nature biotechnology

Article

https://doi.org/10.1038/s41587-023-01821-9

Adenine transversion editors enable precise, efficient A•T-to-C•G base editing in mammalian cells and embryos

Received: 12 January 2023

Accepted: 8 May 2023

Published online: 15 June 2023

Check for updates

Liang Chen[1,8], Mengjia Hong[1,8], Changming Luan[1,8], Hongyi Gao[1], Gaomeng Ru[1], Xinyuan Guo[1], Dujuan Zhang[1], Shun Zhang[1], Changwei Li[2], Jun Wu[1], Peyton B. Randolph[3,4,5], Alexander A. Sousa[3,4,5], Chao Qu[1], Yifan Zhu[1], Yuting Guan[1], Liren Wang[1], Mingyao Liu[1,6], Bo Feng[7], Gaojie Song[1], David R. Liu[3,4,5] & Dali Li[1]

Base editors have substantial promise in basic research and as therapeutic agents for the correction of pathogenic mutations. The development of adenine transversion editors has posed a particular challenge. Here we report a class of base editors that enable efficient adenine transversion, including precise A•T-to-C•G editing. We found that a fusion of mouse alkyladenine DNA glycosylase (mAAG) with nickase Cas9 and deaminase TadA-8e catalyzed adenosine transversion in specific sequence contexts. Laboratory evolution of mAAG significantly increased A-to-C/T conversion efficiency up to 73% and expanded the targeting scope. Further engineering yielded adenine-to-cytosine base editors (ACBEs), including a high-accuracy ACBE-Q variant, that precisely install A-to-C transversions with minimal Cas9-independent off-targeting effects. ACBEs mediated high-efficiency installation or correction of five pathogenic mutations in mouse embryos and human cell lines. Founder mice showed 44–56% average A-to-C edits and allelic frequencies of up to 100%. Adenosine transversion editors substantially expand the capabilities and possible applications of base editing technology.

Human genetic disorders are mainly caused by genetic variants, of which approximately half are pathogenic single-nucleotide variations (SNVs)[1]. Precise DNA sequence conversions can be achieved through gene editing technologies, such as programmable nuclease-induced homology-directed repair (HDR), base editing and prime editing[2]. HDR is a double-strand DNA break (DSB)-dependent process that is inefficient in most therapeutically relevant cell types and requires DNA donor templates. Moreover, DSBs are associated with undesired side effects, such as P53 activation, large DNA fragment deletions and chromosomal abnormalities, such as translocations[3]. Prime editors are highly versatile and precise tools for genome editing, but their application can be limited by the need to test a variety of designs to achieve

[1]Shanghai Frontiers Science Center of Genome Editing and Cell Therapy, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China. [2]Department of Orthopedics, Shanghai Key Laboratory for Prevention and Treatment of Bone and Joint Diseases with Integrated Chinese-Western Medicine, Shanghai Institute of Traumatology and Orthopedics, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. [3]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [5]Howard Hughes Medical Institute, Cambridge, MA, USA. [6]BRL Medicine, Inc., Shanghai, China. [7]School of Biomedical Sciences, Faculty of Medicine, Chinese University of Hong Kong, Hong Kong, China. [8]These authors contributed equally: Liang Chen, Mengjia Hong, Changming Luan. e-mail: dlli@bio.ecnu.edu.cn

efficient editing, especially in primary cells and in vivo[4–6], although recently described engineered prime editing guide RNA (epegRNA)[7] and PE4/5max prime editing systems[8] have greatly improved prime editing efficiencies in many living systems.

For some single-nucleotide conversions, base editing remains the most efficient technology to install base substitutions without inducing DSBs or requiring donor templates. Cytosine base editors (CBEs) efficiently generate C•G-to-T•A conversions and are composed of a Cas9 nickase, cytosine deaminase and uracil glycosylase inhibitor (UGI) to inhibit cytosine transversion side products, including C•G-to-G•C and C•G-to-A•T, which are induced by the uracil DNA N-glycosylase (UNG)-mediated base excision repair (BER) pathway[9,10]. By substituting the CBE UGI with UNG or DNA repair factors, several C•G-to-G•C base editors (CGBEs) that increase the frequency of cytosine transversion editing were developed that mainly induce C-to-G conversions in mammalian cells but C-to-A transversion in *Escherichia coli*[9–13]. Adenine base editors (ABEs), composed of Cas9 nickase and laboratory-evolved TadA deaminases, create A•T-to-G•C conversions with few byproducts. The product purity of ABEs can be very high (≥99% from ABE7.10 at some targets, for example), likely due to a lack of efficient or well-expressed endogenous DNA glycosylases that initiate BER at inosines to induce insertions and deletions (indels) and adenine transversions[14].

Adenine transversions could play an important role in therapeutic gene correction and other applications for genome engineering. The ability to make targeted A•T-to-C•G and A•T-to-T•A substitutions could potentially correct 17% and 8% of known pathogenic SNVs, respectively (Fig. 1a). To date, however, no adenine transversion base editors have been described[14,15]. In addition to corrective gene editing, programmable adenine transversions would increase the diversity of outcomes for mutagenesis-based applications, such as lineage tracing, genetic screening and molecular evolution. Here we describe the development of adenine transversion base editors consisting of a Cas9 nickase, the highly active evolved deoxyadenine deaminase TadA-8e and mouse alkyladenine/3-methyladenine DNA glycosylase (mAAG) variants. Through the laboratory evolution of mAAG and embedding of the mAAG and TadA-8e in Cas9 nickase, A•T-to-C•G base editors (ACBEs) were developed that generate efficient and precise A•T-to-C•G conversion in mammalian cells. ACBEs can install or correct pathogenic SNVs in mouse embryos or in human cells. The development of adenine base transversion editors substantially expands the capabilities and application scope of base editing.

Results

Screening enzymes for inosine excision

Although nucleotide transitions can be achieved through base deamination, transversion is more complicated and typically requires target nucleotide replacement. The BER pathway can achieve this replacement by creation of an apurinic–apyrimidinic (AP) site and subsequent excision and repair of the lesion in DNA in a mutagenic manner[16,17]. We hypothesized that it was possible to achieve target adenine transversion through (1) programable adenine deamination to generate inosine; (2) removal of hypoxanthine by a DNA glycosylase to create an AP site or direct excision of inosine by an endonuclease; (3) repair of the DNA lesion to result in pyrimidine substitution; and (4) installation of a nick on the unedited DNA strand to stimulate DNA repair to retain the mutated nucleotide. Evolved TadA and Cas9 nickase are capable of mediating the first and the fourth step, but it is difficult to process a targeted inosine excision because cellular repair of inosine intermediates created by ABEs is inefficient[14]. We hypothesized that fusion of an active enzyme that catalyzes inosine excision with an ABE would make the resulting protein much more capable of processing inosines to mediate adenine transversion (Fig. 1b).

