Supplementary Information

Title: Engineering TadA ortholog-derived cytosine base editor without motif preference and adenosine activity limitation

Authors list:

Guoling Li[1†], Xue Dong[1†], Jiamin Luo[1†], Tanglong Yuan[2†], Tong Li[1†], Guoli Zhao[3,4†], Hainan Zhang[1], Jingxing Zhou[1], Zhenhai Zeng[3,4], Shuna Cui[1], Haoqiang Wang[1], Yin Wang[1], Yuyang Yu[1], Yuan Yuan[1], Erwei Zuo[2*], Chunlong Xu[5*], Jinhai Huang[3,4*], Yingsi Zhou[1*]

Affiliations:

[1]HuidaGene Therapeutics Co., Ltd., Shanghai 200131, China

[2]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China.

[3]Eye Institute and Department of Ophthalmology, Eye and ENT Hospital, Fudan University, Key Laboratory of Myopia, Chinese Academy of Medical Sciences, Shanghai, 200030 China

[4]Shanghai Research Center of Ophthalmology and Optometry, Shanghai, 200030 China

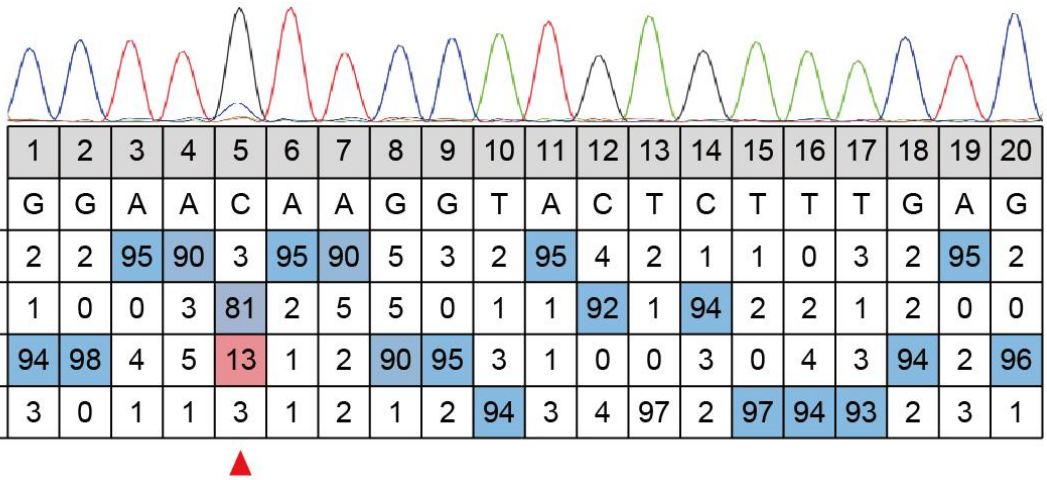
[5]Lingang Laboratory, Shanghai, China.

[†]These authors contributed equally to this work.

[*]Corresponding authors. zuoerwei@caas.cn (E.Z.); xucl@lglab.ac.cn (C.X.); jinhaihuang@fudan.edu.cn (J.H.); yingsizhou@huidagene.com (Y.Z.)

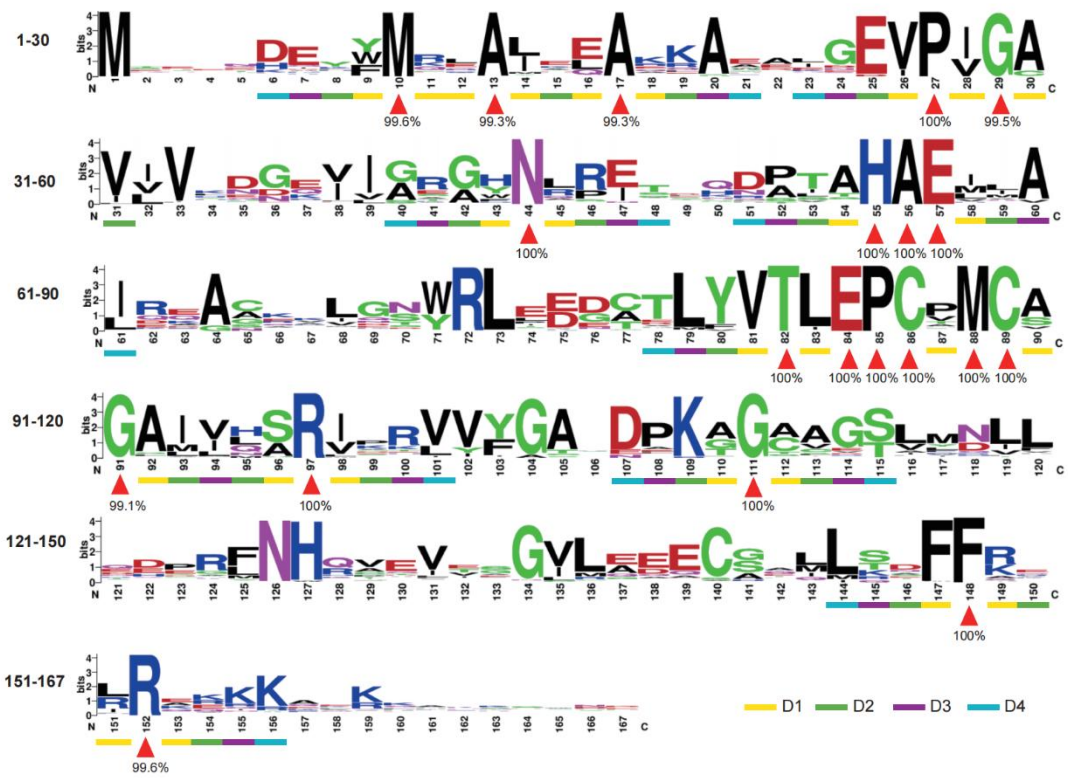
This PDF file includes:

Supplementary Figure 1 to 11



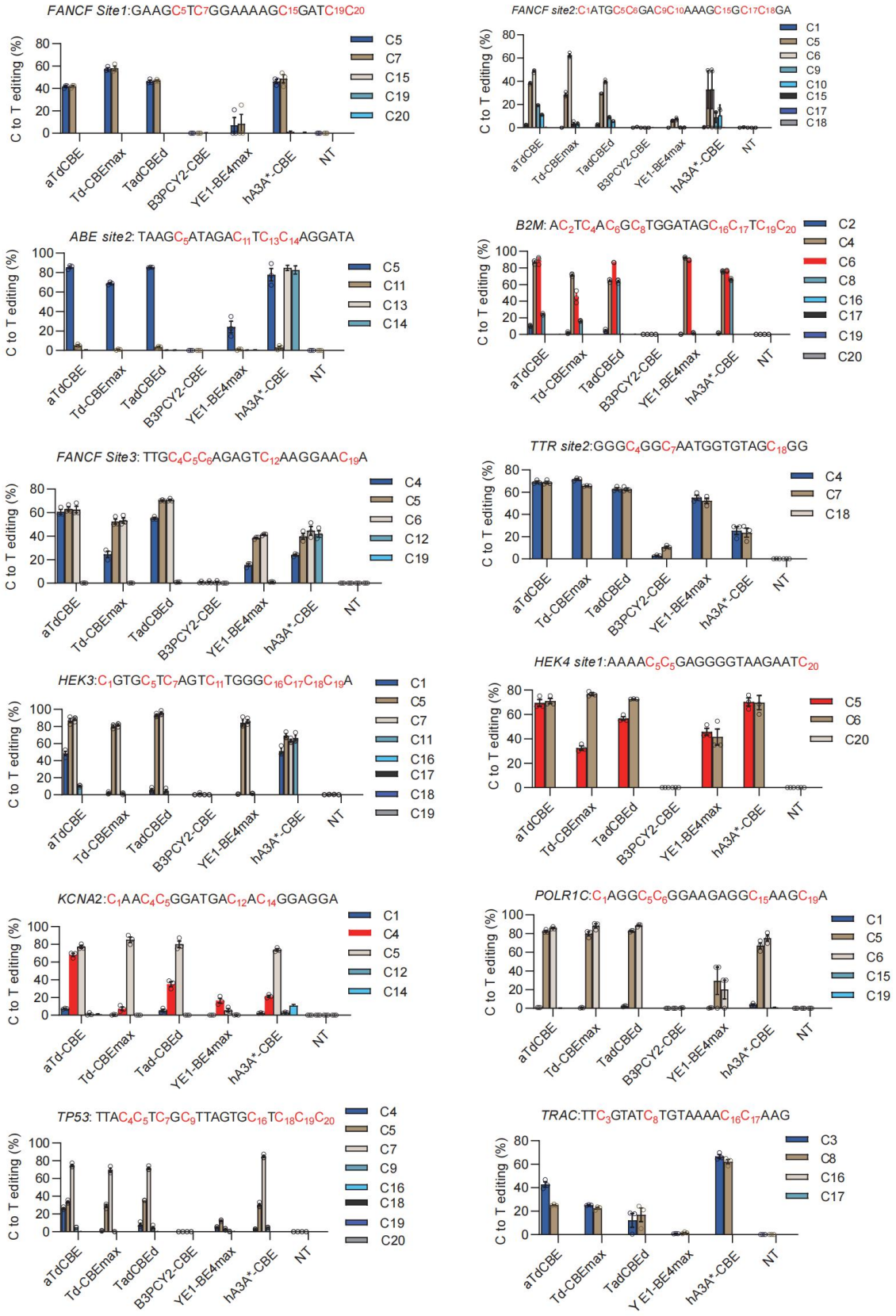
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	G	G	A	A	C	A	A	G	G	T	A	C	T	C	T	T	T	G	A	G
A	2	2	95	90	3	95	90	5	3	2	95	4	2	1	1	0	3	2	95	2
C	1	0	0	3	81	2	5	5	0	1	1	92	1	94	2	2	1	2	0	0
G	94	98	4	5	13	1	2	90	95	3	1	0	0	3	0	4	3	94	2	96
T	3	0	1	1	3	1	2	1	2	94	3	4	97	2	97	94	93	2	3	1

