nature biotechnology

BRIEF COMMUNICATION

https://doi.org/10.1038/s41587-020-0527-y

Dual base editor catalyzes both cytosine and adenine base conversions in human cells

Xiaohui Zhang[1,9], Biyun Zhu[1,9], Liang Chen[1,9], Ling Xie[1], Weishi Yu[1,2], Ying Wang[3], Linxi Li[1], Shuming Yin[1], Lei Yang[1], Handan Hu[1], Honghui Han[4], Yongmei Li[1], Liren Wang[1], Geng Chen[1], Xueyun Ma[1], Hongquan Geng[5], Wanfeng Huang[1], Xiufeng Pang[1], Zuozhen Yang[2], Yuxuan Wu[1], Stefan Siwko[6], Ryo Kurita[7], Yukio Nakamura[8], Li Yang[3], Mingyao Liu[1] and Dali Li[1]✉

Although base editors are useful tools for precise genome editing, current base editors can only convert either adenines or cytosines. We developed a dual adenine and cytosine base editor (A&C-BEmax) by fusing both deaminases with a Cas9 nickase to achieve C-to-T and A-to-G conversions at the same target site. Compared to single base editors, A&C-BEmax's activity on adenines is slightly reduced, whereas activity on cytosines is higher and RNA off-target activity is substantially decreased.

The commonly used base editors BE3 (cytosine base editor (CBE)) and ABE7.10 (adenine base editor (ABE)) typically generate C·G-to-T·A or A·T-to-G·C conversions, respectively, within a ~5-nucleotide editing window on the target DNA[1,2]. Using highly active deaminases and nuclease-impaired Cas9 or Cpf1, base editors catalyze nucleotide conversions very potently and rarely generate double-stranded DNA breaks[3]. However, CBEs or ABEs only catalyze conversion of a single type of nucleotide, either C·G-to-T·A or A·T-to-G·C[1–3], which limits product diversity. We assumed that a tool that converts two types of nucleotides on the same allele would drastically modify the target sequence, broadening its base editing capability.

We hypothesized that fusing cytidine and adenosine deaminases to Cas9n (SpCas9 D10A mutant) could achieve simultaneous C·G and A·T base editing on the same allele, creating a new base editor, A&C-BE (Fig. 1a). To test this, five constructs were generated with combinations of the two deaminases and Cas9n (Supplementary Fig. 1 and Supplementary Sequence 1) and tested with a reporter system that is activated after simultaneous C/A editing to restore luciferase expression (Supplementary Fig. 2a and Supplementary Sequence 2). A&C-BEs with cytidine deaminase fused to the N terminus, but not the other constructs, significantly stimulated reporter activity (Supplementary Fig. 2b). Further analyzing the editing efficiency at the endogenous site (*FANCF-sg3*), we found that TadA functioned only when it was fused to the N terminus adjacent to Cas9n, and cytidine deaminases fused to the N terminus of ABE7.10 induced significant levels of simultaneous A/C mutation in the same allele (Fig. 1b,c and Supplementary Fig. 3). Moreover, cytidine deaminase activity was strongly increased when it was fused to the N terminus of ABE7.10, and hAID exhibited higher activity compared to rAPOBEC1 (Fig. 1b,c and Supplementary Fig. 3).

On the basis of these results, we further optimized the ABE7.10-N-AID construct to improve its efficiency. Consistent with previous reports[4], we optimized codons and added a bipartite nuclear localization signal (NLS) to generate ABE7.10-N-AIDmax, leading to significantly increased CBE activity and slightly elevated ABE activity, as well as simultaneous A/C conversions (Fig. 1b,c and Supplementary Figs. 3 and 4). Through screening five linker sequences, a rigid 15-residue (EAAAKEAAAKEAAAK) linker exhibited better performance (Supplementary Fig. 5). The final version of A&C-BE, A&C-BEmax, was generated by adding two copies of uracil DNA glycosylase inhibitor (UGI)[5] (Supplementary Fig. 1). Through the serial optimization steps above, the base conversion efficiency, product purity and A/C simultaneous conversion activity of A&C-BEmax were significantly increased compared to the original construct (Supplementary Fig. 6). Compared to A&C-BEmax, co-transfection of CBE and ABE7.10 with six single-guide RNAs (sgRNAs) yielded a very low simultaneous A/C conversion ratio for most targets (Fig. 1b,c and Supplementary Fig. 7). Most likely, single base editors competed with each other due to occupation of the target site. Once the target has been edited by one editor, the other base editor will not efficiently function because the sgRNA no longer perfectly matches the target DNA. Our initial study showed that, through fusion of cytidine and adenosine deaminase to Cas9n, A&C-BEmax was able to catalyze A/C base conversion on the same allele with increased cytidine deaminase activity compared to CBEs.