Because in most prokaryotic and eukaryotic species, hypoxanthine in DNA is removed either by the alkyladenine DNA glycosylases (AAGs) or by endonuclease V through alternative excision repair pathways[18], we chose nine enzymes for investigation: endonuclease V from *E. coli*[19] and eight AAGs from different species, including ANPG[20] and its truncated human form[21], APDG from rat[21], AAG from mouse[22], MAG from yeast[21], UDG6 from *Methanosarcina barkeri*[23], AAG from *Bacillus subtilis*[24] and AlkA from *E. coli*[25]. These enzymes were individually fused to the C-terminus of Cas9 nickase, resulting in nine constructs named AH1–9 (Supplementary Fig. 1). After high-throughput sequencing (HTS) analysis at an endogenous site in human cells, we were excited to find that four constructs supported up to 8.7% A-to-Y (Y = C or T) transversion events with an AH4 > AH3 > AH8 » AH1 hierarchy, and low indel events were observed in all groups except AH5 candidate (Fig. 1c and Supplementary Fig. 2a). Although A-to-G transitions were still the major product, it was encouraging to observe programmable adenine transversions independent of donor templates in human cells. As AH4 exhibited the highest efficiency, mAAG was selected for further evaluation. By testing fusions of the mAAG in varied orientations within the construct, we found that AH4 with mAAG on the C-terminus still showed the highest adenine transversion activity at the three tested targets, and we named it AXBE (X = any nucleotide) (Fig. 1d and Supplementary Fig. 2b).

Characterization of AXBE in human cells

To further characterize the performance of AXBE in human cells, 26 endogenous sites were investigated. AXBE induced adenine transversions (averaging 14%, with up to 38% at the *RUNX1*-sg3 site) within a relatively wide editing window from positions 2 to 12 (with 7–9 as the major editing window, counting the PAM sequence as positions 21–23) (Fig. 2a). Notably, the proportion of A-to-C edits (averaging 9.7%, with up to 29%) in the products was 2.5-fold higher on average than that of A-to-T edits (averaging 4.1%, with up to 9.5%), which was consistent at individual targets probably due to the intrinsic mechanisms of DNA repair after inosine excision (Fig. 2a and Supplementary Fig. 3). We also noticed that AXBE selectively generated A-to-Y edits in specific positions—for example, transversions occurred at A_5 and A_9 but not A_7 at *PPP1R12C* site 15, whereas A-to-G edits were observed at all three positions (Fig. 1d). These data suggest that mAAG might catalyze hypoxanthine excision in specific sequence contexts.

To characterize sequence context preferences of AXBE, the efficiency of 93 adenines associated with their sequence context within these 26 target loci was further analyzed. We found that adenines prone to transversion (>10% on average) were located within a YA × R (R = A/G) motif with a CAG > CAA > TAA > TAG hierarchy, and some other motifs (CAT, GAA, GAG and AAG) also demonstrated notable editing efficiency (Fig. 2b,c). These sequence context dependences were further confirmed through in vitro enzyme activity assays, which demonstrated that recombinant mAAG preferentially removed hypoxanthine from inosine in YIR as well as RIR motifs (Fig. 2d and Supplementary Fig. 4). These findings suggest that AXBE selectively generates adenine transversions in specific motifs in a manner dependent on mAAG activity, which is known to be influenced by DNA sequence context[22].

Further investigation of nine additional target sites containing at least one YAR motif demonstrated that AXBE induced efficient (15–32%) A-to-Y conversion within the major window (A_7–A_9), further confirming that the sequence context was critical for adenine transversions (Fig. 2e). After analysis of all 28 tested YAR motif targets, we found that AXBE induced transversion at A_2–A_{10} with the average efficiency up to 27% on A_8, including 19% of A-to-C edits and 7.7% of A•T-to-T•A events (Fig. 2f and Supplementary Fig. 5a,b). Similarly to CGBEs[11], indels were also observed in AXBE-edited products (Supplementary Fig. 5c). Additionally, AXBE also induced robust adenine transversions (up to 46%), including A-to-C editing (15–33%) and A-to-T editing (3.8–14%) in HeLa cells, which was even higher than in HEK293T cells, indicating that AXBE-induced A-to-Y editing was applicable to different mammalian cell lines (Supplementary Fig. 6a,b).

We found that AXBE induced lower Cas9-dependent and Cas9-independent off-target edits than ABE8e through HTS analysis of confirmed or predicted 40 off-target sites[26,27] and orthogonal R-loop