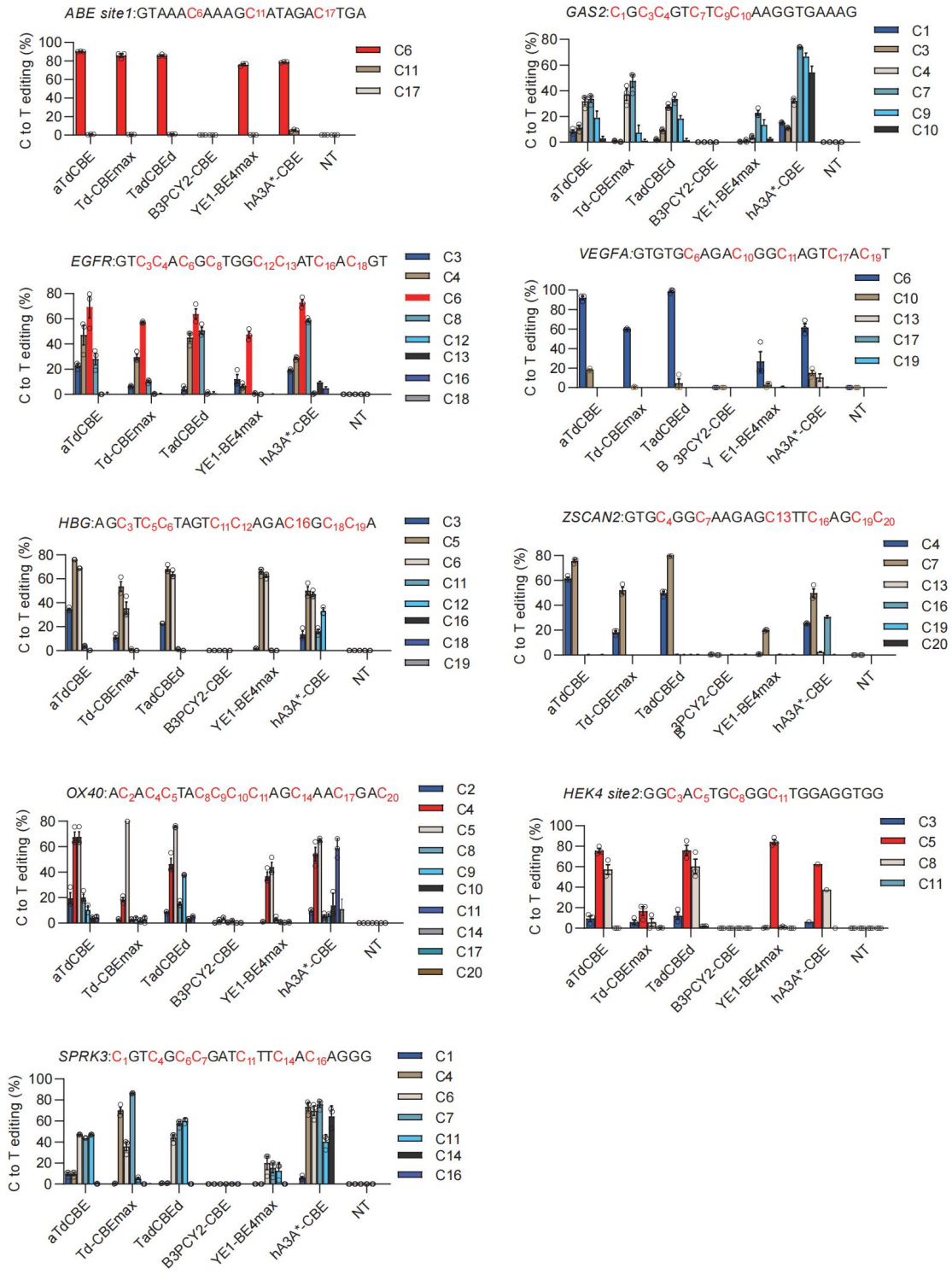
Supplementary Fig.1 Sanger sequencing of AjTadA.v1 at *PIK3CA* loci.



Supplementary Fig.2 Conservative site analysis of TadA protein sequence based on multiple sequence alignment.