To unbiasedly characterize the performance of A&C-BEmax compared to ABEmax and AID-BE4max (with optimized codon, NLS and UGI copies; see Supplementary Sequence 1), 28 endogenous targets (including four sites containing only Cs or As, respectively) were investigated in HEK293T cells (Fig. 1d). The A-to-G editing window of A&C-BEmax was consistent, and the A-to-G editing efficiency was similar or slightly decreased compared to ABEmax at most of the targets (Fig. 1d,e and Supplementary Fig. 8a). Notably, the C-to-T editing window for A&C-BEmax was expanded to 16 nucleotides (positions 2–17) compared to positions 3–13 for AID-BE4max (Fig. 1d,f and Supplementary Fig. 8b).

[1]Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China. [2]Cipher Gene LLC, Beijing, China. [3]CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [4]Bioray Laboratories Inc., Shanghai, China. [5]Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [6]The Institute of Biosciences and Technology, Department of Molecular and Cellular Medicine, Texas A&M University Health Science Center, Houston, TX, USA. [7]Department of Research and Development, Central Blood Institute, Blood Service Headquarters, Japanese Red Cross Society, Tokyo, Japan. [8]Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan. [9]These authors contributed equally: Xiaohui Zhang, Biyun Zhu, Liang Chen. ✉e-mail: dlli@bio.ecnu.edu.cn

The average C-to-T conversion efficiency was similar at positions 2–6, but the activity was increased 1.9~14-fold at positions 7–17 in A&C-BEmax-treated cells compared to AID-BE4max (Fig. 1f). Analyzing the first 20 targets containing both As and Cs in the editing window revealed that the simultaneous A/C mutation rate on the same allele ranged from 2% to 30%, and the proportion of alleles bearing only C-to-T or A-to-G varied from 5.3% to 82.6% and from 0.2% to 10%, respectively (Fig. 1g and Supplementary Fig. 9). Higher simultaneous A/C conversion rates were observed in sites containing adenines at position 6–7 within the target (Fig. 1g and Supplementary Fig. 9). Robust base editing efficiency was observed in HeLa cells at all examined targets, suggesting that A&C-BEmax is a general tool functioning in different cell types (Supplementary Fig. 10).