https://doi.org/10.1038/s41587-023-01821-9

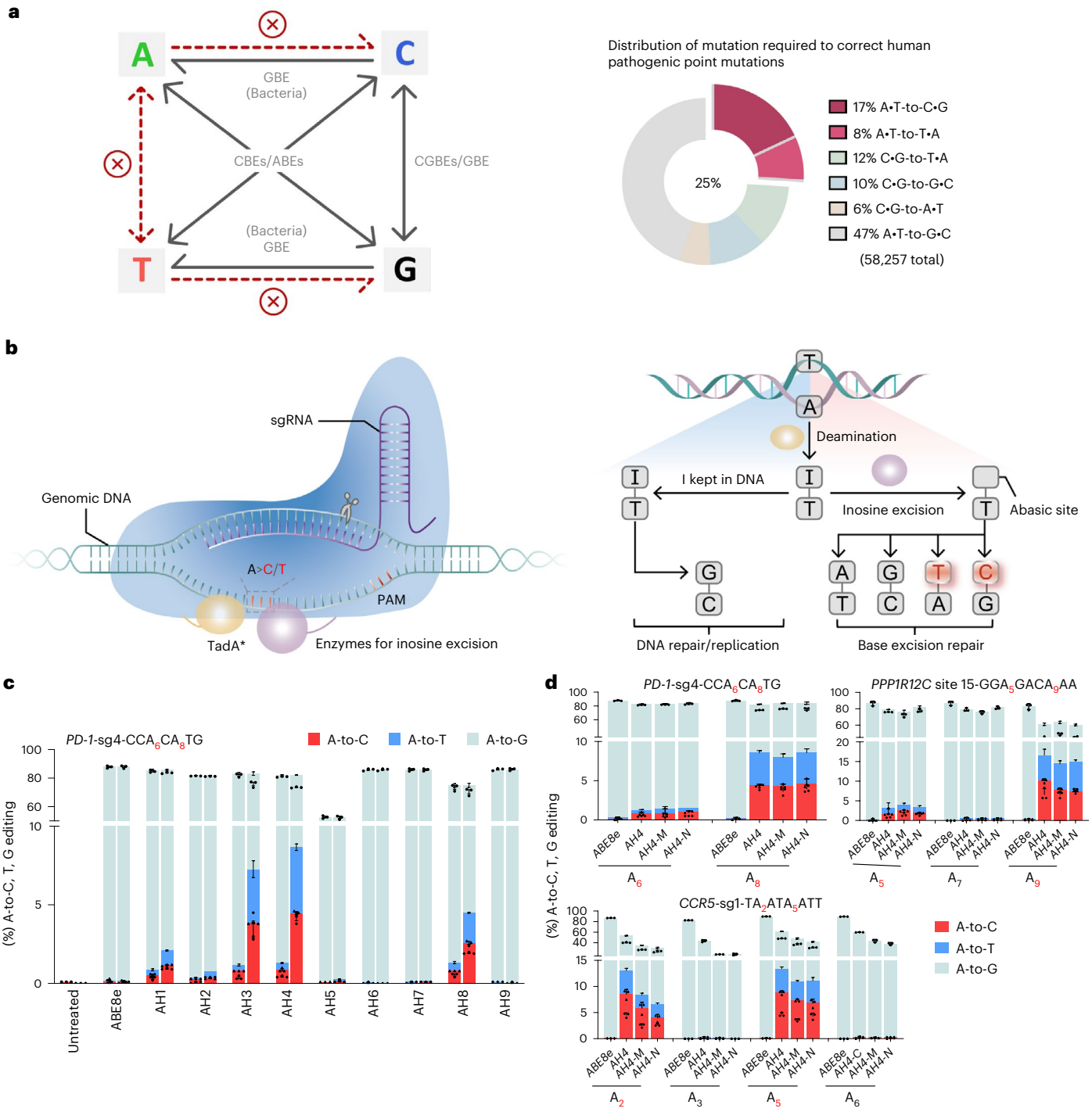
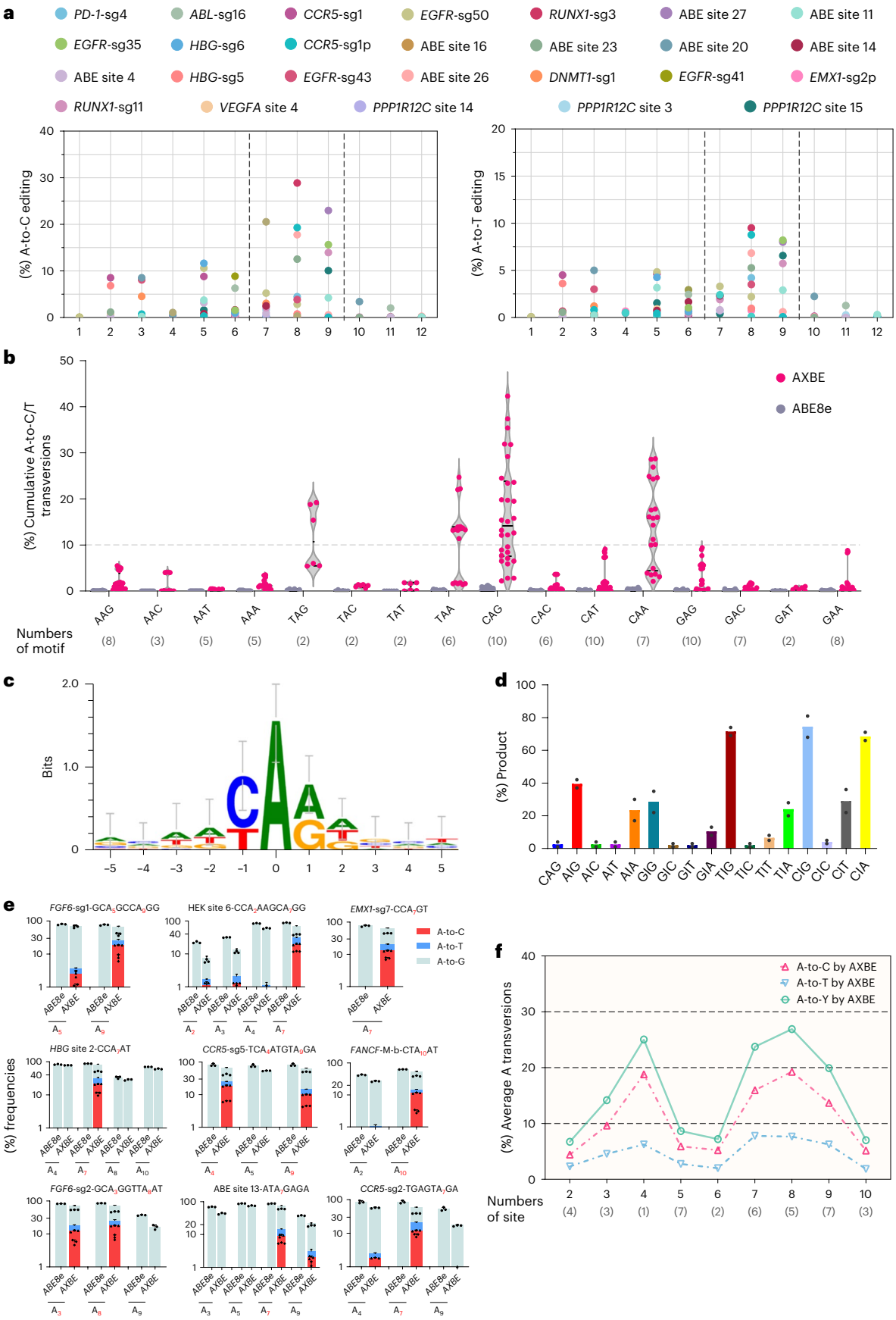