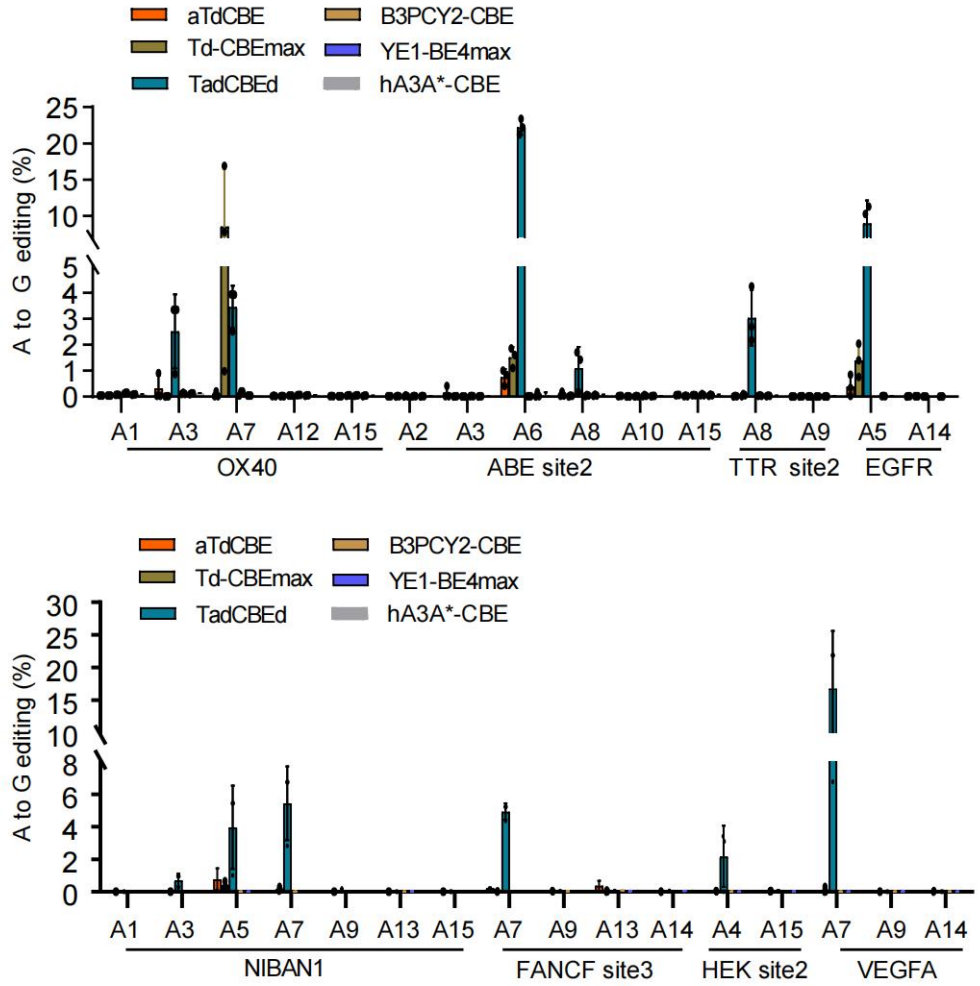
The red triangle indicates highly conserved amino acids. The yellow, green, purple and blue squares represent adjacent amino acids with different distances from highly conserved sites. The amino acid frequency visualized by webogo(https://weblogo.berkeley.edu/logo.cgi).


FANCF Site1:GAAGC5TC7GGAAAAGC15GATC19C20
C5
C7
C15
C19
C20
FANCF site2:C1ATGC5C6GAC9C10AAAGC15GC17C18GA
C1
C5
C6
C9
C10
C15
C17
C18
ABE site2: TAAGC5ATAGAC11TC13C14AGGATA
C5
C11
C13
C14
B2M: AC2TC4AC6GC8TGGATAGC16C17TC19C20
C2
C4
C6
C8
C16
C17
C19
C20
FANCF Site3: TTGC4C5C6AGAGTC12AAGGAAC19A
C4
C5
C6
C12
C19
TTR site2:GGGC4GGC7AATGGTGTAGC18GG
C4
C7
C18
HEK3:C1GTGC5TC7AGTC11TGGGC16C17C18C19A
C1
C5
C7
C11
C16
C17
C18
C19
HEK4 site1:AAAAC5C5GAGGGGTAAGAATC20
C5
C6
C20
KCNA2:C1AAC4C5GGATGAC12AC14GGAGGA
C1
C4
C5
C12
C14
POLR1C:C1AGGC5C6GGAAGAGGC15AAGC19A
C1
C5
C6
C15
C19
TP53: TTAC4C5TC7GC9TTAGTGC16TC18C19C20
C4
C5
C7
C9
C16
C18
C19
C20
TRAC:TTC3GTATC8TGTAAAAC16C17AAG
C3
C8
C16
C17
C to T editing (%)
aTdCBE
Td-CBEmax
TadCBEd
B3PCY2-CBE
YE1-BE4max
hA3A*-CBE
aTd-CBE
Tad-CBEd
NT



Supplementary Fig.3 ATdCBE enables robust C-to-T genomic editing in mammalian cells.

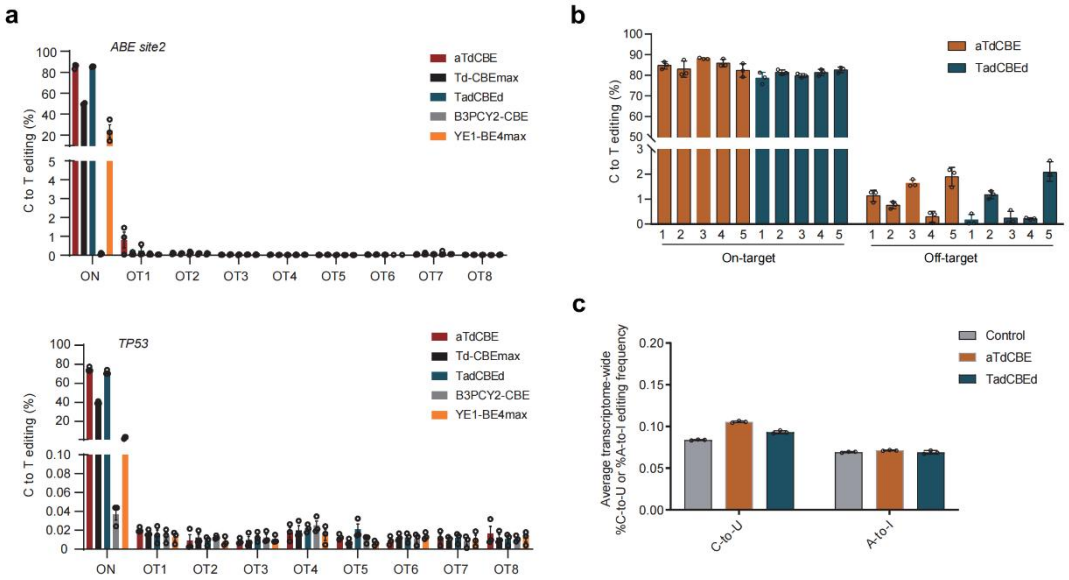
Data are presented as means ± S.D. Values represent n = 3 independent biological replicates. All of the above base editors contain UGI. Source data are provided as a Source Data file.



