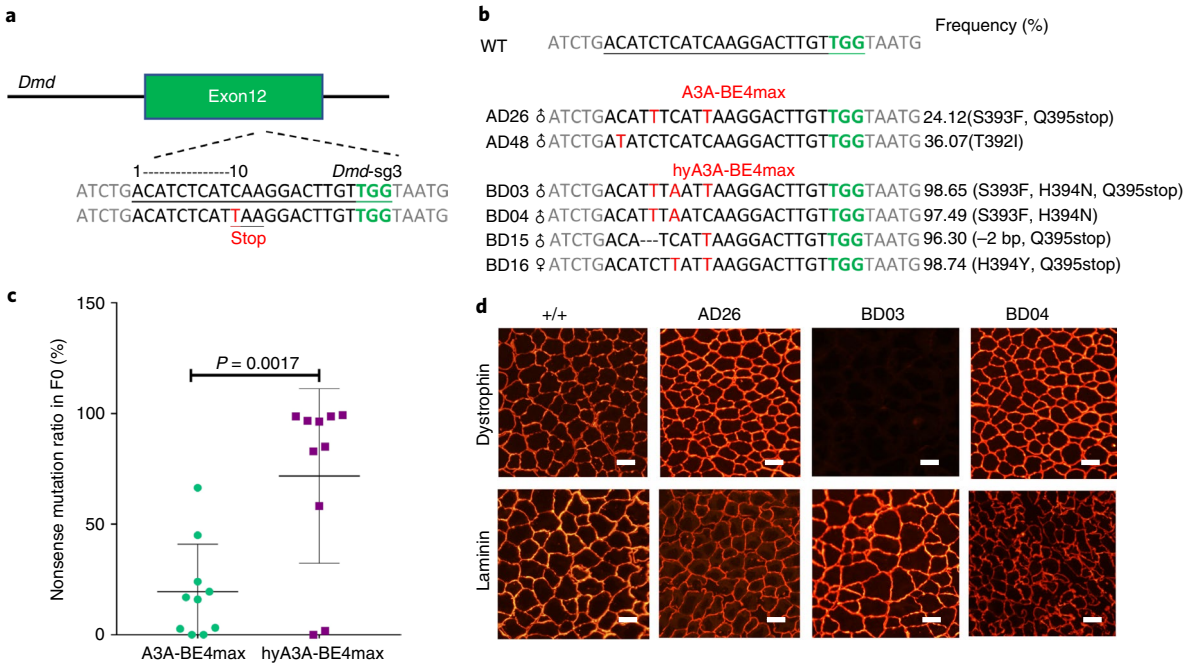


Fig. 3 | Highly efficient base editing by hyA3A-BE4max in mouse embryos. a, The target sequence in the dystrophin (*Dmd*) exon 12 locus. The PAM sequence and sgRNA target sequence are shown in green and bold, respectively. The desired C-to-T (C10) conversion to create the premature stop codon TAA is shown in red. **b**, Genotyping of representative F0 generation pups by A3A-BE4max and hyA3A-BE4max. The frequencies of wild-type (WT) and mutant alleles were determined by analysing HTS using BE-Analyzer. The percentage values on the right represent the frequencies of the indicated mutant alleles, with the corresponding mutation-induced amino acid conversions shown in parentheses. The frequency of the wild-type allele was omitted. **c**, Ratio of HTS reads containing the TAA stop codon of each founder that received A3A-BE4max ($n=10$ mice) or hyA3A-BE4max ($n=10$ mice) injection. Data are means ± s.d. The *P* value was determined by two-tailed Student's *t*-test. **d**, Immunofluorescence staining of tibialis anterior muscle tissue using antibodies against dystrophin and laminin from 6-week-old wild-type and founder (AD26, BD03 and BD04) mice. Histological analysis images are representative of three independent experiments. Scale bars: 50 μm. Statistical source data are provided with the paper.

efficiency of hyeA3A-BE4max is dramatically increased with fusion of Rad51DBD and maintains the specificity to edit cytidines within TC motifs (Fig. 4a,b). Like eA3A-BE4max, hyeA3A-BE4max generated indels at a very low rate (Fig. 4c).

We then performed off-target analysis through HTS of 50 potential off-target sites in total, including 24 predicted off-target sites from haemoglobin γ 1/2-117 (HBG1/2-117) sgRNA using the in silico Cas-OFFinder program[24] and 26 previously investigated off-target sites from EMX1 site 1 and FNACF site 1 using modified digenome sequencing[25]. The analysis did not detect significant off-target mutations in eA3A-BE4max- and hyeA3A-BE4max-treated groups, suggesting they were highly specific compared with A3A-BE4max (Fig. 5a–c). Base editors have been reported to edit numerous unpredictable off-target RNA sequences[26,27]. Consistent with a previous report[28], we confirmed that A3A-BE4max generated much less RNA off-target editing than BE4max, and eA3A-BE4 catalysed almost no cellular RNA substrates, suggesting that fusion of ssDBD to individual CBEs did not increase their RNA editing effects (Fig. 5d). Moreover, the higher activity of hyCBEs was not due to increased protein expression levels (Extended Data Fig. 3b). Importantly ssDBD fusion kept product purity high (Extended Data Fig. 4) and did not exhibit significant cellular or DNA toxicity (Fig. 5e,f).

Extremely high efficiency of hyeA3A-BE4max in mouse embryos to target a TC motif outside the conventional editing window. To investigate the efficacy of hyeA3A-BE4max at introducing mutations in TC motifs in vivo, we attempted to introduce a C-to-T conversion to create a premature stop codon in the *Dmd* gene (Fig. 6a). The target cytidine is at position C13 of the sgRNA outside the conventional editing window of eA3A-BE4max. After microinjection of the zygotes, 13 of 15 (86.7%) F0 mice receiving hyeA3A-BE4max contained the desired mutation (Fig. 6b and Supplementary Table 2). In contrast, in the eA3A-BE4max-treated group, only 1 of 16 (6%) F0 mice (with a 22.5% C-to-T editing efficiency) was born with the desired mutation (Fig. 6b and Supplementary Table 2). HTS analysis also showed that in hyeA3A-BE4max-treated mice, six out of 15 (40%) F0 mice were homozygous, and the average editing efficiency of the 15 founders was 81% (Fig. 6c). The absence of *Dmd* gene expression in the homozygous founders was confirmed by immunostaining (Fig. 6d). Through HTS analysis, no mutation was identified in 12 predicted off-target sites tested in three homozygous founders (Fig. 6e). Moreover, we performed whole-genome sequencing (WGS) analysis on wild-type and hyeA3A-BE4max-treated homozygous founder (DD11) mice following a reported method[29]. Among 175,058 NRG (R = A/G) PAM-containing sites bearing up to seven mismatches in the protospacer sequence, only one site was identified as a potential off-target site that was further determined as a false positive site through targeted NGS analysis of various tissues (Extended Data Fig. 5). These data showed that hyeA3A-BE4max has superior activity and specificity in C-to-T conversions in TC motifs outside the conventional activity window, suggesting that hyeA3A-BE4max expanded the targeting scope of eA3A-BE3 both in vitro and in vivo.

Efficient and accurate base conversion of hyeA3A-BE4max strongly activates γ-globin expression. We further examined whether hyeA3A-BE4max allows for precision base editing in an expanded target window for therapeutic purposes. β-haemoglobinopathies, such as β-thalassaemia and sickle cell disease, are caused by mutations in the haemoglobin β subunit