Fig. 1 | 3-methyladenine DNA glycosylase-derived base editors mediate programmable adenine transversions. a, Overview of site-specific base transitions and transversions induced by reported base editors (unachievable adenine transversions marked with 'x') (left). Distribution of human pathogenic SNPs needed to reverse diseases in the ClinVar database (accessed 24 May 2021) is shown (right). **b**, Schematic illustration of the conceptual design to induce programmed A-to-C/T (left) and a potential molecular mechanism of intracellular adenine conversions (right). **c**, Screening of nine candidate fusion constructs, each composed of nickase Cas9 (nCas9), adenine deaminase (TadA-8e) and a C-terminal enzyme that could potentially act on inosine, at an endogenous target site (*PD-1*-sg4) in HEK293T cells. The enzymes to be screened in each AH series were: AH1: ANPG from human; AH2: truncated ANPG from human; AH3: APDG from rat; AH4: AAG (3-methyladenine DNA glycosylase) from mouse; AH5: endonuclease V from *E. coli*; AH6: AlkA from *E. coli*; AH7: UDG family 6 from *M. barkeri*; AH8: AAG from *B. subtilis*; and AH9: MAG from yeast. **d**, Effects of mAAG orientation of the AH4 construct on adenine transversion at three endogenous target sites. In **c** and **d**, single triangle, diamond and dot represent A-to-C, A-to-T and A-to-G editing of an individual replicate, respectively. Data represent the mean ± s.d. of $n = 3$ independent replicates.

assays[28] (lower efficiency at four of six sites), respectively (Fig. 3a,b). Notably, AXBE displayed a 90% reduction of RNA A-to-I off-target effects, suggesting that mAAG in AXBE decreased the RNA editing activity of TadA-8e because the protein level of AXBE and ABE8e was similar (Fig. 3c and Supplementary Fig. 6c). Because alkylated bases and hypoxanthine in DNA chain are the only reported substrates, and

https://doi.org/10.1038/s41587-023-01821-9



no publications suggest that AAGs can bind or catalyze RNA substrates, we speculated that fusion of mAAG to ABE8e disrupted the binding affinity of deaminase to RNA substrates; however, the detailed mechanisms still need further investigations. These data demonstrated that AXBE displayed much less off-target effects than ABE8e at both the DNA and RNA level.

Fig. 2 | Characterization of AXBE in vivo and in vitro. a, Dot plots showing the distribution of A-to-C and A-to-T editing of AXBE across the protospacers (PAM is at positions 21–23) from the 26 endogenous genomic loci in HEK293T cells 3 d after transfection of base editor plasmids. The dashed lines show the major activity window by AXBE. Each dot represents the mean of three independent biological replicates. **b**, Cumulative adenine transversions of AXBE at 93 NAN motifs from 26 target sites in **a** with ABE8e as control. Each motif and corresponding numbers are indicated on the bottom. Each dot represents a biological replicate ($n = 3$). **c**, The sequence context of adenine transversions (>10%) identified by HTS. Sequence conservation at positions from −5 to +5 is shown, with the mutated A at position 0. **d**, The fraction of inosine excision mediated by mAAG in vitro. Inosine was embedded in 16 different motifs (NIN) and incubated with mAAG for 15 min to cleave the 5′ 6-mer fragment of each oligonucleotide (additional details in Methods). Data represent the mean of two independent biological experiments. **e**, Bar plots showing the on-target DNA base editing frequencies of AXBE and ABE8e at nine additional endogenous target sites bearing a YAR (Y = C/T, R = A/G) motif. Single triangle, diamond and dot represent A-to-C, A-to-T and A-to-G editing of an individual replicate, respectively. Data represent the mean ± s.d. of $n = 3$ independent replicates. **f**, Average A-to-C, A-to-T and A-to-Y editing efficiencies of AXBE across the protospacers from 28 endogenous genomic loci bearing a YAR motif in **a** and **e**. Corresponding numbers of target sites are listed on the bottom. Data represent the mean of three biologically independent experiments.

AXBE induced conversion of adenosine to the other three nucleosides, which could generate more codon variations at a given target site than canonical ABEs. A-to-G conversion of the 64 triplet codons could theoretically generate 122 codon and 32 amino acid variants, whereas AXBE could induce 558 codon and 147 amino acid variants (Fig. 3d,e and Supplementary Table 4). This suggests that AXBE substantially increases the number of variants achievable with a given triplet code compared to ABEs, indicating that it could be used as a mutator for molecular evolution, functional genetic screening or generation of genetic barcodes for lineage tracing.

Structure-guided engineering of mAAG

Next, we sought to use structure-guided laboratory evolution of mAAG to improve the efficiency, product purity and editing scope of AXBE. The structure of mAAG predicted by AlphaFold is almost identical to the experimentally determined structure of human AAG (hAAG, with Cα root mean square deviation (RMSD) of 0.2 Å), and they share high sequence homology to each other especially within the substrate binding pocket[29]. We, thus, harnessed the structure of human AAG in complex with double-strand DNA (dsDNA) containing an alkylated adenine (1,N^6-ethenoadenine) for reference (Fig. 4a). For the first round of mutation screening, 15 residues that might participate in substrate recognition or catalytic reaction were selected. For example, E125 (mAAG E145) is predicted to deprotonate the bound water for nucleophilic attack on the substrate glycosylic bond, and a tyrosine (Y159 in hAAG or Y179 in mAAG) is predicted to form a hydrogen bond with the backbone of dsDNA as well as a key π−π stacking interaction with the flipped base within the active site. Moreover, due to π−electron stacking, AAG likely recognizes diverse substrate nucleosides through aromatic side chains in the binding pockets[30], and varying the size and hydrophobicity of the side chains could alter engagement of adenine and inosine by mAAG variants. In total, 38 variants were individually mutated to modulate the performance of mAAG (Supplementary Fig. 7a).

After evaluation of 38 AXBE variants with individual mAAG point mutations at the *FGF6*-sg2 target, nine variants with individual mutations at five amino acids (R165, Y179, G183, M184 and Y185) displayed reduced A-to-G edits and increased A-to-Y efficiency compared to AXBE (Supplementary Fig. 7a,b). To generate more active variants, a second round of screening was performed to increase the diversity of substitutions on the above five residues and other additional adjacent residues. After evaluating 25 variants for editing at the *EMX1* site, the AXBE-R165E mutation exhibited the highest editing efficiency and a 4.9-fold Y:G outcome ratio compared to AXBE (Fig. 4b,c). To further increase AXBE activity, combinational mAAG pair mutations of R165E, Y179F, G183Q, G183H, M184F, M184W and Y185F were generated and analyzed. Among the tested 17 combination mutant AXBE variants, AXBE-R165E•Y179F (AXBE-EF) displayed the highest A-to-Y transversion efficiency (52.8%), and the A-to-Y versus A-to-G efficiency ratio was 3.8, which was 8.7-fold higher than that of AXBE. A-to-C editing was the major outcome, ~2.8-fold more frequent than either A-to-G or A-to-T outcomes (Fig. 4b,c). Similar results were also observed at an additional target site (*CCR5*-sg1) (Supplementary Fig. 7c). This suggested that AXBE containing an mAAG-EF mutation exhibited a much higher adenine transversion efficiency. This variant was named AXBEv2.