77

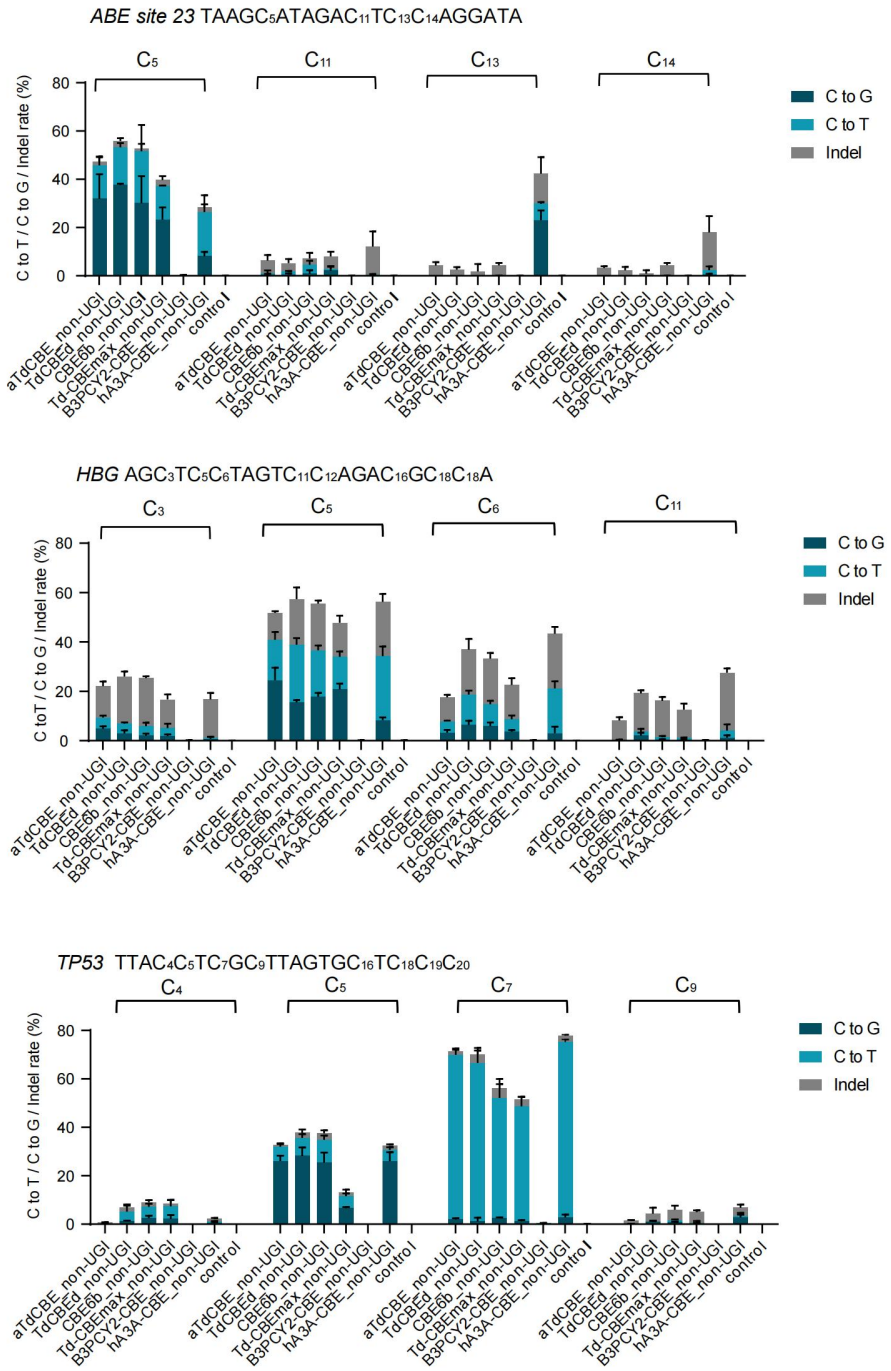
78 **Supplementary Fig.4 A-to-G genomic editing of base editors in**
79 **mammalian cells.**

80 Data are presented as means ± S.D. Values represent n = 3 independent
81 biological replicates. All of the above base editors contain UGI.
82 Source data are provided as a Source Data file.