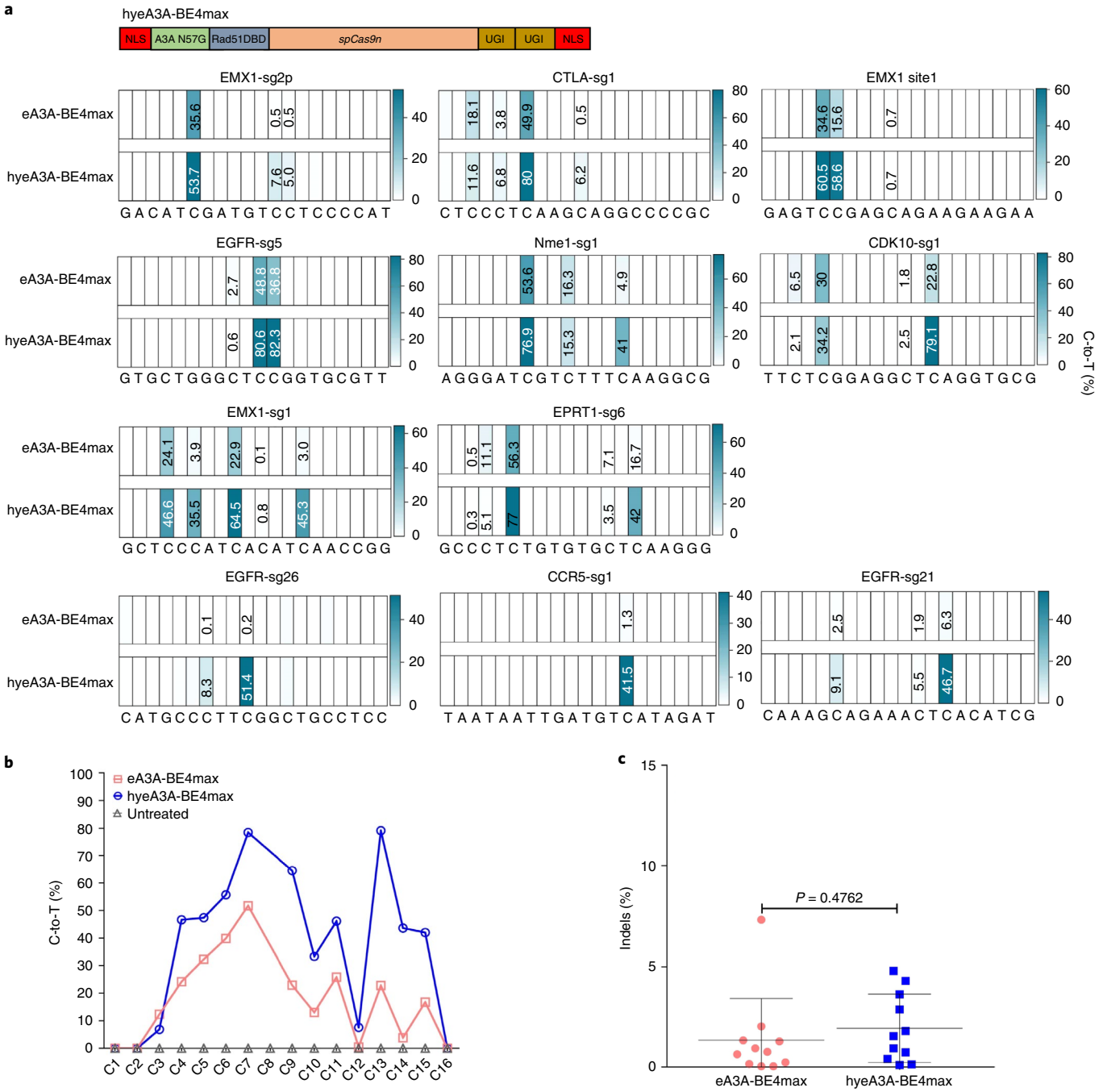


Fig. 4 | Characterization of hyeA3A-BE4max in HEK293T cells. a, The editing efficiencies of eA3A-BE4max and hyeA3A-BE4max were examined at 11 endogenous genomic loci in HEK293T cells. The average mutation percentage derived from three independent experiments of hyeA3A-BE4max and eA3A-BE4max at the same site is listed. **b**, Average C-to-T editing efficiency at different positions of the same 11 target sites shown in **a**. Data represent means from three independent experiments. **c**, Frequency of indel formation by eA3A-BE4max and hyeA3A-BE4max at the same 11 endogenous genomic loci shown in **a**. Each data point represents the average indel frequency at each target site, as calculated from three independent experiments. The error bars and *P* value were derived from these ten data points. Data are means ± s.d. The *P* value was determined by two-tailed Student's *t*-test. Statistical source data are provided with the paper.

gene (*HBB*; encoding adult β-globin) locus[30,31]. Reactivation of γ-globin to form foetal haemoglobin with α-globin to functionally substitute for adult haemoglobin is an attractive strategy to treat β-haemoglobinopathies, as demonstrated by several pre-clinical studies using CRISPR–Cas9 or a base editor to modify genomic elements and activate γ-globin genes[32–34]. It was reported that a heterozygous point mutation (G-to-A) 117 base pairs (bp) upstream (−117) of the ^{A}γ-globin gene produced 10–20% foetal haemoglobin in Greek patients with hereditary persistence of foetal haemoglobin[35] and this mutation disrupted the binding motif (TGACC) of BCL11A—a transcriptional repressor to silence HBG in adults[36]. We first examined whether A3A-BE4max and hyA3A-BE4max can generate a precise −117G-to-A mutation in the HBG promoter in HEK293T cells (Fig. 7a). A3A-BE4max generated G-to-A conversions at three sites (C3, C11 and C16, corresponding to −109, −117 and −122 in the HBG promoter), which were also edited

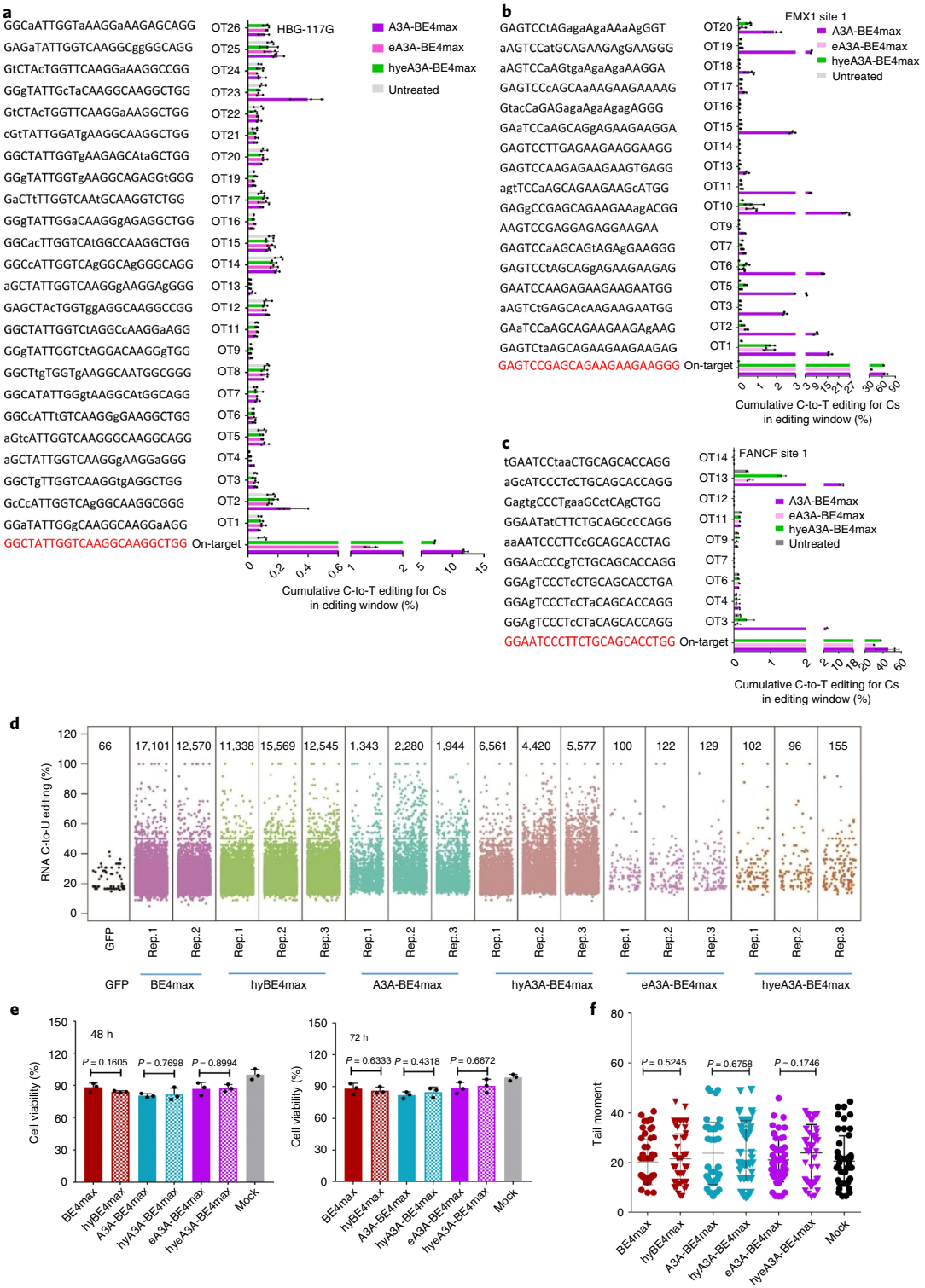


Fig. 5 | Investigation of the off-target editing activity and potential toxicity of hyCBEs. a–c, DNA on- and off-target analysis of the indicated targets (HBG-117, EMX1 site 1 and FANCF site 1) by A3A-BE4max, eA3A-BE4max and hyeA3A-BE4max in HEK293T cells. Data are means ± s.d. ($n = 3$ independent experiments). Mismatched nucleotides in off-targeting sequences are indicated in lowercase. **d**, Comparison of off-target RNA editing activity by BE4max versus hyBE4max, A3A-BE4max versus hyA3A-BE4max and eA3A-BE4max versus hyeA3A-BE4max. Jitter plots from RNA-Seq experiments in HEK293T cells show the efficiencies of C-to-U conversions (*y* axis) with BE4max, hyBE4max, A3A-BE4max, hyA3A-BE4max, eA3A-BE4max and hyeA3A-BE4max expression or a GFP negative control. Total numbers of modified bases are listed on the top. Each biological replicate (rep.) is listed on the bottom. **e**, Comparison of cell viability induced by BE4max, hyBE4max, A3A-BE4max, hyA3A-BE4max, eA3A-BE4max and hyeA3A-BE4max using the MTS assay 48 or 72 h after transfection in HEK293T cells. Data are means ± s.d. ($n = 3$ independent experiments). *P* values were determined by two-tailed Student's *t*-test. **f**, Analysis of DNA damage induced by BE4max, hyBE4max, A3A-BE4max, hyA3A-BE4max, eA3A-BE4max and hyeA3A-BE4max 72 h after transfection in HEK293T cells. Comet assay results show the tail moments of cells ($n = 50$ cells from three independent experiments) transfected with the indicated base editors. Data are means ± s.d. *P* values were determined by two-tailed Student's *t*-test. Statistical source data are provided with the paper.