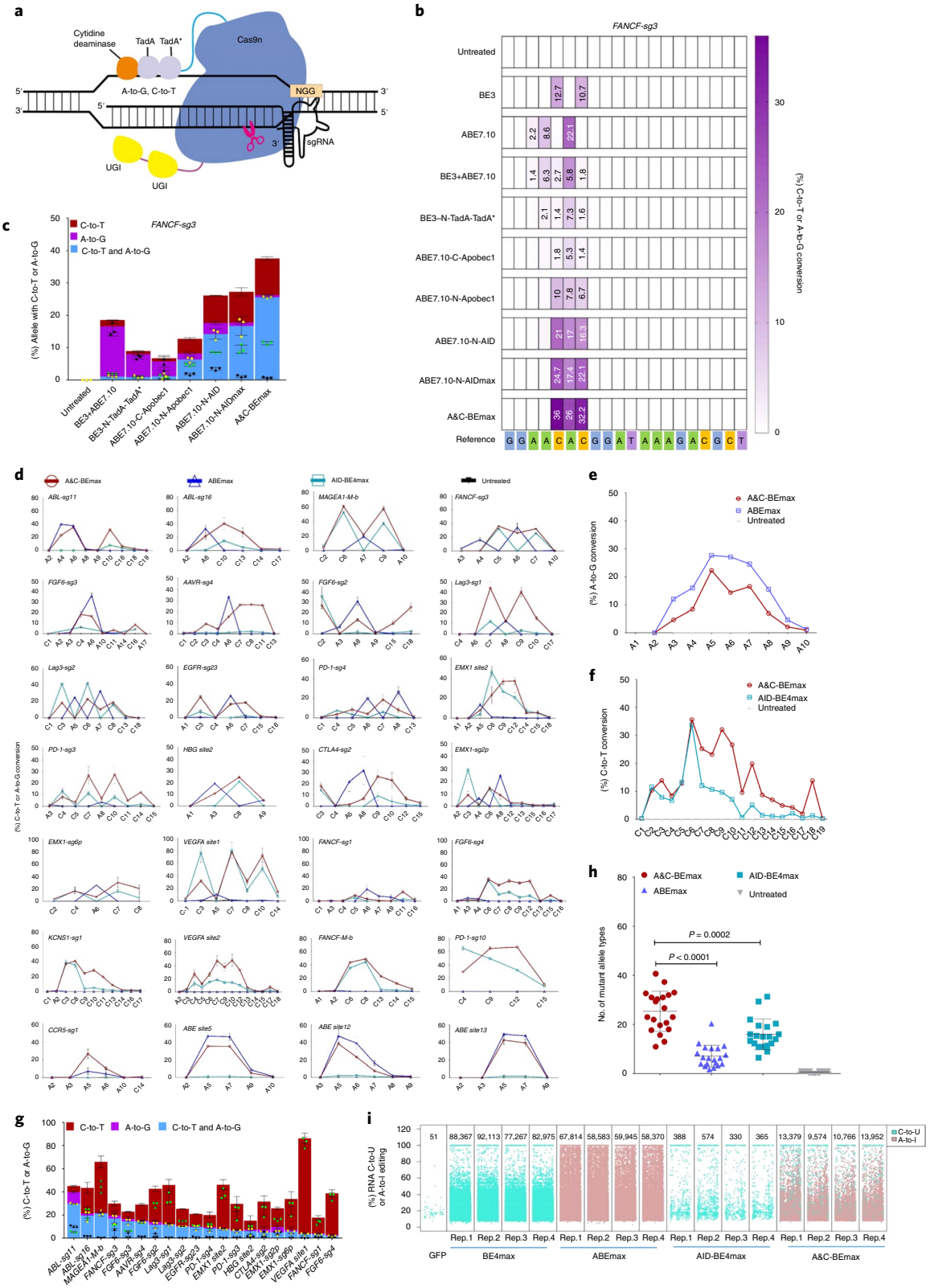
By further comparing the mutation spectra created by A&C-BEmax, ABEmax or AID-BE4max, we found that A&C-BEmax generated more mutant allele types (Fig. 1h). Moreover, NGS studies showed that A&C-BEmax did not generate more DNA indels than AID-BE4max (Supplementary Fig. 11) and exhibited similar off-target effects with AID-BE4max and fewer than Cas9 as determined after analyzing 88 potential off-target sites, which were selected either by Cas-OFFinder software prediction[6] or experimental identification by GUIDE-seq[1], ChIP-seq[3] or Digenome-seq[7] through NGS studies (Supplementary Fig. 12). However, further studies should investigate whether A&C-BEmax also induces significant unpredictable DNA off-targeting effects similar to BE3 (refs. [8,9]), because a different cytidine deaminase was incorporated. As recent studies demonstrated that base editors (except for AID-BE3) induced extensive RNA single-nucleotide variants (SNVs)[10,11], transcriptome-wide off-target effects of A&C-BEmax were determined. Consistently, BE4max and ABEmax induced tens of thousands of RNA SNVs, whereas AID-BE4max generated fewer RNA mutations. A&C-BEmax had greatly reduced A-to-I RNA off-targeting, less than 20% of that of the ABEmax-treated group (Fig. 1i). The above data demonstrate that A&C-BEmax has some advantages over traditional base editors, including simultaneous A/C mutation on the same allele, higher editing activity, a wider window and less RNA off-targeting.

To further investigate the application of A&C-BEmax for gene therapy, a β-hemoglobinopathy model was employed. It is well accepted that reactivation of fetal hemoglobin (HbF) is a feasible strategy for the treatment of sickle cell disease and β-thalassemia[12]. Patients with β-thalassemia with mutations −114C-to-T or −113A-to-G in the promoter of the γ-globin genes (*HBG1* and *HBG2*) have been identified with increased HbF production and reduced symptoms[12]. The −114 or −115 C-to-T mutation disrupts the BCL11A binding site, which is a strong transcription repression element responsible for *HBG1* and *HBG2* inhibition in adults[13]. The −113A-to-G mutation does not disrupt the BCL11A site but creates a new GATA1 binding site that has been demonstrated to activate *HBG1* transcription[14]. To test the feasibility of A&C-BEmax to disrupt the BCL11A binding site (−114 or −115) and create a GATA1-site (−113) simultaneously, an sgRNA targeting the *HBG1* and *HBG2* promoter was designed and tested in HEK293T cells (Fig. 2a). A&C-BEmax induced variant mutations besides −115 to −113 and exhibited a 1.8-fold and 1.7-fold increase in the editing efficiency at −114C (41.2% versus 14.7%) and −115C (39.2% versus 14.8%), respectively, over AID-BE4max and similar efficiency at −113A with ABEmax, yielding an 8.76% rate of A/C simultaneous mutation (including −113A with either −114C or −115C) on the same allele (Fig. 2b and Supplementary Fig. 13). However, co-transfection of ABE and CBE hardly ever generated A/C simultaneous mutation at either the −115/−113 or the −114/−113 sites (Fig. 2b and Supplementary Fig. 13). Next, an erythroid precursor cell line HUDEP-2 (ref. [15]) was employed to test the physiological function of the mutation generated by lentivirus-packaged A&C-BEmax (Supplementary Fig. 14 and Supplementary Sequence 3). A&C-BEmax induced highly efficient base editing, strongly promoting HBG mRNA induction (A&C-BEmax, 73.7% versus AID-BEmax, 45.9%) as a combined effect of several types of mutation in the promoter that disrupted the BCL11A binding site and/or generated a GATA1 site in the promoter (Fig. 2c,d and Supplementary Fig. 15). To further verify whether simultaneous A/C mutation induced higher HBG reactivation, single clones of HUDEP-2($\Delta^{G}\gamma$) cells, which carry only one γ-globin gene to facilitate genotype analysis[16], were established after A&C-BEmax/sgRNA delivery. After analysis of single clones, we demonstrated that simultaneous −113A/−114C conversion (#B-7) induced the highest HBG expression, and additional mutations at other sites (#B-7 versus #B-1) did not show extra induction (Fig. 2e,f). These data constitute also a proof of principle showing that A&C-BEmax was able to generate sequence variety to investigate the relationship of genome type with function, suggesting that A&C-BEmax could be used as a tool to dissect the function of a given sequence at single-nucleotide resolution.