Expanded targeting scope of AXBEv2 variants

We speculated that increased activity might also lead to an expanded targeting scope of AXBEv2. In vitro enzyme activity assay suggested that mAAG-EF recombinant protein displayed a marked enhancement of catalytic velocity with an average cleavage rate of 0.367 min^{-1}, whereas that of wild-type mAAG was only 0.056 min^{-1} among tested motifs (Supplementary Fig. 7d). For example, the activity of mAAG-EF showed an 11-fold increase on CIG substrates and significant improvement at other motifs, such as CIC, GIG and GIA, that were very poorly processed by wild-type mAAG (Fig. 4d and Supplementary Fig. 7d,e). The R165E + Y179F mutations in mAAG thus markedly enhanced the catalytic activity and sequence context compatibility of the glycosylase.

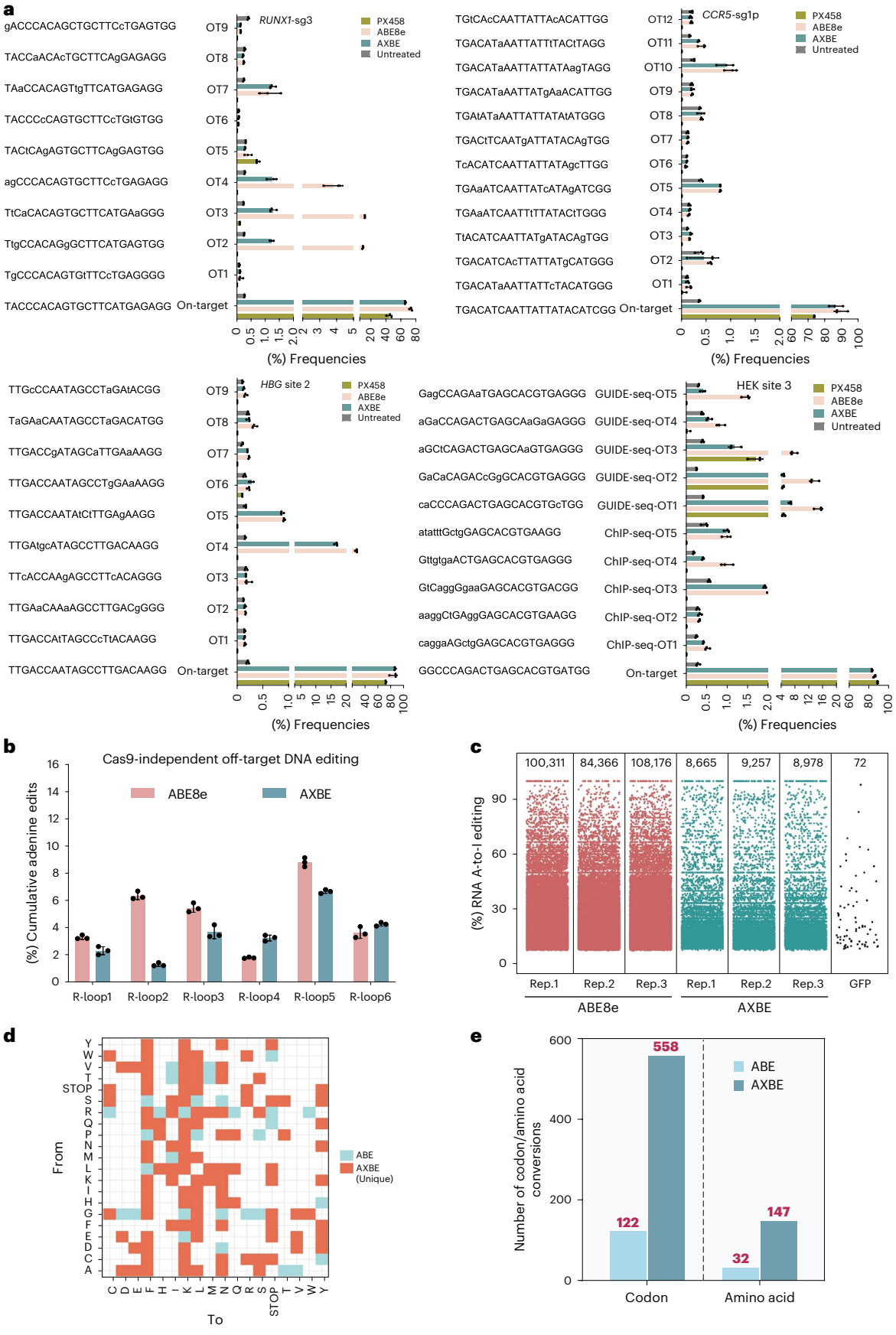
To investigate whether AXBEv2 had expanded targeting scope in cultured cells, 16 target sites each containing one of distinct 16 NAN motifs in the major editing window were tested. The average A-to-Y conversion efficiency of AXBEv2 was 3.2-fold (up to 5.8-fold) comparing to AXBE at all these 16 motifs, whereas the A-to-C efficiency was increased an average of 3.7-fold (up to 9.8-fold) (Fig. 4e and Supplementary Fig. 7f). Notably, AXBEv2 was also compatible with PAM-relaxed SpCas9 variants at all tested targets with variant PAM sequences, such as SpCas9-NG[31] and near-PAMless SpRY[32], supporting A-to-Y conversion efficiencies of up to 64% (highest C outcomes with frequencies of up to 43%) and 52% (up to 36% A-to-C edits), respectively (Fig. 4f,g). These data suggest that AXBEv2 substantially increased A-to-Y transversion activity, sequence context tolerance and compatibility with PAM-relaxed Cas9 variants.