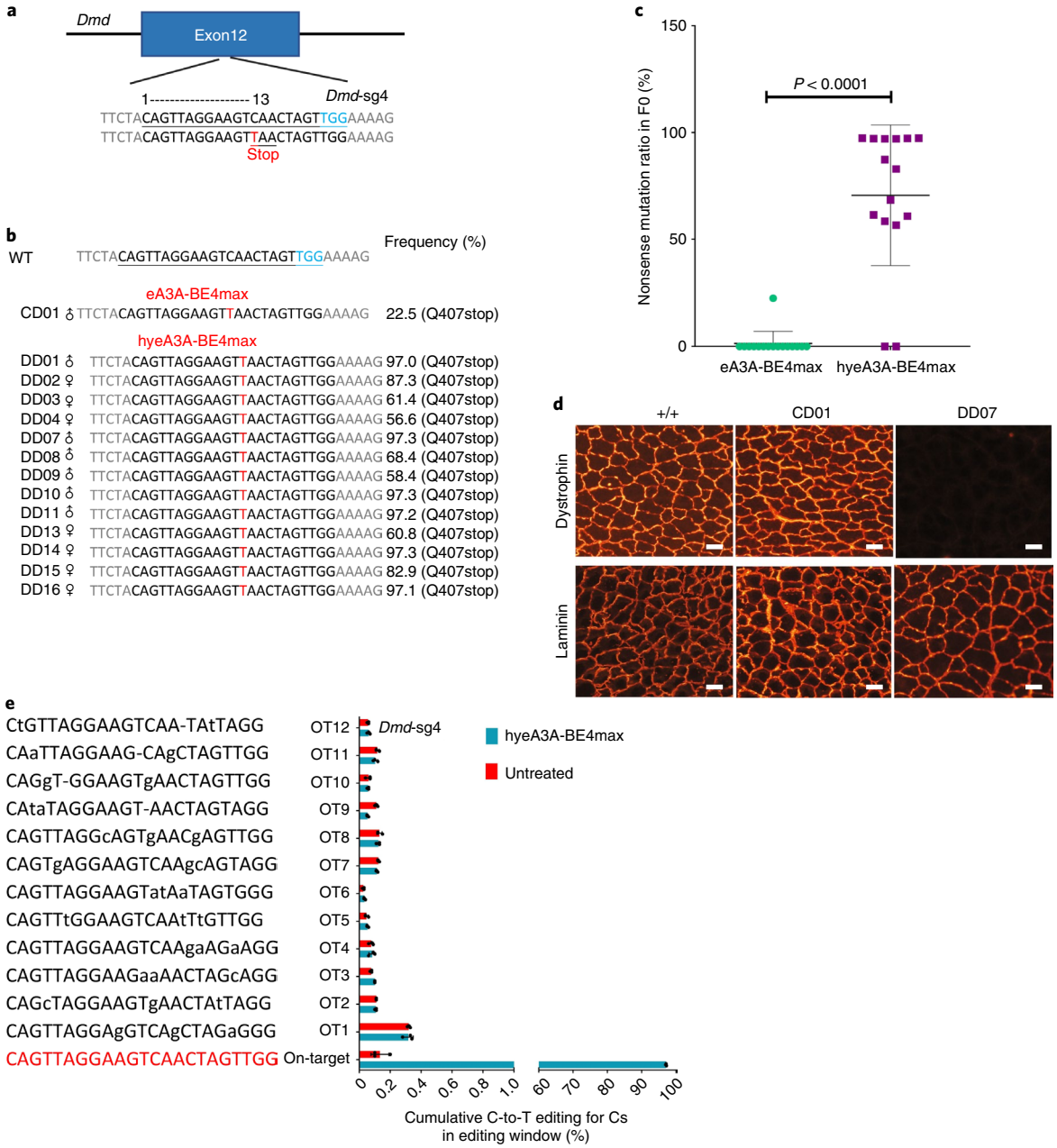


Fig. 6 | Highly efficient base editing by hyeA3A-BE4max in mouse embryos. a, Target sequence on the target of exon 12 in the dystrophin (*Dmd*) gene locus. The PAM sequence is shown in blue. The desired C-to-T mutation to create a premature stop codon is shown in red. **b**, Genotyping of all F0 generation pups by eA3A-BE4max and hyeA3A-BE4max. The allele frequency was determined by analysing HTS using BE-Analyzer. The percentage values on the right represent the frequencies of the indicated mutant alleles, with the corresponding mutation-induced amino acid conversions shown in parentheses. The frequency of the wild-type allele was omitted. **c**, Ratio of HTS reads containing the desired TAA stop codon of each founder that received eA3A-BE4max ($n=16$ mice) or hyeA3A-BE4max ($n=15$ mice) injection. Data are means ± s.d. The *P* value was determined by two-tailed Student's *t*-test. **d**, Immunofluorescence staining of tibialis anterior muscle tissue using antibodies against dystrophin and laminin from 6-week-old wild-type, CD01 and DD07 founders. The histological analysis images are representative of three independent experiments. Scale bars: 50 μm. **e**, Off-target analysis of three homozygous founders through HTS on 12 putative off-target sites predicted by Cas-OFFinder software. Data are means ± s.d. ($n=3$ mice). Statistical source data are provided with the paper.

by hyA3A-BE4max with much higher activity. The editing efficiency at C11 and C16 was increased 3.1- and 14-fold, respectively, by hyA3A-BE4max (Fig. 7b). To increase the specificity, we tested eA3A-BE4max and hyeA3A-BE4max, since C11 was in the cognate TCR motif. Although eA3A-BE4max precisely induced C-to-T conversion on C11 without bystander mutations, the efficiency was very low (Fig. 7b). However, hyeA3A-BE4max exhibited a threefold increase compared with eA3A-BE4max and did not generate detectable mutations at C3 and C16, suggesting its specificity (Fig. 7b). To further confirm that introduction of the −117G-to-A mutation could increase HBG expression, HUDEP-2($\Delta^G\gamma$) cells[37] that carry only one γ-globin gene were employed to facilitate the genotype analysis. HUDEP-2($\Delta^G\gamma$) cells were treated with lentiviral hyA3A-BE4max or hyeA3A-BE4max and then the stable

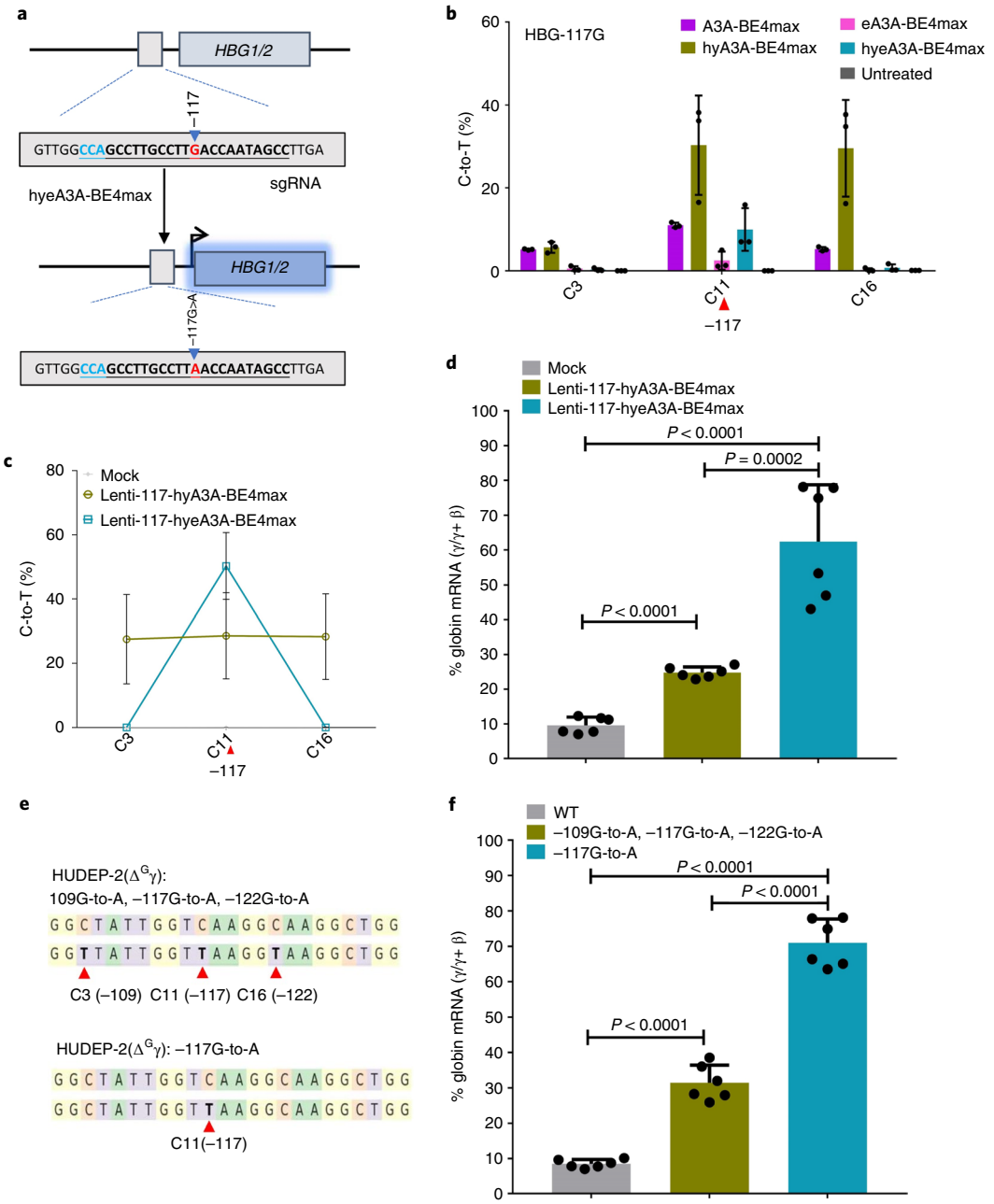


Fig. 7 | hyeA3A-BE4max generates accurate mutation in the HBG promoter in HUDEP-2 cells. a, Schematic of the critical sequence in the promoter to regulate HBG1/2 expression. The −117G-to-A mutation is shown in red. The core sequence of the transcriptional repressor BCL11A binding site is boxed. The PAM sequence is in blue. The G-to-A conversion abolishes BCL11a binding and activates HBG1/2 expression in adult erythrocytes. **b**, Base editing efficiency of A3A-BE4max, hyA3A-BE4max, eA3A-BE4max and hyeA3A-BE4max at the HBG1/2 promoter (−117G) site in HEK293T cells. Data are means ± s.d. ($n = 3$ independent experiments). **c**, Base editing efficiency of Lenti-117G-hyA3A-BE4max-P2A-GFP, Lenti-117G-hyeA3A-BE4max-P2A-GFP or Lenti-BE4max-P2A-GFP (mock; without sgRNA) at the HBG1/2 promoter (−117G) target in HUDEP-2($\Delta^G\gamma$) cells. The cells were infected with the indicated recombinant virus and the stably infected pooled cells were analysed. Data are means ± s.d. ($n = 3$ independent experiments). **d**, γ-globin mRNA expression relative to β-like (γ + β) globin mRNA in the HUDEP-2($\Delta^G\gamma$) cells infected with Lenti-117G-hyA3A-BE4max-P2A-GFP, Lenti-117G-hyeA3A-BE4max-P2A-GFP or mock infection after erythroid differentiation. Data are means ± s.d. ($n = 6$ independent experiments). *P* values were determined by two-tailed Student's *t*-test. **e**, Genotypes for each of the HUDEP-2($\Delta^G\gamma$) single-cell clones. Red arrowheads indicating substitutions comparing with wilde-type reference sequence (upper line). **f**, γ-globin mRNA expression relative to β-like (γ + β) globin mRNA in individual single clones of HUDEP-2($\Delta^G\gamma$) cells. Data are means ± s.d. ($n = 6$ independent experiments). *P* values were determined by two-tailed Student's *t*-test. Statistical source data are provided with the paper.

cells were pooled (Extended Data Fig. 6a). HTS data showed that hyeA3A-BE4max efficiently induced precise −117G-to-A mutation in HUDEP-2 ($\Delta^G\gamma$) cells and exhibited higher activity at this site compared with hyA3A-BE4max (Fig. 7c and Extended Data Fig. 6b). γ-globin mRNA levels were substantially elevated in hyA3A-BE4max- and hyeA3A-BE4max-treated cells compared with the parental HUDEP-2($\Delta^G\gamma$) cells, and importantly hyeA3A-BE4max treatment exhibited threefold higher HBG mRNA

induction compared with hyA3A-BE4max (Fig. 7d and Extended Data Fig. 6c), suggesting that bystander mutations at −122 or −109 sites could be detrimental to −117G-to-A conversion for reactivation of HBG expression. This was further confirmed through generation of single-cell clones containing single or triple G-to-A mutations and subsequent HBG mRNA evaluation (Fig. 7e,f). These data show that hyeA3A-BE4max could specifically generate the –117G-to-A mutation, indicating its advantages towards future clinical application.