To explore the therapeutic potential of A&C-BEmax to correct pathogenic mutations, we computationally profiled all clinically relevant variants in Clin Var[17] and single-nucleotide polymorphisms (SNPs) in dbSNP[18] potentially able to be targeted by A&C-BEmax using our recently reported strategy[19]. Our in silico analysis found 203 target sites containing known pathogenic A-to-G mutation(s) and C-to-T mutation(s) that could potentially be reverted by A&C-BEmax through a single sgRNA, albeit the SNVs might be located on different alleles (Supplementary Fig. 16 and Supplementary Table 1). The

Fig. 1 | Design and optimization of A&C-BE. a, Schematic diagram of A&C-BE construction. Cytidine deaminase refers to rat APOBEC1 or human AID. mTadA, evolved TadA (TadA*). Cas9n, Cas9D10A. **b**, Heat maps showing C-to-T or A-to-G base editing efficiency for base editors at the endogenous *FANCF-sg3* target site in HEK293T cells. Data represent the means from three independent experiments. **c**, Comparison of the products distribution among edited DNA sequencing reads of the *FANCF-sg3* target edited by variant base editors. The individual data points are shown as light green (only C-to-T), black (only A-to-G) and yellow (simultaneous C-to-T and A-to-G) dots. Values and error bars reflect the means and s.d. of three independent experiments. **d**, Base editing efficiencies of ABEmax, AID-BE4max and A&C-BEmax at 28 endogenous human genomic loci. **e**, Merged data of average C-to-T editing efficiency at 24 targets in **d** (except *CCR5-sg1*, *ABE* site5, *ABE* site12 and *ABE* site13 with no Cs in the editing window) edited by AID-BE4max or A&C-BEmax. Data represent the means from three independent experiments. Values and error bars reflect the means and s.d. of three independent experiments. **f**, Merged data of average A-to-G editing efficiency at 24 targets in **d** (except *KCNS1-sg1*, *VEGFA* site2, *FANCF-M-b* and *PD-1-sg10* with no As in the editing window) edited by AID-BE4max or A&C-BEmax. Data represent the means from three independent experiments. **g**, The composition of A&C-BEmax base editing products at 20 endogenous human genomic loci. The individual data points are shown as light green (only C-to-T), black (only A-to-G) and yellow (simultaneous C-to-T and A-to-G) dots. Values and error bars reflect the means and s.d. of three independent experiments. Statistical source data are provided in Source Data Fig. 1. **h**, Mutation allele types yielded by ABEmax, AID-BE4max and A&C-BEmax at 20 endogenous target sites. Each data point represents average mutation allele types at each target site (except target sites with only As or only Cs in the editing window) calculated from three independent experiments. Data are means ± s.d. *P* value was determined by two-tailed Student's *t*-test. **i**, Jitter plots from RNA-sequencing experiments in HEK293T cells showing efficiencies of C-to-U or A-to-I conversions (*y* axis) with BE4max, ABEmax, AID-BE4max and A&C-BEmax expression or a GFP negative control. Total number of modified bases is listed at the top. Each biological replicate is listed on the bottom.

targeting scope increased by 2.8-fold (~573) when the Cas9-NG variant[20] was leveraged for the analysis (Supplementary Fig. 16 and Supplementary Table 1). Moreover, if we expand the category to one pathogenic mutation plus an SNP with unknown function within the editing window, 3,831 targets are potentially restorable by A&C-BEmax (10,784 targets for Cas9-NG variant), suggesting broad applications of A&C-BEmax for gene therapy. Additionally, A&C-BEmax efficiently generates simultaneous A/C conversions that significantly increase the variants number of a given triplet code compared to ABE or CBE (Supplementary Fig. 17 and Supplementary Table 2).