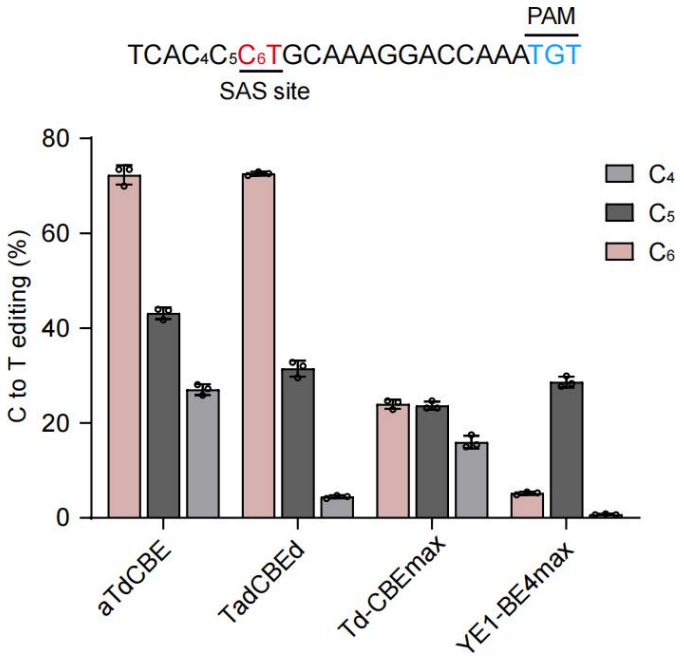
Supplementary Fig.5 AjTadA-derived base editors with high specificity in mammalian cells.

a. The gRNA-dependent off-target levels at the potential off-target sites. b**.** The gRNA-independent off-target activity at five R-loops formed by dSaCas9. **c.** Transcriptome-wide off-target analysis of aTdCBE and TadCBEd. Data are presented as means ± S.D. Values represent n = 3 independent biological replicates. All of the above base editors contain UGI. Source data are provided as a Source Data file.



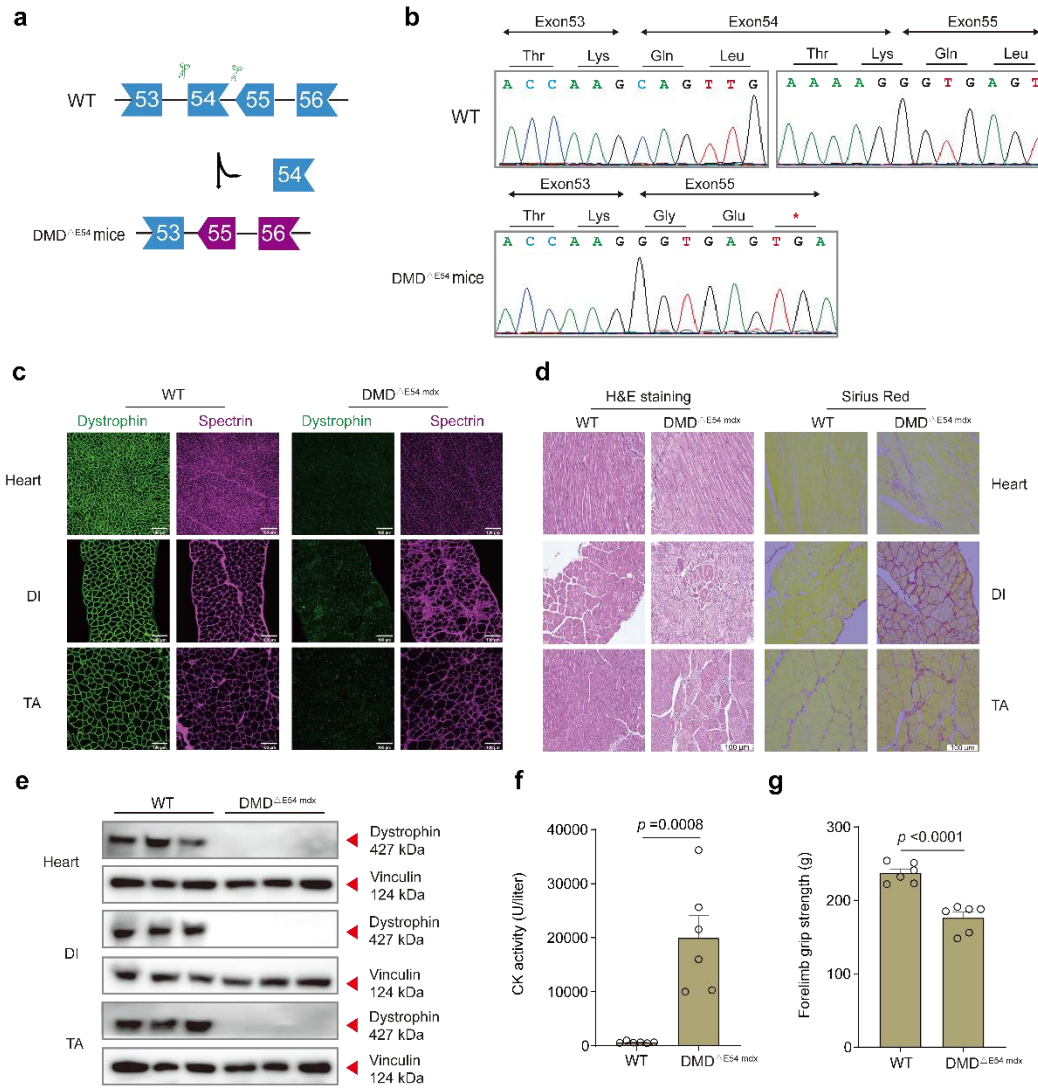
93

94 **Supplementary Fig.6 Compare the editing products of the base**
95 **editors without UGI at three endogenous loci.**
96 Data are presented as means ± S.D. Values represent n = 3 independent
97 biological replicates. All of the above base editors do not contain UGI.
98 Source data are provided as a Source Data file.