Discussion

In this study, we generated hyCBEs (hyBE4max, hyA3A-BE4max and hyeA3A-BE4max) through fusion of a non-sequence-specific ssDBD from Rad51 protein. These advanced CBEs have higher activity and a broader editing window than previously reported tools. We also showed their ability to generate accurate mutations in animal models to mimic human disease and their potential applications for gene therapy. Our study increases the toolbox of base editing, and these featured tools will facilitate our basic research as well as multiple applications including gene therapy.

Since nucleotide deaminases usually catalyse ssDNA substrates, we hypothesized that increasing the affinities of CBEs with their substrates would improve the editing activity. Previous studies have shown that the N terminus of Cas9 is relatively static while the C terminus undergoes dramatic conformational changes during functioning. Based on this, fusion of deaminase or FokI to the N terminus of Cas9n is more efficient than to the C terminus[2,3,38,39]. Thus, we tested ten non-sequence-specific ssDBDs that were fused to the N terminus of BE4max. The RPA70-C domain from human RPA70 protein and Rad51DBD facilitated BE4max activity, but it is not clear why the other ssDBDs reduced its efficiency (Fig. 1c). It is possible that the binding activity of distinct ssDBDs varies. If the affinity is too strong, it will compete with the deaminase and inhibit its activity. If the affinity is too weak, it will have no effect. We also found that the base editor activity was further increased when the Rad51DBD was fused between rApobeC1 and Cas9n (Fig. 1c). It is likely that Rad51DBD functions as a long linker sequence in addition to ssDNA binding. Since it has been reported that a short 32-amino-acid linker between eA3A and nCas9 does not increase editing activity or alter editing window length[10], we believe that the much longer linker may set free the deaminase for better interaction with substrates, increasing its activity and targeting range. Based on the above speculation, more ssDBDs of various lengths can be tested to generate more active CBEs. Moreover, ssDBD fusion had no adverse effects on indel activity (Figs. 1f, 2c and 4c), protein expression (Extended Data Fig. 3b), product purity (Extended Data Fig. 4) or potential cell and DNA toxicity (Fig. 5e,f). Although RNA off-targeting editing in hyBE4max or hyA3A-BE4max still exists (Fig. 5d), this issue could be rapidly addressed with the development of new deaminase variants through the introduction of point mutations[26,27]. Moreover, no significant DNA off-target mutations were identified through both HTS at predicted off-target sites and WGS analysis, suggesting that hyCBEs are accurate tools. Since ssDBD increases the DNA affinity of hyCBEs, it is possible that potential sequence-independent off-target editing is also elevated. Taking into consideration that a typical sequencing depth (30–40×) cannot detect rare off-target mutations in cells or organisms[40], a more stringent technique, such as GOTI (genome-wide off-target analysis by two-cell embryo injection)[41], or an assay based on in-transition deamination within R-loops generated by an orthogonal Cas9 homologue[42], would be required to identify truly unpredictable DNA off-target mutations.

We also noticed that the effect of the ssDBD in hyBE4max and hyA3A-BE4max was slightly different. As shown in Fig. 1d,e, the activity of hyBE4max was much higher, but the editing window was comparable to BE4max. However, both the activity and the editing window were increased in hyA3A-BE4max compared with A3A-BE4max. A more dramatic activity increase was observed at positions C11–C16 which were usually untargetable for A3A-BE4max, suggesting a substantial increase of the targeting range (Fig. 2a,b). This suggests that insertion of ssDBDs into the base editors would be a general engineering strategy to increase some performance features of CBEs, but the effect may vary with different cytidine deaminases. Further studies to examine the fusion of ssDBD to other deaminases, such as hAID[21], PmCDA1 and their family members[14], will be interesting to explore more CBE variants with upgraded characteristics. Since the evolved CBEs[20] are efficient in all sequence contexts, fusion of ssDBD to evoCBEs (evoAPOBEC1-BE4max) would generate still more active CBEs. Additionally, the targeting scope of the CBEs will be further expanded through substitution of Cas9 variants with distinct PAM sequence requirements, such as xCas9 (ref. [43]), SpCas9-NG[44], SaCas9 and other variants[45,46].

In summary, this study demonstrates a strategy to engineer CBEs, which is compatible with other methods, such as codon optimization, linker sequence optimization and deaminase mutation/substitution. These CBEs significantly expand the targeting range and increase the activity both in vivo and in vitro, suggesting their broad use for research and therapeutic applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-020-0518-8.

Received: 16 October 2019; Accepted: 7 April 2020;
Published online: 11 May 2020

References

1. Collins, F. S., Brooks, L. D. & Chakravarti, A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
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. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
5. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
6. Li, X. et al. Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* **36**, 324–327 (2018).
7. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
8. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
9. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A–Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
10. Gehrke, J. M. et al. An APOBEC3A–Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
11. 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).
12. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
13. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv* **3**, eaao4774 (2017).
14. Cheng, T.-L. et al. Expanding C–T base editing toolkit with diversified cytidine deaminases. *Nat. Commun.* **10**, 3612 (2019).
15. Wang, L. et al. Enhanced base editing by co-expression of free uracil DNA glycosylase inhibitor. *Cell Res.* **27**, 1289–1292 (2017).
16. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).

17. 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).
18. Huang, T. P. et al. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* **37**, 626–631 (2019).
19. Zafra, M. P. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
20. Thuronyi, B. W. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
21. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Mehods* **13**, 1036–1042 (2016).
22. Jiang, W. et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res.* **28**, 855–861 (2018).
23. Dickey, T. H., Altschuler, S. E. & Wuttke, D. S. Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* **21**, 1074–1084 (2013).
24. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
25. Kim, D. et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
26. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
27. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
28. Grünewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).
29. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
30. Traxler, E. A. et al. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat. Med.* **22**, 987–990 (2016).
31. Wienert et al. Wake-up sleepy gene: reactivating fetal globin for β-hemoglobinopathies. *Trends Genet.* **34**, 927–940 (2018).
32. Ye, L. et al. Genome editing using CRISPR–Cas9 to create the HPFH genotype in HSPCs: an approach for treating sickle cell disease and β-thalassemia. *Blood* **113**, 10661–10665 (2016).
33. Antoniani, C. et al. Induction of fetal hemoglobin synthesis by CRISPR/Cas9-mediated editing of the human β-globin locus. *Proc. Natl Acad. Sci. USA* **131**, 1960–1973 (2018).
34. Wang, L. et al. Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. *Cell Res.* **30**, 276–278 (2020).
35. Dedoussis, G., Sinopoulou, K., Gyparaki, M. & Loutradis, A. Fetal hemoglobin expression in the compound heterozygous state for −117 (G→A) A_γ HPFH and IVS-1 nt 110 (G→A) β^+ thalassemia: a case study. *Eur. J. Haematol.* **65**, 93–96 (2000).
36. Martyn, G. E. et al. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat. Genet.* **50**, 498–503 (2018).
37. Wienert, B. et al. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood* **130**, 803–807 (2017).
38. Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582 (2014).
39. Tsai, S. Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576 (2014).
40. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR–Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
41. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
42. 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.* https://doi.org/10.1038/s41587-020-0414-6 (2020).
43. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
44. Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
45. Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
46. Kleinstiver, B. P. et al. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–484 (2015).

Methods

Plasmid construction. The primers and DNA sequences used in this study are listed in Supplementary Tables 1–5 and the Supplementary Information. PCR was performed using KOD -Plus- Neo DNA Polymerase (Toyobo; KOD-401). Human codon-optimized Apobec3A and Apobec3A (N57G) were synthesized by Genewiz. Human-derived RPA70-A, RPA70-B, RPA70-C, RPA32-D, BRCA2-OB2, BRCA2-OB3, Rad51DBD (amino acids 1–114), the HNRNPK KH domain and the PUF60 RNA recognition motif were amplified from human 293T complementary DNA (cDNA). BE4max (112096) and lentiCRISPR version 2 (52961) were purchased from Addgene. ssDBD-BE4max, hyBE4max, hy(e)A3A-BE4max and BE4max-C-Rad51 plasmids and Lenti-117-EFS-hy(e)A3A-BE4max-P2A-GFP were constructed using a ClonExpress MultiS One Step Cloning Kit (Vazyme)[47] (Supplementary Information). sgRNA expression plasmids were constructed as described in the step-by-step protocol of generation of site-specific point mutations by serial hyCBEs that can be found at the Nature Protocol Exchange[47]. Briefly, oligonucleotides listed in Supplementary Table 1 were denatured at 95 °C for 5 min, followed by slow cooling to room temperature. Annealed oligonucleotides were ligated into BbsI-linearized U6-sgRNA(sp)-EF1α-GFP for sgRNA expression (Thermo Fisher Scientific)[47,48].