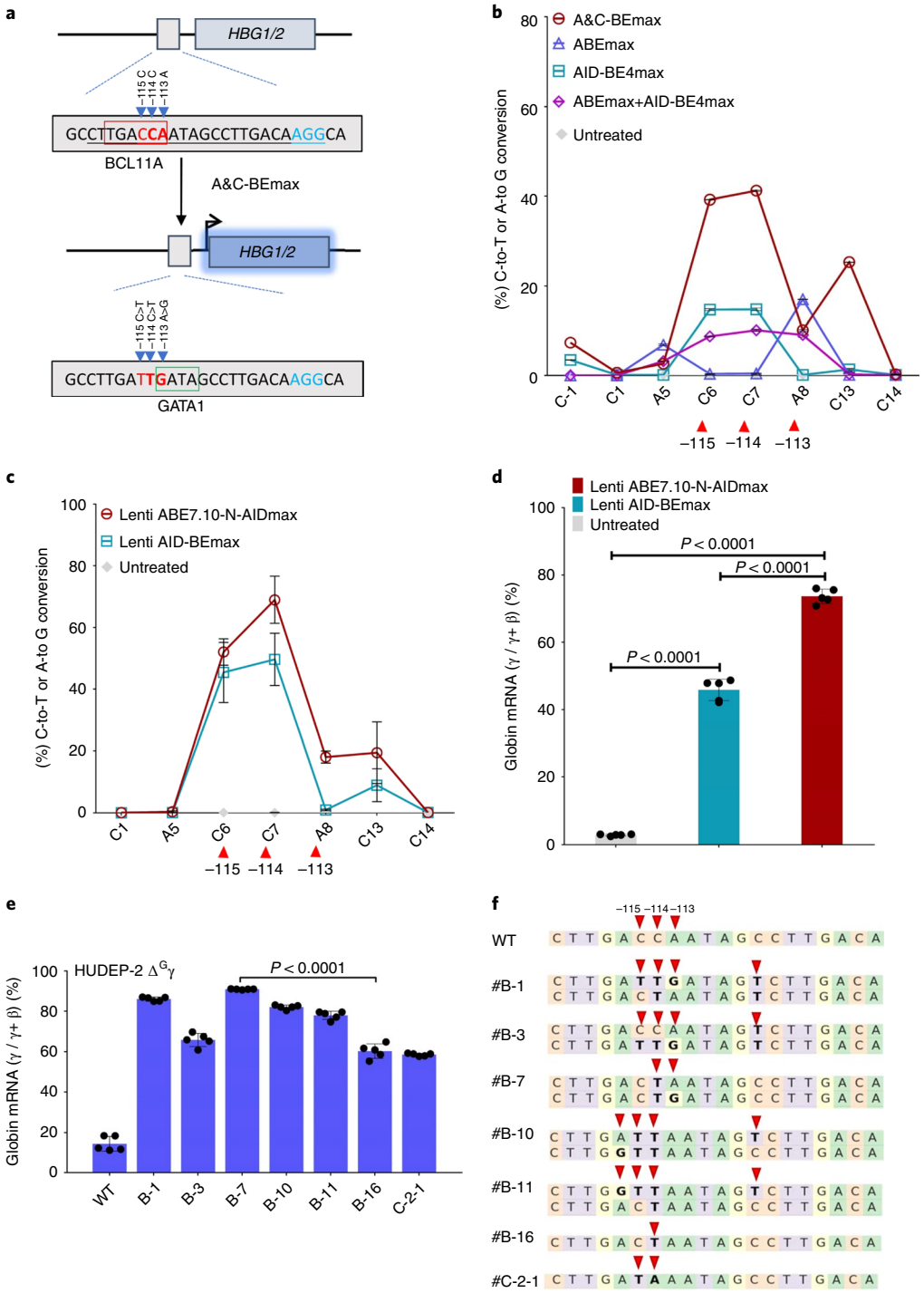


Fig. 2 | Efficient editing of *HBG* promoter by A&C-BEmax in HUDEP-2 cells. a, Schematic diagram of the critical sequence regulating *HBG1* and *HBG2* expression. The core sequence of the BCL11A binding site is boxed in red. The GATA1 binding site is boxed in green. The PAM sequence of the target site is in blue. **b**, Editing efficiencies of A&C-BEmax, ABEmax+AID-BE4max, ABEmax or AID-BE4max in the *HBG1* and *HBG2* promoter in HEK293T cells. Values and error bars reflect the means and s.d. of three independent experiments. **c**, Base editing efficiency at the *HBG1* and *HBG2* promoter site in pooled HUDEP-2 cells transduced with lentiviral ABE.7.10-N-AIDmax or AID-BEmax after puromycin selection. Values and error bars reflect the means and s.d. of three independent experiments. **d**, Comparison of γ-globin mRNA expression relative to β-like globin mRNA via ABE7.10-N-AIDmax or AID-BEmax treatment in HUDEP-2 cells after differentiation. Values and error bars reflect the means and s.d. of five independent experiments. *P* value was determined by two-tailed Student's *t*-test. **e**, γ-globin mRNA expression relative to β-like globin mRNA in individual single clones of HUDEP-2($\Delta^{G}\gamma$) cells treated with the A&C-BEmax and sgRNA vector. Values and error bars reflect the means and s.d. of five independent experiments. *P* value was determined by two-tailed Student's *t*-test. The genotype of each clone is presented in **f**. The comparison between clone #B-7 and #B-16 indicates that a de novo GATA1 site is critical for HBG induction.