Supplementary Fig.7 Comparing the editing efficiency of CBEs at the exon 55 SAS of *DMD* gene in HEK293T cells.

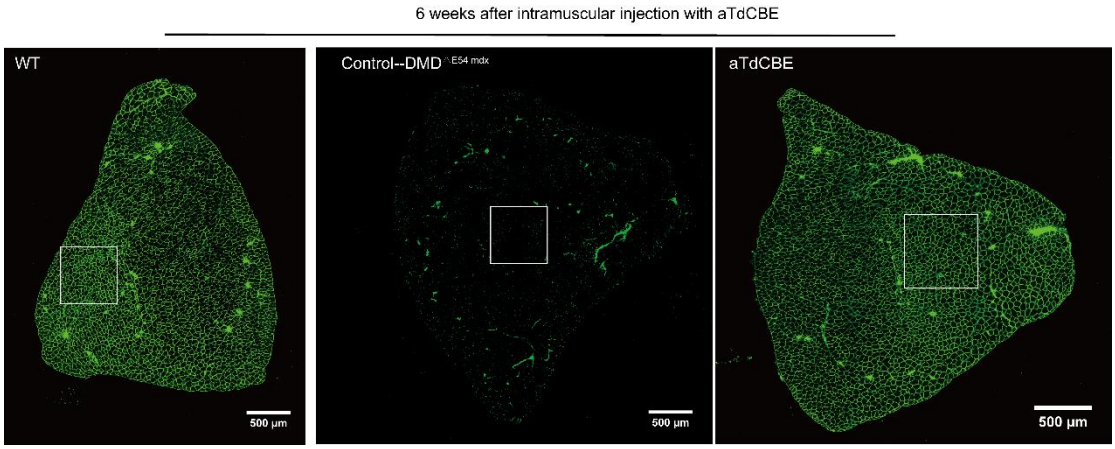
SpG Cas9 were used to targeting the SAS-containing sequence with 3'-TGT PAM. Data are presented as means ± S.D. Values represent n = 3 independent biological replicates. All of the above base editors contain UGI. Source data are provided as a Source Data file.



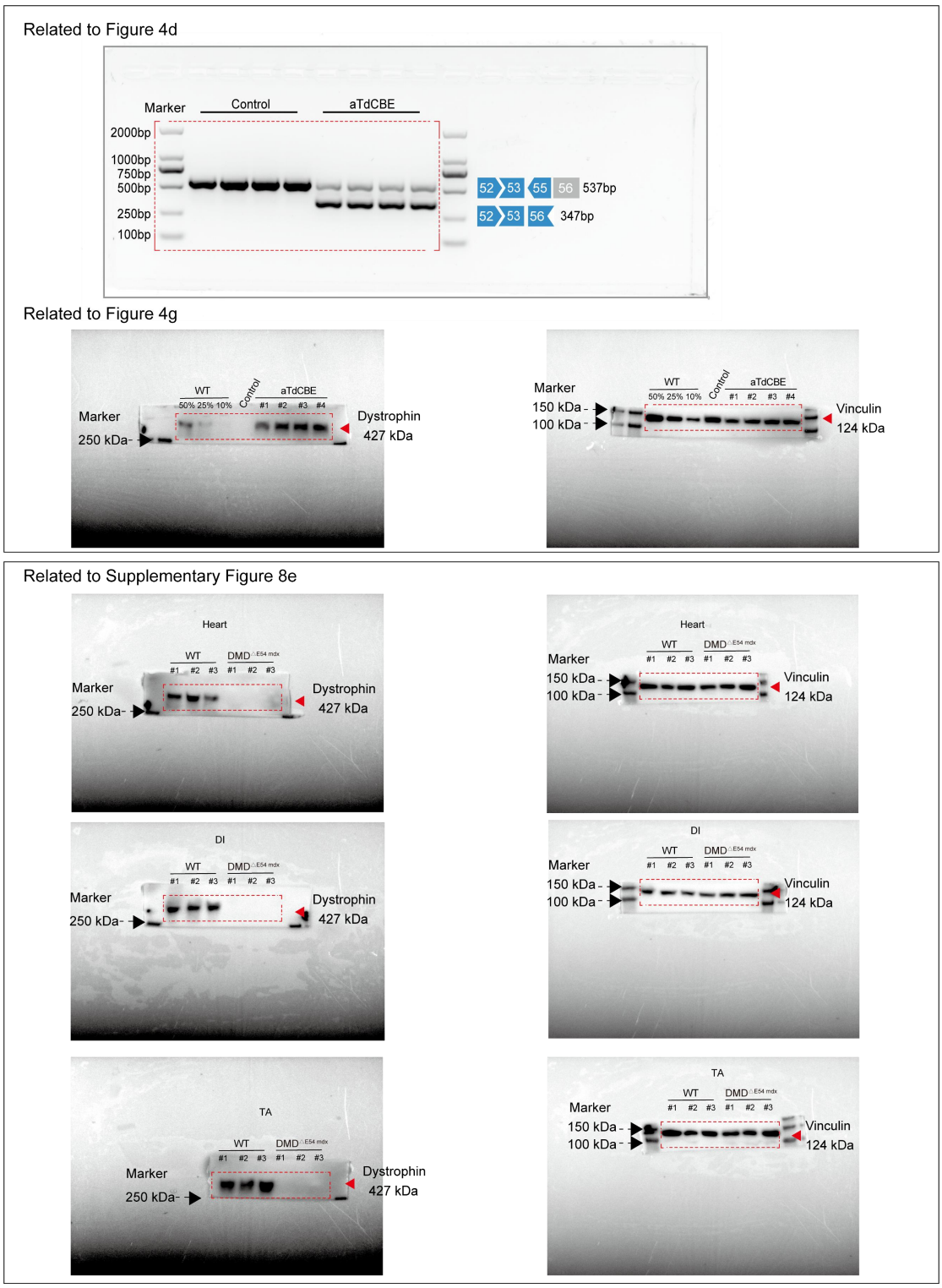
Supplementary Fig.8 Establishment and characterization of a humanized DMD mouse model.