Human cell culture. The HEK293T (ATCC; CRL-3216) cell line was cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% (*v*/*v*) foetal bovine serum (Gibco). HUDEP-2 cells were maintained and expanded in serum-free expansion medium (Stem Cell Technologies) supplemented with human stem cell factor (50 ng ml^{-1}; PeproTech), erythropoietin (3 IU ml^{-1}; PeproTech), dexamethasone (1 μM; Sigma–Aldrich), doxycycline (1 μg ml^{-1}; Takara Bio) and 2% Penicillin-Streptomycin (Gibco). All cell lines used were maintained at 37 °C and 5% CO_2 in an incubator[47].

Cell transfection and genomic DNA or total mRNA preparation. For DNA base editing, HEK293T cells were seeded into 24-well plates (Corning) and transfected at approximately 80% confluency. A total of 750 ng ssDBD-BE4max/BE4max, hy(e)A3A-BE4max/BE4max-C-Rad51 and 250 ng sgRNA expression plasmids were co-transfected using polyethyleneimine (PEI; Polysciences) following the manufacturer's recommended protocol. Three days later, transfected cells were digested with 0.25% trypsin (Gibco) for genomic DNA extraction. Cell lines or mouse tail tip genomic DNA were isolated using the TIANamp Genomic Kit (TIANGEN Biotech) according to the manufacturer's instructions[47]. For RNA off-target analysis, HEK293T cells were seeded into 10-cm dishes and transfected with 30 μg Cas9n-P2A-GFP, BE4max, hyBE4max, A3A-BE4max, hyA3A-BE4max, eA3A-BE4max and hyeA3A-BE4max using PEI at approximately 80% confluency. Three days later, transfected cells were digested with 0.25% trypsin (Gibco) for fluorescence-activated cell sorting (FACS). FACS was carried out on a FACSAria III (BD Biosciences) using FACSDiva version 8.0.2 (BD Biosciences). Cells were gated on their population via forward/sideward scatter after doublet exclusion (Supplementary Information). Around 500,000 cells (top 15% green fluorescent protein (GFP) signal) were collected, and RNA was extracted according to the standard protocol.

Western blotting. For western blots, HEK293T cells were lysed 72 h after transfection using RIPA buffer supplemented with proteinase and phosphatase inhibitors. Total protein was quantified using the BCA Protein Assay kit (Thermo Fisher Scientific). Total protein (10 μg per well) was loaded into a 15-well 8% Tris gel, separated by electrophoresis, and transferred to a nitrocellulose membrane before blocking with TBST containing 1% BSA. Nitrocellulose membranes were incubated with a 1:10,000 dilution of Anti-GAPDH (Abcam; ab9485) and a 1:5,000 dilution of Anti-CRISPR–Cas9 (Abcam; ab189380) overnight. Then, membranes were incubated with a 1:10,000 dilution of IRDye 800CW Goat anti-Rabbit IgG (H + L)(Abcam; ab216773) for 1 h and visualized using an Odyssey imager.

HUDEP-2 cell differentiation and quantitative PCR (qPCR). HUDEP-2 cells were differentiated in erythroid differentiation medium (IMDM; Corning) supplemented with 2% human blood type AB plasma (Sera Care), 1% L-glutamine, 2 IU ml^{-1} heparin, 10 μg ml^{-1} recombinant human insulin, 3 IU ml^{-1} erythropoietin, 330 μg ml^{-1} human holo-transferrin (Sigma–Aldrich), 100 ng ml^{-1} stem cell factor, 1 μg ml^{-1} doxycycline and 2% penicillin/streptomycin. Differentiated HUDEP-2 cells were surface stained with a 1:50 dilution of anti-CD235a-FITC (BioLegend; 349103) and a 1:50 dilution of anti-CD49d-APC (Miltenyi; 304307) to confirm the differentiation stage. A minimum of 10,000 cell events were collected for each sample on a FACSAria III (BD Biosciences) using a FACSDiva version 8.0.2 (BD Biosciences) and analysed with FlowJo version 10 software. On day 8 of differentiation, cells were harvested for total mRNA isolation. Isolated mRNA was reverse transcribed using HiScript II Q RT SuperMix (Vazyme). qPCR was performed on the QuantStudio 3 real-time PCR system (Applied Biosystems). HBG and HBB mRNA sequences were quantified by SYBR Green qPCR. qPCR primers are listed in Supplementary Table 4.

Lentivirus production and transduction of cell lines. Lentivirus production was performed as described in the protocol at the Nature Protocol Exchange[47]. Briefly, HEK293T cells were seeded into a 10-cm dish 1 d before transfection. At approximately 85% confluency, cells were co-transfected with 10 μg transfer plasmid (Lenti-117-EFS-hyA3A-BE4max-P2A-GFP or -117-EFS-hyeA3A-B E4max-P2A-GFP), 5 μg pMD2.G and 7.5 μg psPAX2 using PEI. Virus-containing supernatant was harvested at 48 and 72 h after transfection. Supernatant was centrifuged at 8,000 r.p.m. for 10 min at 4 °C to precipitate the cell debris, filtered by passing through a 0.45-μm low protein binding membrane (Millipore) and then centrifuged at 25,000 r.p.m. for 2.5 h at 4 °C to concentrate the lentivirus. The lentivirus titre was determined with FACS. Briefly, lentiviruses were titrated and produced in triplicates after transduction of HEK293T cells. For each viral construct, 1×10^4 HEK293T cells were transduced in suspension with 0.0001, 0.001, 0.01, 0.1, 1 or 10 μl viral supernatant in wells of a 96-well plate. In each well, culture medium was added to make the final volume 300 μl. Cells without any virus added were also plated in a 96-well plate (three wells as a control). Three days post-transduction, cells transduced with lentivirus were analysed to quantify the percentage of enhanced GFP using a Fortessa Flow Cell Analyzer (BD Biosciences). For each virus construct, the following score was calculated: titre (TU ml^{-1}) = cell number × % EGFP × 10^3/virus stock volume (μl). Equal titres (multiplicity of infection: 40) of Lenti-117-EFS-hA3A-BE4max-P2A-GFP, Lenti-117-EFS-heA3A-BE4max-P2A-GFP and lenti EFS-hA3A-BE4max-P2A-GFP in the presence of polybrene (8 μg μl^{-1}) were used to transduce HUDEP-2($\Delta^G\gamma$) cells. Then, 12 h after transduction, the medium was replaced with fresh culture medium. Three days after transduction, FACS was performed to sort the GFP$^+$ HUDEP-2($\Delta^G\gamma$) cells for the next culture. Cells were gated on their population via forward/sideward scatter after doublet exclusion[46] (Supplementary Information).

Preparation of sgRNA and mRNA. sgRNA and mRNA preparation was performed as previously described[48]. The annealed sgRNA sequences were cloned into the T7-sgRNA sp-scaffold. The T7 promoter and different sgRNA templates were amplified using the primers IVT-PCF-F/R (sp) (Supplementary Table 4). The sgRNA sequences were then transcribed using the in vitro T7 Transcription Kit (MEGAshortscript Kit; Ambion). The T7 promoter was introduced into the (e)A3A-BE4max and hy(e)A3A-BE4max templates by PCR using the primers T7-mRNA(hCBE)-F/R (Supplementary Table 4). (e)A3A-BE4max and hy(e) A3A-BE4max mRNA sequences were transcribed using the in vitro RNA transcription kit (mMESSAGE mMACHINE T7 ULTRA Kit; Ambion). Both sgRNA and mRNA were purified using a MEGAclear Kit (Ambion).

Animals and microinjection of zygotes. Animal manipulation was performed as previously described[49]. In brief, C57BL/6J and ICR mouse strains purchased from the Shanghai Laboratory Animal Center were housed in standard cages in a specific pathogen-free facility on a 12 h light/ 12 h dark cycle with ad libitum access to food and water. C57BL/6J and ICR mouse strains were used as embryo donors and foster mothers, respectively. 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. For microinjection, solutions containing complexes of CBE or hyCBE mRNA (100 ng μl^{-1}) and sgRNA (200 ng μl^{-1}) were diluted in nuclease-free water and injected into cytoplasm using an Eppendorf TransferMan NK2 micromanipulator. Injected zygotes were transferred into pseudopregnant female mice immediately after injection.

Immunofluorescent staining. Tibialis anterior muscle from 5-week-old wild-type or *Dmd* mutant mice (CD07; DD01) was frozen in liquid nitrogen-cooled isopentane and immunostained with laminin or dystrophin antibodies. Dystrophin and laminin were detected with a 1:500 dilution of anti-dystrophin (Abcam; ab15277) or a 1:500 dilution of anti-laminin rabbit polyclonal antibody (Abcam; ab11575), respectively, followed by a 1:1,000 dilution of Alexa Fluor 594 Donkey anti-rabbit secondary antibody (Thermo Fisher Scientific). Fluorescence of sections was observed with a Leica DMI4000 B fluorescence microscope.

MTS assay and comet assay. HEK293T cells were seeded into 96-well plates (Corning) and transfected with 100 ng base editors at approximately 80% confluency. Then, 48 and 72 h after transfection, cell viability was measured by Promega CellTiter 96 AQueous One Solution Cell Proliferation Assay (MTS) according to the manufacturer's instructions (Promega). A comet assay was performed as previously described[50]. In brief, HEK293T cells were seeded into 24-well plates (Corning) and transfected with 750 ng base editors at approximately 80% confluency. Then, 72 h after transfection, the cells were collected and resuspended in phosphate-buffered saline at a concentration of 3×10^5 cells per ml. The comet assay then was performed according to the manufacturer's instructions (4250–050-K; Trevigen). DNA damage was measured in terms of tail moments using comet score software (casplab_1.2.3b2).

HTS and data analysis. On- and off-target genomic regions of interest were amplified from ~50–100 ng genomic DNA by PCR with the primers listed in Supplementary Tables 2 and 3. HTS amplification libraries were prepared by PCR using KOD -Plus- Neo DNA polymerase and site-specific primers containing an adaptor sequence (forward: 5′-GGAGTGAGTACGGTGTGC-3′; reverse:

5′-GAGTTGGATGCTGGATGG-3′) at the 5′ end[46] (Supplementary Table 4 and 5). The above products were then subjected to a second round of PCR using primers containing different barcode sequences. The resulting libraries were mixed and sequenced on an Illumina HiSeq platform. The C-to-T, A and G conversions and indels in the HTS data were analysed using BE-Analyzer[51] or CRISPResso2 (ref. [52]).