In summary, through tethering two base deaminases, we developed a dual-functional base editor, A&C-BEmax, which can induce simultaneous C-to-T and A-to-G conversions with increased CBE activity and reduced RNA off-targeting (compared to ABEmax). A&C-BEmax is a valuable tool not only for dissecting genomic sequence function at a single base resolution but also for the therapy of genetic disorders.

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/s41587-020-0527-y.

Received: 8 August 2019; Accepted: 17 April 2020;
Published online: 01 June 2020

References

1. 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).
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. Rees, H. A. & Liu, D. R. J. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
4. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
5. 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).
6. 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).
7. Gehrke, J. M. et al. An APOBEC3A–Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
8. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
9. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
10. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
11. Grünewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).
12. Wienert, B., Martyn, G. E., Funnell, A. P., Quinlan, K. G. & Crossley, M. J. Wake-up sleepy gene: reactivating fetal globin for β-hemoglobinopathies. *Trends Genet.* **34**, 927–940 (2018).
13. 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).
14. Martyn, G. E. et al. A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. *Blood* **133**, 852–856 (2019).
15. Kurita, R. et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PloS ONE* **8**, e59890 (2013).
16. Wienert, B. et al. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood* **130**, 803–807 (2017).
17. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
18. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
19. Wang, Y. et al. Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. *Genome Biol.* **20**, 218 (2019).
20. Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).

Methods

Plasmid construction. Primers and DNA sequences used in this article are listed in the Supplementary Tables and Supplementary Sequences. Human codon-optimized AID was synthesized by Genewiz. ABE7.10 (no. 102919), BE3 (no. 73021), ABEmax (no. 112095), PX458 (no. 48138) and lentiCRISPR v2 (no. 52961) were purchased from Addgene. PCR was performed using KOD-Plus-Neo DNA Polymerase (Toyobo, code no. KOD-401). BE plasmids generated in this study were constructed using ClonExpress MultiS One Step Cloning Kit (Vazyme) (Supplementary Sequences 2 and 3). sgRNA expression and reporter plasmids were constructed as previously described[21]. Briefly, oligonucleotides listed in Supplementary Table 3 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 or into NheI- and BamHI-linearized CMV-T2A-non-ATG-luciferase for reporter plasmids (Thermo). Plasmid sequences of U6-sgRNA(sp)-EF1α-GFP and CMV-T2A-non-ATG-luciferase are listed in Supplementary Sequence 1. A step-by-step protocol for generation of site-specific point mutations by A&C-BEmax can be found at the Nature Protocol Exchange[22].

Cell culture. HEK293T (ATCC CRL-3216) and HeLa (ATCC CCL-2) cell lines were kept in DMEM (Gibco) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Gibco). HUDEP-2 cells were maintained and expanded in serum-free expansion medium (Stem Cell Technologies) supplemented with human Stem Cell Factor (SCF, 50 ng ml^{-1}, PeproTech), erythropoietin (EPO, 3 IU ml^{-1}, PeproTech), dexamethasone (1 μM, Sigma), doxycycline (1 μg ml^{-1}, Takara Bio) and 2% penicillin–streptomycin (Gibco). All cell lines used were maintained at 37 °C, 5% CO_2 in the incubator.

Cell transfection and fluorescence-activated cell sorting. For reporter assays and DNA on-target or off-target base editing experiments, HEK293T or HeLa cells were seeded into 24-well plates (96-well plates for reporter assay) (Corning) and transfected at approximately 80% confluency. Next, 750 ng of A&C-BE and 250 ng of sgRNA expression plasmids (90 ng A&C-BE with sgRNA and 10 ng reporter plasmids for reporter assay) were co-transfected using polyethyleneimine (PEI, Polysciences) following the manufacturer's recommended protocol. Two days later, reporter luciferase activity was detected. Three days later, transfected cells were digested with 0.25% trypsin (Gibco) for genomic DNA extraction. Genomic DNA was isolated using the TIANamp Genomic Kit (Tiangen Biotech) according to the manufacturer's instructions. For RNA off-target analysis, HEK293T cells were seeded into 10-cm dishes and transfected with 30 μg of Cas9n-P2A-GFP, BE4max-P2A-GFP, ABEmax-P2A-GFP, AID-BE4max-P2A-GFP and A&C-BEmax-P2A-GFP 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 an 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 Note). Then, ~500,000 cells (top 8% GFP signal) were collected, and RNA was extracted according to standard protocols.