Fig. 3 | Off-target assessment of AXBE. a, Cas9-dependent DNA on-target and off-target analysis of cumulative adenine editing frequencies at the indicated targets induced by ABE8e and AXBE using PX458 as control. Mismatched nucleotides in off-target sequences are indicated in lowercase letters. Data represent the mean ± s.d. of $n = 3$ independent replicates. **b**, Analysis of Cas9-independent off-target adenine editing at the *FGF6*-sg2 target site induced by ABE8e and AXBE using orthogonal R-loop assays (additional details in Methods). Data represent the mean ± s.d. of $n = 3$ independent replicates. **c**, Off-target RNA editing activities induced by ABE8e and AXBE. Jitter plots from RNA sequencing experiments showing efficiency of A-to-I conversion (*y* axis) 3 d after transfection of ABE8e, AXBE base editor and GFP control plasmids, respectively (additional details in Methods). Total number of modified bases is listed at the top. Each biological replicate is listed on the bottom. **d**, Amino acid changes can be induced by ABE (cyan) or AXBE (orange + cyan) with unique conversions in orange. **e**, The total number of codon and amino acid changes induced by ABE or AXBE. ChIP-seq, chromatin immunoprecipitation followed by sequencing; GUIDE-seq, genome-wide unbiased identification of double-stranded breaks enabled by sequencing.

https://doi.org/10.1038/s41587-023-01821-9

Engineering of A-to-C base editors

AXBEv2 mediated highly efficient A-to-Y editing with cytosines as dominant products for most target sites, but it also induced A-to-G edits. We sought to develop ACBEs with a narrower editing window to minimize unwanted edits. Because embedding the deaminase of base editors into the Cas9 domain can alter the editing window and increase



activity[33], we first generated several constructs with TadA-8e and/or mAAG-EF inserted into the CTD domain of Cas9, because insertion of deaminase into this position showed higher adenine conversion with narrow editing window (Supplementary Fig. 8). Almost all constructs displayed elevated A-to-C efficiency compared to AXBEv2, among which four constructs (AXBEv2-Ce01/Ce02/Ce07/Ce08) showed ~2-fold and ~5-fold increase in the A_5-to-C frequency and a substantial decrease in A-to-G editing efficiency at A_2 and A_3 compared to AXBEv2 and AXBE, respectively (Fig. 5a). As AXBEv2-Ce07 displayed high A-to-C efficiency and reduced A-to-G activity, we designated the prototype ACBE. To further reduce bystander editing, we used our recently evolved TadA-8e-N108Q deaminase variant, which shows similar activity but offers a narrowed editing window[34]. After introducing N108Q into the above four constructs, all bystander edits were substantially reduced, and AXBEv2-Ce07Q (ACBE-Q) displayed efficient A-to-C editing but induced only minimal A-to-G editing at A_3 and A_{12} (Fig. 5a and Supplementary Fig. 9a,b).

We compared AXBEs and ACBEs at 13 endogenous targets. ACBE exhibited a much higher A-to-C conversion efficiency, averaging 35% (up to 45%) with transversion efficiency up to 73% (Fig. 5b and Supplementary Fig. 9c). Editing activity was enhanced on average four-fold and 1.4-fold compared to AXBE and AXBEv2, respectively, with a similar level of indel events (~14%) (Supplementary Fig. 9d,e). In addition to the activity increase, ACBE also showed an A-to-C product purity averaging 48% (up to 69%), which was much higher than AXBE and AXBEv2, averaging 15% and 39% product purity, respectively (Fig. 5b). We also observed that, at some non-YAR motif targets, ACBE showed higher activity compared to AXBE or AXBEv2, such as the *RUNX1*-sg1 (GAG motif) and *EGFR*-sg22 (GAA motif) (Supplementary Fig. 9c). Similarly, ACBE-Q displayed an increased A-to-C efficiency, averaging 27% (up to 45%) (Fig. 5b,c and Supplementary Fig. 9c) and broader editing sequence context (Supplementary Fig. 9f). Compared to ACBE, ACBE-Q had a very steep and narrow A_4–A_6 editing window leading to greatly reduced bystander A-to-G byproducts (Supplementary Fig. 9c,g,h), suggesting that ACBE-Q was more accurate than ACBE. To compare the relative precision ratio of these two ACBEs, we calculated the ratio of desired single A-to-C reads divided by all the reads of editing outcomes, including indels of the 12 target sites. ACBE-Q induced much higher ratios of precisely edited alleles compared to ACBE (averaging 43-fold, ranging from 1.9-fold to 171-fold) (Fig. 5d). Similar results were also obtained in HeLa cells, suggesting that the DNA repair processes responsible for A-to-C editing outcomes were similar, although the minor differences in efficiency and purity were observed at distinct target sites of these two cell lines (Fig. 5e,f and Supplementary Fig. 10). Analysis of Cas9-independent off-target events via orthogonal R-loop assays revealed that ACBE (1.2–2%) induced fewer off-targeting edits compared to ABE8e (2.1–7.2%) at five of six sites. Notably, the off-target editing of ACBE-Q was lowered to near-background levels (mean <0.3%) among all evaluated targets, demonstrating that ACBEs, especially ACBE-Q, enable accurate editing with minimal Cas9-independent DNA off-targets (Fig. 5g).

Another precision gene editing technology capable of installing or correcting transversion mutations is prime editing[35]. Achieving high levels of prime editing efficiency often requires optimization of several components, including the lengths and sequences of primer binding sites and reverse transcriptase templates, the position of pegRNA and nicking sgRNA protospacers and the combination of silent and non-silent edits to be made[8,13]. Because prime editing is the state-of-the-art technology to install targeted A-to-C conversions, we sought to directly compare ACBEs with a minimally optimized application of the PE4/5max system with epegRNAs[7,8]. Thirty-six combinations of prime editing parameters were evaluated at three target sites for precise A-to-C editing, and base conversions up to 33% (*ABL*) and 22% (*EGFR)* were observed, although limited prime editing was detected at the third target (*FBN1*) (Supplementary Fig. 11a). In comparison to the above prime editing strategies, ACBEs demonstrated significantly higher efficiency in A-to-C editing (≥40% at all the three sites), and ACBE-Q displayed enhanced A_6-to-C editing purity (21–40%) (Supplementary Fig. 11). These results suggest that ACBEs are better suited for installing A-to-C edits with minimal optimization in an efficient manner at these tested target sites. As expected, given the unique mechanism of prime editing, the PE4/5max prime editing strategies yield higher product purities than ACBEs (Supplementary Fig. 11b), consistent with other comparisons between base editors and prime editors[3,11,13]. ACBEs offer a more convenient and efficient option for certain targets, whereas prime editing offers a higher product purity. Thus, they offer complementary strengths.