RNA sequencing (RNA-Seq) experiments. A total of 3 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using a NEBNext Ultra RNA Library Prep Kit for Illumina (NEB) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5×). First-strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H–). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, a NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridization. To select cDNA fragments of preferentially ~250–300 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter). Then, 3 μl USER Enzyme (NEB) was incubated with size-selected, adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. Then, PCR was performed with Phusion High-Fidelity DNA polymerase, universal PCR primers and Index (X) Primer. Finally, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent 2100 Bioanalyzer system. The clustering of index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina), according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform, and 125 bp/150 bp paired-end reads were generated.

RNA sequence variant calling and quality control. The analysis of RNA-Seq sequencing data was performed as follows. Raw data (raw reads) of FASTQ format were first processed through in-house Perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter and trimming low-quality bases with Trimmomatic. At the same time, the Q20, Q30 and GC contents of the clean data were calculated. All the downstream analyses were based on the clean data with high quality. An index of the reference genome was built using HISAT2 version 2.0.5 and paired-end clean reads were aligned to the reference genome (Ensemble GRCh38) using HISAT2 version 2.0.5. We selected HISAT2 as the mapping tool because it can generate a database of splice junctions based on the gene model annotation file and thus a better mapping result than other non-splice mapping tools. GATK (version 4.0) software was used to perform single-nucleotide polymorphism calling. Variant loci in base editor overexpression experiments were filtered to exclude sites without high-confidence reference genotype calls in the control experiment. The read coverage for a given single-nucleotide variant in a control experiment should be greater than the 90th percentile of the read coverage across all single-nucleotide variants in the corresponding overexpression experiment. Additionally, these loci were required to have a consensus of at least 99% of reads containing the reference allele in the control experiment. RNA edits in Cas9n-P2A-GFP controls were filtered to include only loci with ten or more reads and with greater than 0% of reads containing alternate alleles. Base edits labelled as C-to-U comprise C-to-U edits called on the positive strand as well as G-to-A edits sourced from the negative strand.

WGS. WGS used mouse genomic DNA extracted from mouse tails, and was performed at a sequencing depth of 30× to 40× using an MGI2000 sequencer. We mapped the sequencing data using the BWA tool (version 0.7.17-r1188) of Sentieon (version 2019.11) software with a mouse reference genome (GRCm38/mm10). Variants were identified using the Haplotyper tool of Sentieon software and the NCBI mouse Single Nucleotide Polymorphism Database was sorted out from the identified variants using Annovar (version 2018-04-16). To figure out the potential off-target sites, we picked out that C and G converted to the other bases among the remaining variants. Then, we excluded the common variants between the wild-type and the hyeA3A-BE4max-treated sample. The putative off-target sites were compared with the candidates from Cas-OFFinder considering mismatches up to 7 bp in length.

Statistics and reproducibility. Data are presented as means ± s.d. from independent experiments. All statistical analyses were performed on at least $n = 3$ biologically independent experiments or three biologically independent samples unless otherwise noted in the figure captions. The experiments involving HUDEP-2 differentiation and globin mRNA detection were performed six times with similar results. The experiments involving western blotting, immunofluorescence staining, genotyping of F0 or F1 mice and differentiation-stage evaluation of HUDEP-2 cells were performed three times with similar results and the representative data are shown. An unpaired two-tailed Student's *t*-test was used to determine the significance of the differences between two groups, using GraphPad Prism 6 (GraphPad Software). Specific *P* values are indicated in the figure captions. $P < 0.05$ was considered significant.

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

Data availability

HTS data have been deposited in the NCBI Sequence Read Archive database under accession codes PRJNA566262, PRJNA566253 and PRJNA602779. RNA-Seq data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA599328. WGS data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA610447. Source data for Figs. 1–7 and Extended Data Figs. 2–6 are presented with the paper. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

47. Zhu, B., Chen, L., Zhang, X. & Li, D. Efficient generation of site-specific point mutations in cell lines by hyper active CBEs (hyCBEs). *Protoc. Exch.* https://doi.org/10.21203/rs.3.pex-841/v1 (2020).
48. 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).
49. Li, D. et al. Heritable gene targeting in the mouse and rat using a CRISPR–Cas system. *Nat. Biotechnol.* **31**, 681–683 (2013).
50. Liu, H. et al. Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* **563**, 131–136 (2018).
51. Hwang, G.-H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* **19**, 542 (2018).
52. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

Acknowledgements

We thank M. Crossley from the School of Biotechnology and Biomolecular Sciences, University of New South Wales, Australia, who kindly provided the HUDEP-2($\Delta^{G}\gamma$) cells and edited the manuscript, and X. Huang from ShanghaiTech University for discussions and suggestions in relation to this study. We thank H. Jiang from the Core Facility and Technical Service Center for SLSB, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University for technical support with FACS analysis. This work was partially supported by grants from the National Key Research and Development Program of China (2019YFA0110802 and 2019YFA0802800), National Natural Science Foundation of China (81670470 and 81873685) and Shanghai Municipal Science and Technology Commission (18411953500), and from the Innovation program of the Shanghai Municipal Education Commission (2019-01-07-00-05-E00054). This work was also supported by Fundamental Research Funds for the Central Universities, and the ECNU Public Platform for innovation (011).

Author contributions

X.Z. and D.L. designed the experiments and analysed the data. X.Z., L.C., B.Z., L.W., C,C., M.H., Y.H., H.H., B.C., W.Y., S.Y., L.Y., Z.Y., Meizhen Liu, Y.Z., H.L. Z.M. and Y.W. performed the experiments. D.L., X.Z., L.C. and B.Z. analysed the data and wrote the manuscript with the input from all of the authors, Mingyao Liu and D.L. supervised the research.

Competing interests

The authors have submitted a patent application (application numbers 2019113109698, 2019113125440 and 2019113125370) based on the results reported in this study.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-020-0518-8.

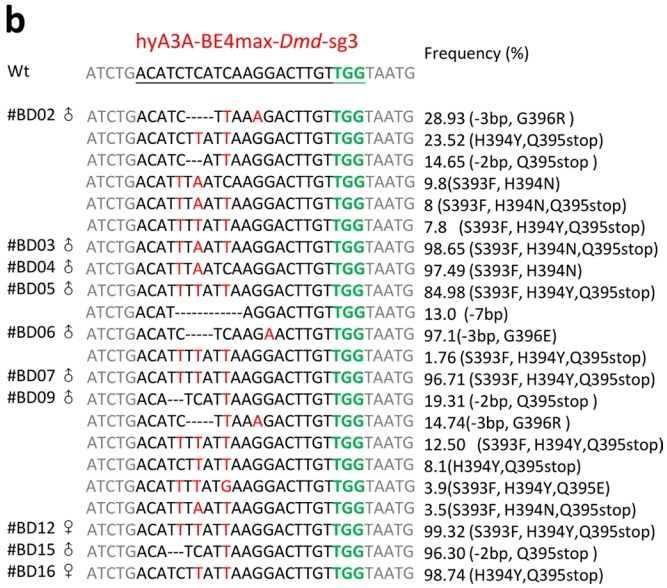
Supplementary information is available for this paper at https://doi.org/10.1038/s41556-020-0518-8.

Correspondence and requests for materials should be addressed to D.L.

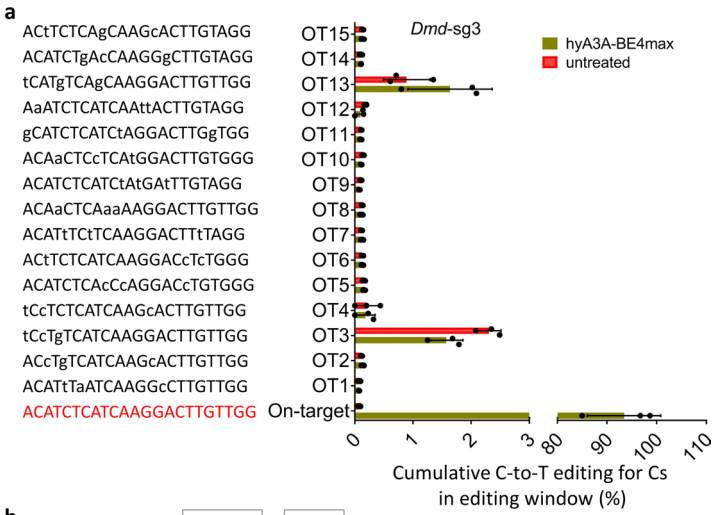
Reprints and permissions information is available at www.nature.com/reprints.