High-throughput DNA sequencing and data analysis. On- and off-target genomic regions of interest were amplified from 50~100 ng of genomic DNA by PCR with primers listed in Supplementary Tables 4 and 5. High-throughput sequencing (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′; backward 5′-GAGTTGGATGCTGGATGG-3′) at the 5′ end (Supplementary Tables 4 and 5). The above products were then subjected to a second-round PCR using primers containing different barcode sequences. The resulting libraries were mixed and sequenced on an Illumina HiSeq platform. For sequencing data analysis, the reference sequence was set from 10 bp upstream of the protospacer to 10 bp downstream of the protospacer-adjacent motif (PAM) sequence. The analysis method was described previously[1]. The alleles containing combined (C-to-T and A-to-G) or exclusive (only C-to-T or A-to-G) conversions and indels were quantitated using BE-analyzer[23] or CRISPResso2 (ref. [24]).

RNA-sequencing experiments. A total amount of 3 μg of RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using the NEBNext UltraTM 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, NEBNext Adaptor with 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 an AMPure XP system (Beckman Coulter). Then, 3 μl of USER Enzyme (NEB) was used 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 Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) 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-sequencing data was performed as previously described[10] 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 base with Trimmomatic. At the same time, Q20, Q30 and GC content of the clean data were calculated. All the downstream analyses were based on the clean data with a high quality. The index of the reference genome was built using Hisat2 v2.0.5, and paired-end clean reads were aligned to the reference genome (Ensemble GRCh38) using Hisat2 v2.0.5. We selected Hisat2 as the mapping tool because Hisat2 can generate a database of splice junctions based on the gene model annotation file and can thus produce a better mapping result than other non-splice mapping tools. GATK (v4.0) software was used to perform SNP 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 SNV in a control experiment should be >90th percentile of the read coverage across all SNVs 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 more than 0% of reads containing an alternate allele. Base edits labeled 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. Base edits labeled as A-to-I comprise A-to-I edits called on the positive strand as well as T-to-C edits sourced from the negative strand.

HUDEP-2 cell differentiation. HUDEP-2 cells were differentiated in erythroid differentiation media (IMDM, Corning) supplemented with 2% human blood type AB plasma (SeraCare), 1% L-glutamine, 2 IU ml^{-1} heparin, 10 μg ml^{-1} recombinant human insulin, 3 IU ml^{-1} EPO, 330 μg ml^{-1} holo-human transferrin (Sigma-Aldrich), 100 ng ml^{-1} SCF, 1 μg ml^{-1} doxycycline and 2% penicillin–streptomycin. On day 8 of the differentiation, cells were harvested for total mRNA isolation. The detailed protocol is available at the Nature Protocol Exchange[22].

mRNA preparation and quantitative PCR. For HEK293T cells, total mRNA was isolated 5 d after A&C-BE transfection using RNAiso Plus (Takara). For HUDEP-2 cells, 1 μg ml^{-1} puromycin was added into the expansion medium 48 h after lentiviral transduction. Cells were screened for 15 d, followed by total mRNA isolation. Isolated mRNA was reverse transcribed using HiScript II Q RT SuperMix (Vazyme). qPCR was performed on the QuantiStudio 3 real-time PCR system (ABI), and mRNA expression levels were calculated by normalizing to β-actin using the ΔΔCt method. Base editor mRNA levels were further adjusted by transfection efficiency as determined by qPCR amplification of the BGH polyadenylation sequence present on the base editor plasmids. HBG and HBB mRNAs were quantified by SYBR Green qPCR. qPCR primers are listed in Supplementary Table 6.

Western blotting. For western blot, HEK293T cells were lysed 5 d after transfection using RIPA buffer supplemented with proteinase and phosphatase inhibitors. Total protein was quantified using the BCA kit (Thermo Fisher Scientific). Ten μg per well of total protein was separated by electrophoresis using a 15-well 10% Tris gel and dry-transferred to a nitrocellulose membrane for 7 min at 20 V before blocking with Tris-buffered saline with 0.05% Tween-20 containing 1% bovine serum albumin. 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 goat anti-rabbit IgG H&L (IRDye 800CW) (Abcam, ab216773) for 1 h and visualized using an Odyssey imager (Supplementary Fig. 4b).

Lentivirus production and transduction of cell lines. Lentivirus production was performed as previously described[25]. 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 ABE7.10-N-AIDmax or Lenti AID-BEmax), 5 μg pMD2.G and 7.5 μg psPAX2. Virus-containing supernatant was harvested at 48 h and 72 h after transfection. Supernatant was centrifuged at 8,000 r.p.m. for 10 min at 4 °C to precipitate cell debris, filtered by passing through a 0.45-mm 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.