During peer review of this manuscript, a study reported the independent development of an adenine transversion base editor, AYBE, by fusion of ABE8e with engineered hAAG (also known as hypoxanthine excision protein N-methylpurine DNA glycosylase (MPG))[36]. We directly compared the performance of AYBEv3 with our AXBEv2 and ACBE-Q at six endogenous sites. HTS data revealed that AYBEv3 and AXBEv2 showed a similar adenine transversion activity in specific sequence contexts as well as simultaneous A-to-G bystander mutations (Supplementary Fig. 12a). ACBE-Q also showed similar A-to-C activity as AYBEv3 and AXBEv2, but it greatly reduced A-to-G bystander edits (Supplementary Fig. 12). These data suggest that variant AAGs can be applied to adenine base transversion and that ACBE-Q is advantageous for efficiently converting A-to-C with minimal bystander A-to-G mutations.

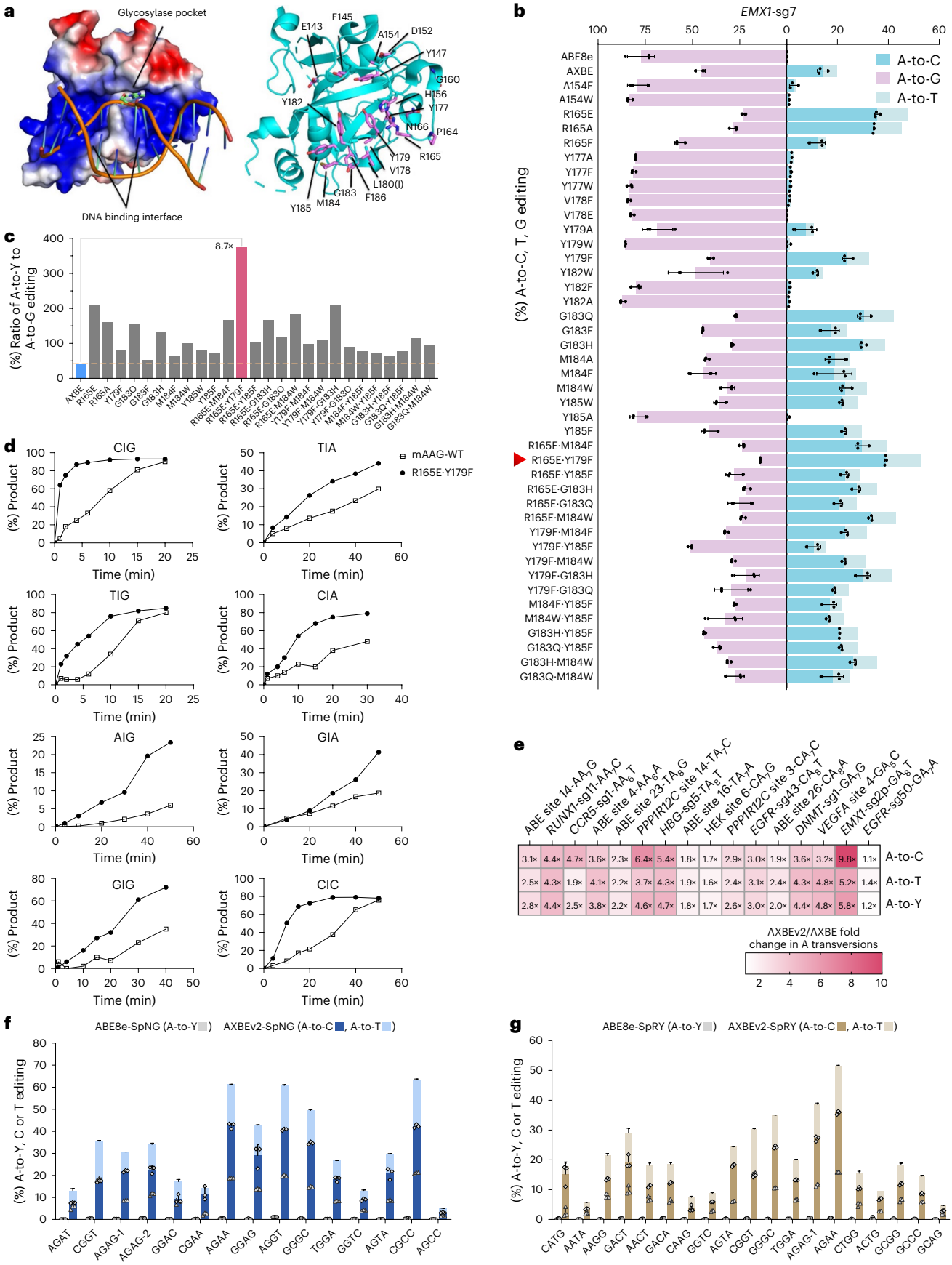
Generation of disease-relevant mouse models

To further evaluate ACBE activity in vivo, we used ACBE-Q to install precise base conversions in mouse embryos to model human diseases. Through targeting the splicing acceptor site in intron 26 of the dystrophin gene (Fig. 6a), 94% (30/32) of F0 pups contained mutations at this site, and 70% (21/30) of the mutants harbored A-to-C conversion with the desired A_6-to-C editing efficiency (the rate of reads containing on-target A-to-C edits among total reads), averaging 56% (up to 99.5%) and purity (the frequency of A-to-C yields divided by adenine mutation efficiencies) up to 99.8%. Among these 21 mutant pups, only seven of them had edits on the A_8 position, demonstrating that ACBE-Q also had a narrow editing window in embryos (Fig. 6b–d and