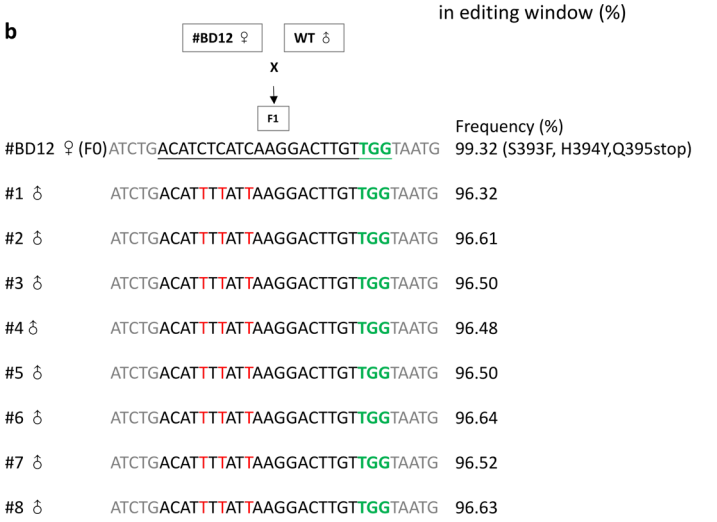
a

b

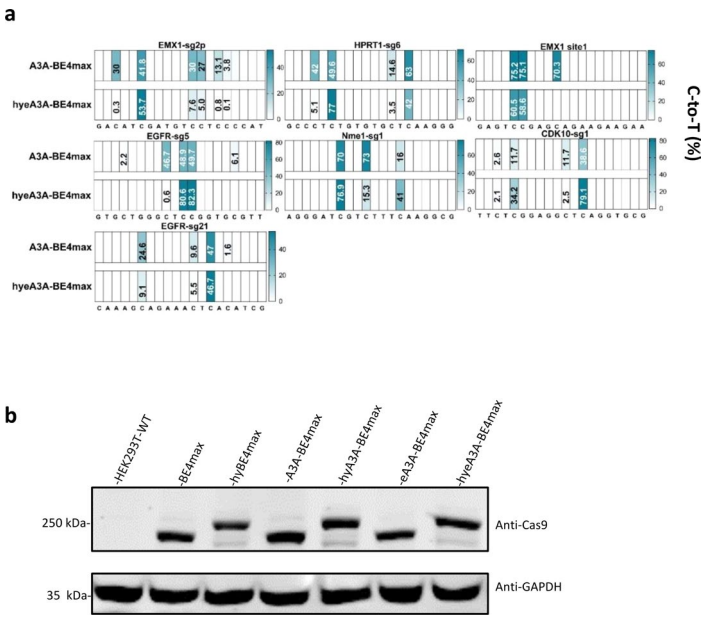


Extended Data Fig. 1 | Highly efficient base editing by A3A-BE4max or hyA3A-BE4max in mouse embryos. (**a, b**) Genotyping of F0 generation pups by A3A-BE4max and hyA3A-BE4max. The frequencies of WT and mutant alleles were determined by analyzing HTS using BE-analyzer. The percentage on the right represents the frequency of the indicated mutant allele with the corresponding mutation-induced amino acid conversion shown in parentheses. The frequency of the wild-type allele was omitted. Wt, wild-type.

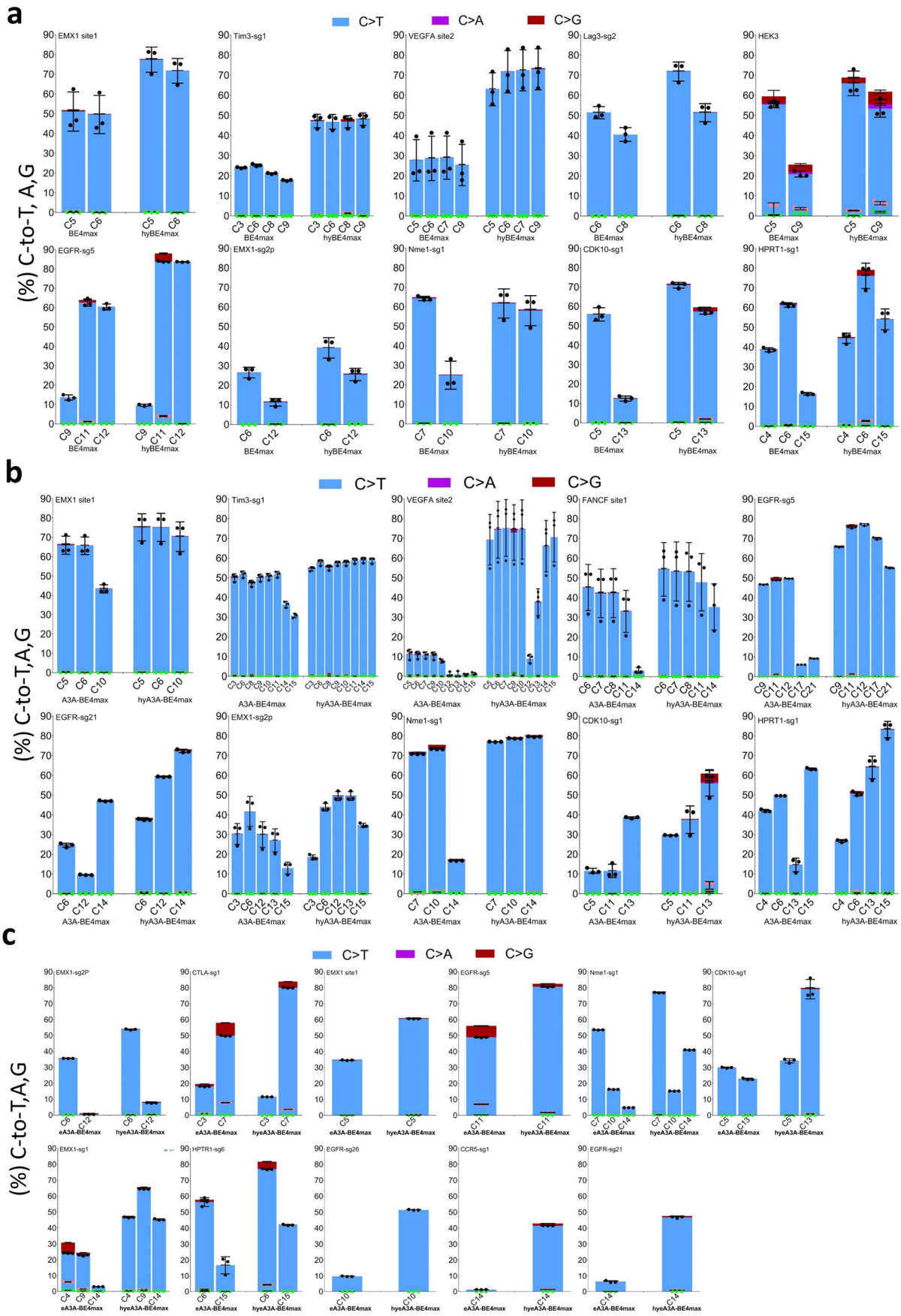




Extended Data Fig. 2 | Off-target analysis and germline transmission of the founders derived from hyA3A-BE4max injection. (**a**) HTS was performed with mouse tails to determine editing efficiencies at 15 potential off-target sites in three Dmd mutant F0 mice (#BD03, #BD04 and #BD07). Mismatched nucleotide letters are indicated in lowercase. Data are means ± SD (n = 3 mice).(**b**) HTS alignments of mutant sequences from F1 generated by mating founder #BD12(♀) with Wt (♂). The column on the right indicates frequencies of mutant alleles. Wt, wild-type.Statistical source data are provided in Source Data Extended Data Fig. 2.



Extended Data Fig. 3 | Comparison of base editing efficiency and protein levels by CBEs and hyCBEs in HEK293T cells. (**a**)Comparison of base editing efficiency induced by A3A-BE4max or hyeA3A-BE4max in HEK293T cells. The average mutation percentage derived from three independent experiments of A3A-BE4max and hyeA3A-BE4max at the same site is listed. Some of the data (hyeA3A-BE4max) are the same as presented in Fig. 4a. Statistical source data are provided in Source Extended Data Fig. 3. (**b**) The protein levels of BE4max, hyBE4max, A3A-BE4max, hyA3A-BE4max, eA3A-BE4max and hyeA3A-BE4max were determined by Western blotting in HEK293T cells 3 days after transfection of similar amounts of plasmid DNA. Specific antibodies against Cas9 (top) or GAPDH (bottom) were used. Western blotting images are representative of three independent experiments. Unprocessed blots are shown in Source Data Extended Data Fig. 3.



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

Extended Data Fig. 4 | Comparison of base editing product purity induced by variant base editors in HEK293T cells. (**a**) Comparison of base editing products induced by BE4max vs hyBE4max. HTS data were analyzed and the ratio of each type of nucleotides was listed on each target position. Data are means ± SD (n = 3 independent experiments). (**b**) Comparison of base editing products induced by A3A-BE4max vs hyA3A-BE4max. HTS data were analyzed and the ratio of each type of nucleotides was listed on each target position. Data are means ± SD (n = 3 independent experiments) (**c**) Comparison of base editing product induced by eA3A-BE4max vs hyeA3A-BE4max. HTS data were analyzed and the ratio of each type of nucleotides was listed on each target position. The individual data points are shown as black (C > T), light green (C > A) and light red (C > G) dots. Data are means ± SD (n = 3 independent experiments). Statistical source data are provided in Source Data Extended Data Fig. 4.

a

	Control	Treated	Uniquely assigned to #
	WT	hyeA3A-BE4max #DD11	hyeA3A-BE4max #DD11
SNPs	62,359	82,573	24,755
SNPs after excluding dbSNP	20,361	37,523	20,387
C>T SNPs	3009	5305	2749
C>G SNPs	846	1491	749
C>A SNPs	1280	1950	853
G>A SNPs	2734	4894	2587
G>C SNPs	1019	1737	870
G>T SNPs	1373	2095	932

b

Mismatch	No. of on-/off-target sites
0	1
3	20
4	374
5	2,869
6	23,225
7	148,569

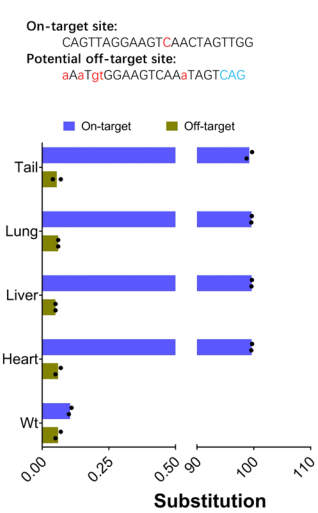
c

Sample	Total Bases	Cover Bases	Cover Ratio(%)	Depth
WT	2730871774	2651281235	97.09	36.94 x
#DD11	2730871774	2651948450	97.11	37.7 x