Lentiviral titration. Virus stock was diluted via five serial ten-fold dilutions with DMEM (10% FBS). For each viral construct, 1×10^4 HEK293T cells were first digested and suspended. Cells were spun down, resuspended with the diluted virus (100 µl) and seeded into 96-well plates. Control cells were resuspended with DMEM (10% FBS) only. Three days after transduction, cells were analyzed by checking the EGFP fluorescence via Fortessa Flow Cell Analyzer (BD Biosciences). Virus titration was calculated as follows: titer (TU per ml) = cell number × (%) EGFP × 10^3 per virus stock volume (µl).

Statistics and reproducibility. All statistical analyses were performed on at least $n = 3$ biologically independent experiments using an unpaired two-tailed Student's *t*-test through GraphPad Prism 6 (GraphPad Software). Detailed information on exact samples sizes and experimental replicates can be found in the individual figure legends and the Life Sciences Reporting Summary that is attached to this article. *P* values smaller than 0.05 were considered significant.

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

Data availability

Targeted amplicon sequencing data have been deposited in the NCBI Sequence Read Archive database under accession codes PRJNA558944, PRJNA559260, PRJNA559237, PRJNA559051, PRJNA592341 and PRJNA592333. The RNA-sequencing data used in this study have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA592597. Plasmids encoding A&C-BEmax, AID-BE4max, Lenti ABE7.10-N-AIDmax and Lenti AID-BEmax are available from Addgene.

References

21. 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).
22. Biyun, Z. et al. *Protoc. Exch.* https://doi.org/10.21203/rs.3.pex-877/v1 (2020).
23. Hwang, G.-H. et al. Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* **19**, 542 (2018).
24. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
25. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

Acknowledgements

We thank M. Crossley at the University of New South Wales, Australia, for providing the HUDEP-2($\Delta^G\gamma$) cell line and H. Yang and C. Zhou at the Institute of Neuroscience, Chinese Academy of Sciences, for the suggestions of RNA off-targeting analysis. This work was partially supported by grants from the National Key R&D Program of China (2019YFA0110802 to D.L. and 2019YFA0802800 to M.L.), the National Natural Science Foundation of China (81670470 and 81873685 to D.L., and 31925011 and 91940306 to L.Y.), grants from the Shanghai Municipal Commission for Science and Technology (18411953500 to D.L.), the Innovation Program of the Shanghai Municipal Education Commission (2019-01-07-00-05-E00054 to D.L.), the support of the ECNU Public Platform for Innovation (011) and Fundamental Research Funds for the Central Universities.

Author contributions

X.Z., D.L. and M.L. designed the experiments and analyzed the data. X.Z., B.Z., L.C., L.X., L.L., W.H., S.Y., L.Y., Ha.H., Ho.H., Y.L., L.W., X.M., X.P. and Y.W. performed experiments. X.Z., W.Y., Z.Y. and G.C. analyzed the transcriptome data. R.K. and Y.N. generated HUDEP-2 cells. Y.W. and Li.Y. performed bioinformatic analysis to profile A&C-BEmax-editable variants and SNPs. D.L. and M.L. supervised the research. D.L., X.Z., S.S., H.G. and B.Z. analyzed the data and wrote the manuscript with input from all the authors.

Competing interests

The authors have submitted patent applications (nos. 201810929391.3, 201810930979.0 and 201810930980.3) based on the results reported in this study.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0527-y.

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

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

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. 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 or A>G or C>T and A>G) and indels efficiencies. Potential DNA off-targets site for A&C-BEmax 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 . 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

Targeted amplicon sequencing data have been deposited in the NCBI Sequence Read Archive database under Accession Code PRJNA558944, PRJNA559260, PRJNA559237, PRJNA559051, PRJNA592341 and PRJNA592333. The RNA-seq data used in this study have been deposited in the NCBI Sequence

Read Archive database under Accession Code PRJNA592597.

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 was excluded.
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 1:10000 dilution of Anti-GAPDH (Abcam, ab9485) a 1:5000 dilution of anti-CRISPR-Cas9 (Abcam, ab189380).
Validation	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).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK 293T (purchased from ATCC CRL-3216) Hela (purchased from ATCC® CCL-2™) HUDEP-2 cell line (Ryo Kurita and Yukio Nakamura, Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki,Japan) HUDEP-2(ΔGγ) cell line (University of New South Wales, Merlin Crossley 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.

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 passed through a 45µm cell strainer cap before sorting (72h after transfection).
Instrument	FACSAria III (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 ~ 50-60% GFP+ (of gated population = % parent in BD FACSDiva).
Gating strategy	Gates were established using uninfected control cells and GFP positive control. Gates were drawn to collect subsets of GFP-expressing cells. cells with top 8% of GFP signal were sorted, after gating for the cell population (~8% of parent). Please see the Supplementary Information for gating strategies in different biology replicates.

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