a. Strategy for generating a humanized DMD mouse model. CRISPR-Cas9 editing was employed to delete human *DMD* exon54. **b.** RT-PCR analysis of TA muscles to validate deletion of human exon 54. **c.** Dystrophin immunohistochemistry from indicated muscles of WT and DMD$^{\Delta E54\ mdx}$ mice. WT mice were derived from crosses between STOCK Tg (DMD) 72Thoen/J mice (#018900) and mdx mice, which carry a c.2977C>T, p.Gln993* mutation in exon 23 on Chr.X. Dystrophin (Abcam, ab15277) and spectrin (Millipore, MAB1622) are shown in green and mangenta, respectively. **d.** Western blot confirming the absence of dystrophin (Sigma, D8168) in indicated muscle tissues. **f.** Serum CK, a marker of muscle damage and membrane leakage, was measured in WT and DMD$^{\Delta E54\ mdx}$ mice. **e.** Sirius red staining and HE staining of TA, DI, and heart muscle of WT and DMD$^{\Delta E54\ mdx}$ mice. **g.** WT and DMD$^{\Delta E54\ mdx}$ mice were subjected to forelimb grip strength testing to measure muscle performance. All mice were 4 weeks old at the time of the experiment. Data are represented as mean ± s.e.m (n=6 independent biological replicates).

Each dot represents an individual mouse. *P*-value was calculated using an unpaired two-tailed Student's t-test. Scale bar, 100 μm. Source data are provided as a Source Data file.

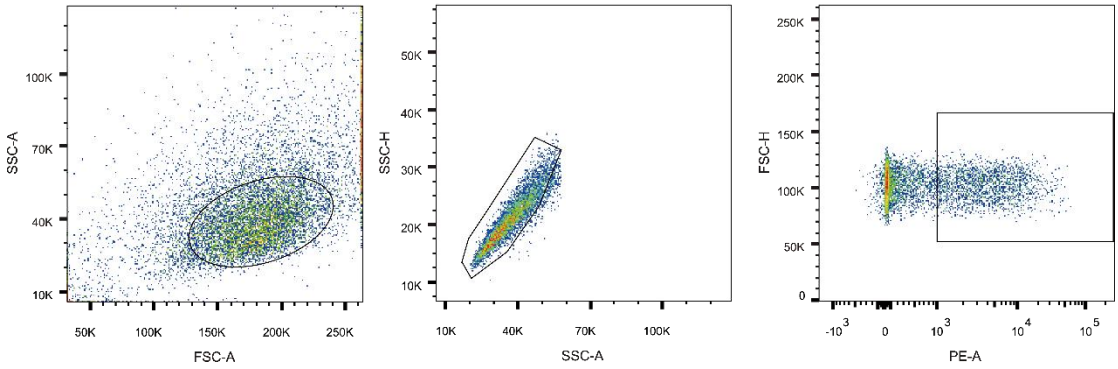


Supplementary Fig.9 Rescue of dystrophin expression following intramuscular injection of aTdCBE after 6 weeks.
Dystrophin immunohistochemistry of TA muscle. Control mice were injected with saline. Images shown in both Fig. 4e and Supplementary Fig.8 were obtained from the same tissue at 20x magnification. Fig. 4e showed the local region staining image rather than the reconstituted whole-tissue scanning image in Supplementary Fig.8, and highlighted with white boxes. Dystrophin is shown in green. Scale bar, 500 μm.



Supplementary Fig.10 Uncropped images.

The red rectangles indicate the cropping location.



Supplementary Fig.11 Flow cytometry gating strategy.

Cell singletons were first gated out via forward scatter (FSC) and side scatter (SSC) parameters. Fluorescent cells were then gated for gene editing analysis.