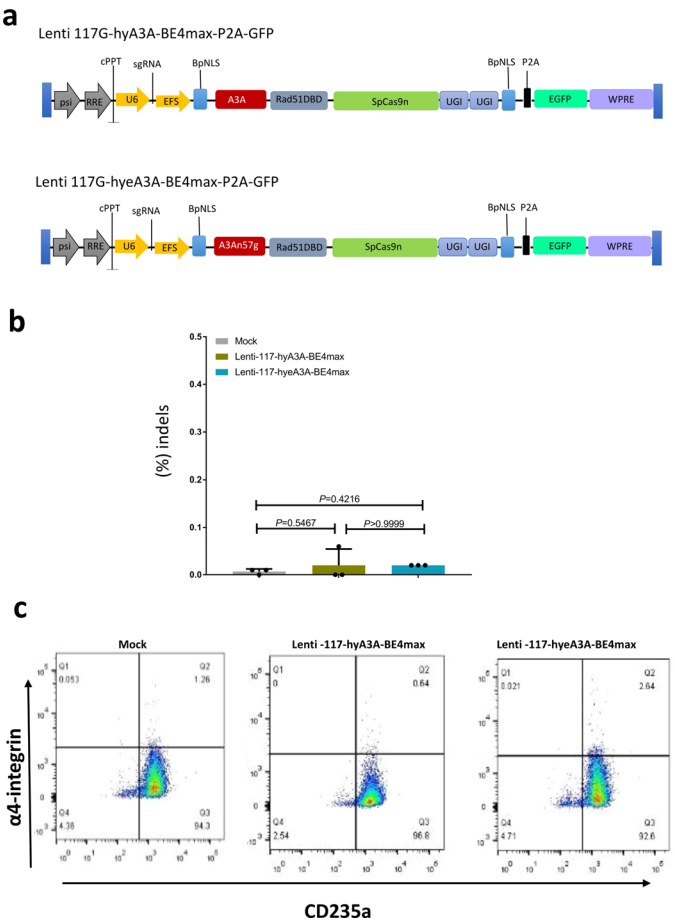
d

Sample	On-target	off-target
wt	0	0
#DD11	1	1

e



Extended Data Fig. 5 | Whole genome sequencing of *Dmd* F0 (#DD11) and wild-type (Wt) mice. (**a**) Summary of genome sequencing analysis. WGS for a Dmd mutant mouse (#DD11) and a wild type mouse (Wt) were performed. A total of 82,573 and 62,359 SNPs were identified for #DD11 and Wt, respectively. After filtering out dbSNP (naturally occurring variants in the SNP database), 20,387 SNPs were obtained in the #DD11 genome. Then the sequences at the remaining SNP sites were compared with all on-/off-target sequences (20 bp). (**b**) Summary of on-/off-target site information. A total of 175,058 sites, including 1 on-target site and 20; 374; 2,869; 22,335; and 148,569 off-target sites with 3, 4, 5, 6, or 7 mismatch/es, respectively, were analyzed. (**c**) Summary of the whole-genome sequencing. (**d**) Summary of off-target analysis. After comparing the sequences at the remaining SNP sites with the 175,058 on-/off-target sequences (20 bp), the C-to-T substitution was only detected within the on-target sequencing in #DD11. (**e**) Validation the off-target candidate site determined in (d) using targeted deep sequencing of genomic DNA isolated from various #DD11 organs (heart, liver, lung and tail). Mismatched nucleotides and PAM sequences are shown in red and in blue, respectively. Data represent mean from two independent experiments. Statistical source data are provided in Source Data Extended Data Fig. 5.



Extended Data Fig. 6 | Indels and differentiation stage evaluation of HUDEP-2 ($\Delta^{G}\gamma$) cells after viral infection. (**a**) Schematic representation of lentivirus constructs for HUDEP-2 infection.Psi+, Psi packaging signal; RRE, Rev response element; cPPT, central polypurine tract; EFS, elongation factor 1a short promoter; Bp-NLS, bipartite nuclear localization signals; A3A, derived from human Apobec3A; Rad51DBD, derived human rad51 single strand DNA binding protein domain; spCas9n, Cas9 D10A; P2A, 2 A self-cleaving peptide; WPRE, post-transcriptional regulatory element; UGI, Uracil DNA glycosylase inhibitor; EGFP, a maker for FACS. (**b**) Comparison of indels generated by lenti-hyA3A-BE4max or lenti-hyeA3A-BE4max treated HUDEP-2(ΔGγ) cells. Data are means ± SD (n = 3 independent experiments). P value was determined by two-tailed Student's t test. (**c**) Erythroid differentiation validation of HUDEP-2 (ΔGγ) cells evaluated by anti α4-integrin(APC) and anti CD235a(FITC) surface markers, 8 days after differentiation. FACS data analysis are representative of three independent experiments. Statistical source data are provided in Source Data Extended Data Fig. 6.

nature research

Corresponding author(s): Dali Li

Last updated by author(s): Mar 21, 2020

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Targeted amplicon sequencing data data was collected and demultiplexed by an Illumina HiSeq X Ten instrument. RNA-seq data was collected and demultiplexed by an Illumina NovaSeq 6000 instrument. Whole genome sequencing(WGS) was performed at a sequencing depth of 30 × to 40 × using an MGI2000 sequencer (Wuhan, China) FACS gating data was collected on a FACSAria III (BD Biosciences) using FACSDiva version 8.0.2 (BD Biosciences)
Data analysis	High-throughput sequencing data was analyzed by BE-Analyzer (http://www.rgenome.net/be-analyzer/#!)(Hwang G-H et al, BMC Bioinformatics, 2018) or CRISPResso2 (http://crispresso.pinellolab.partners.org/)(Clement, K. et al.Nat Biotechnol. 2019)for base editing(C>T ,C>G and C>A) and indels efficiencies. Potential DNA off-targets site for hyCBEs were predicated using cas-OFFinder web (http://www.rgenome.net/cas-offinder/). RNA-seq data were analyzed using in-house Perl scripts, Hisat2 v2.0.5 and GATK (v4.0) software. WGS data were analyzed using BWA (version 0.7.17-r1188) aligner of Sentieon (version 2019.11) software, Haplotyper of Sentieon software and Annovar (version 2018-04-16). DNA damage was measured in terms of tail moments using comet score software (casplab_1.2.3b2) . FACS data was analyzed using FlowJo v.10. Prism 7 was also used to analyze 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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

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

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

High–throughput sequencing reads have been deposited in the NCBI Sequence Read Archive under(PRJNA566253 ,PRJNA566262 and PRJNA602779) . RNA-seq data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA599328. Whole-genome sequencing data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA610447.All Plasmids sequences or target site sequences mentioned in the manuscript are provided in supplementary information file. There are no restrictions on data availability.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample size. Experiments were performed in biological triplicate n=3 unless otherwise noted.
Data exclusions	No data were excluded from the analyses.
Replication	Three independent biological replicates were performed on different days. All replications were successful.
Randomization	Samples were randomly distributed into groups.
Blinding	One person collected samples. Another person analyzed without any description.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	a 1:500 dilution of anti-dystrophin(Abcam, ab15277) a 1:500 dilution of anti- laminin (Abcam, ab11575), a 1:10000 dilution of anti-GAPDH (Abcam, ab9485) , a 1:5000 dilution of anti-CRISPR-Cas9 (Abcam, ab189380). a 1:50 dilution of anti-CD235a-FITC (BioLgend,349103). a 1:50 dilution of anti-CD49d-APC(Miltenyi,304307).
Validation	The dystrophin antibody has been validated by mouse triceps muscle(https://www.abcam.com/dystrophin-antibody-ab15277.html). The laminin antibody has been validated by mouse anterior tibialis(https://www.abcam.com/laminin-antibody-ab11575.html). The GAPDH antibody has been validated by western blot in HEK293T cells lysate(https://www.abcam.com/gapdh-antibody-

loading-control-ab9485.html).
The Cas9 (ab204448, Abcam) antibody has been validated by western blot in HEK293T cells transfected with CRISPR-Cas9 (Q99ZW2, Streptococcus pyogenes serotype M1) with GFP-Myc tag (https://www.abcam.com/crispr-cas9-antibody-epr18991-ab189380.html).
The CD235a-FITC antibody has been validated by immunofluorescent staining with flow cytometric analysis (https://www.biolegend.com/en-us/products/fitc-anti-human-cd235a-glycophorin-a-antibody-6701).
The CD49d-APC antibody has been validated immunofluorescent staining with flow cytometric analysis(https://www.biolegend.com/en-us/products/apc-anti-human-cd49d-antibody-582).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK 293T cells (purchased from ATCC CRL-3216) HUDEP-2 (ΔGγ)(from Merlin Crossley`s lab
Authentication	HEK293T cells were directly purchased from ATCC with certification; HUDEP-2(ΔGγ) cells were confirmed by PCR and differentiation studies followed by FACS 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 organisms

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

Laboratory animals	C57BL/6J and ICR mouse strain purchased from Shanghai Laboratory Animal Center were housed in standard cages in a specific pathogen-free facility on a 12 h light/dark cycle with ad libitum access to food and water. 4-6 week-old female wild-type C57BL/6J mouse and 6-12 week-old female wild-type ICR mouse strain were used as embryo donors and foster mothers, respectively.
Wild animals	No studies with wild animals were performed.
Field-collected samples	No studies with field-collected samples were performed.
Ethics oversight	All animal experiments conformed to the regulations drafted by the Association for Assessment and Accreditation of Laboratory Animal Care in Shanghai and were approved by the East China Normal University Center for Animal Research.

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

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	Cell culture and transfection procedures are described in the online methods. Cells were washed and filtered through a 45µm cell strainer cap before sorting (72h after transfection).
Instrument	FACSAria III (BD Biosciences) or Fortessa Flow Cell Analyzer (BD Biosciences)
Software	BD FACSDiva Software Diva8.0.2
Cell population abundance	HEK293T Cell population abundances after gating for target populations were similar in different biology replicates. HEK293T cells infected with base editors described in the supplement usually were ~ 60-75% GFP+ (of gated population = % parent in BD FACSDiva). HUDEP-2 (ΔGγ) Cell population abundances after gating for target populations were similar in different biology replicates. HUDEP-2 (ΔGγ) cells infected with lentivirus described in the supplement usually were ~ 20-30% GFP+ (of gated population = % parent in BD FACSDiva).

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. cells with top 15% of GFP signal were sorted, after gating for the cell population (~15% of parent). Exemplifying the gating strategy is provided in the Supplementary Figure 2.
For HUDEP-2 (ΔGγ) cells ,gates were established using uninfected control cells. Gates were drawn to collect subsets of GFP-expressing cells. cells with top 10% of GFP signal were sorted, after gating for the cell population (~10% of parent). Exemplifying the gating strategy is provided in the Supplementary Figure 3.

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