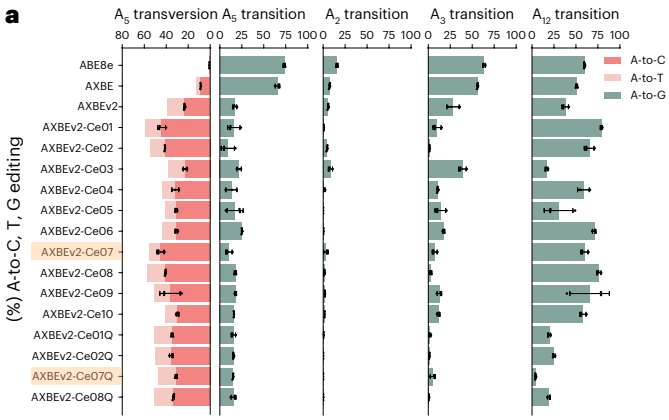
Fig. 4 | Evaluation of evolved mAAG in vivo and vitro. a, Crystal structure of hAAG in complex with DNA containing a 1,N^6-ethenoadenine lesion (left) and the substrate binding site of hAAG (right) (Protein Data Bank: 1F4R). For clarity, all residues are numbered according to their aligned counterpart in mAAG, which is essentially identical to hAAG for the labeled residues except at L180 (I160 in hAAG). **b**, Comparison of adenine conversion efficiency of AXBE and AXBE variants at the *EMX1*-sg7 site in HEK293T cells using ABE8e as control 3 d after transfection of conducts. Data represent the mean ± s.d. of $n = 3$ independent replicates. **c**, The ratio of A-to-Y to A-to-G editing efficiency induced by AXBE and 26 AXBE variants at the same target in **b**. An optimal R165E•Y179F variant in pink was chosen for further evaluation, and original AXBE editor is highlighted in blue. Data represent the mean of three independent experiments. **d**, The cleavage activity of mAAG-WT and mAAG-R165E•Y179F for the indicated motifs as a function of time. **e**, Heat map showing AXBEv2/AXBE fold changes in the frequencies of A-to-C, A-to-T and A-to-Y editing at 16 target sites with different NAN motifs in the canonical window, respectively. The specific motif and corresponding target information are listed on the top with the fold change values labeled in the box. All data represent the mean of three independent biological experiments. **f**,**g**, The on-target transversion efficiency induced by AXBEv2-SpNG and AXBEv2-SpRY using indicated sgRNAs with PAMs, which are compatible with SpCas9-NG (NGN) and SpCas9-RY (NRN/NYN) variants in HEK293T cells (ABE8e-SpNG and ABE8e-SpRY as a control). Single triangle and diamond represent A-to-C and A-to-T editing of an individual replicate, respectively. Data represent the mean ± s.d. of $n = 3$ independent replicates.

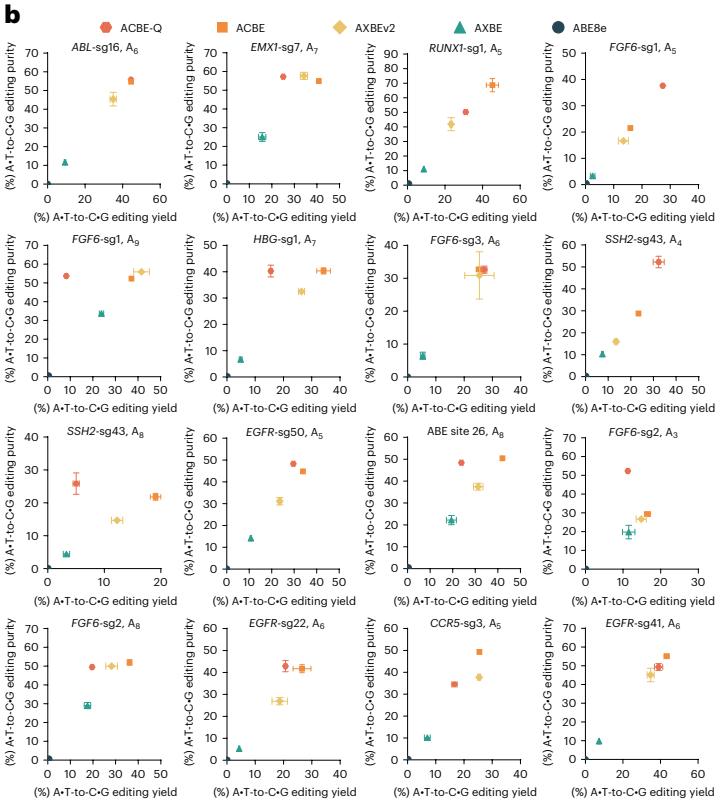
https://doi.org/10.1038/s41587-023-01821-9

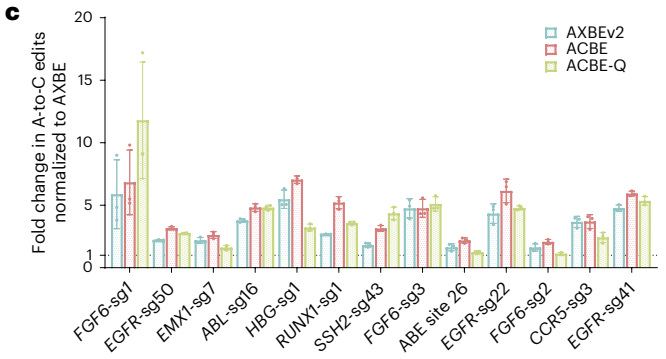
Supplementary Fig. 13a–c), and considerable indels (18% on average) were also observed among the seven pups (Supplementary Fig. 13d). The expression of dystrophin was disrupted in highly mutated founders, which efficiently transmitted the mutations to the F1 generation (Fig. 6e,f), indicating successful generation of a Duchenne muscular dystrophy (DMD) disease model.

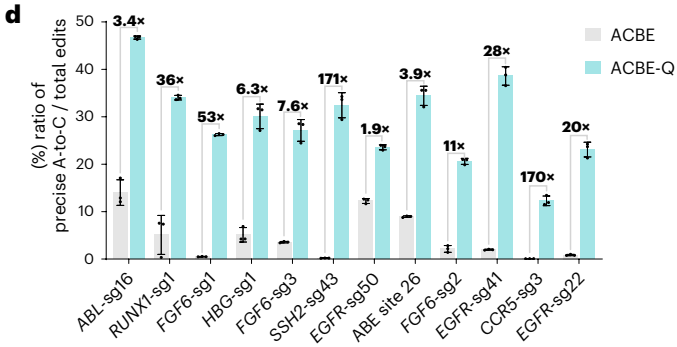


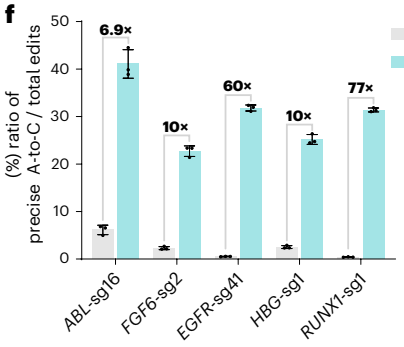
https://doi.org/10.1038/s41587-023-01821-9

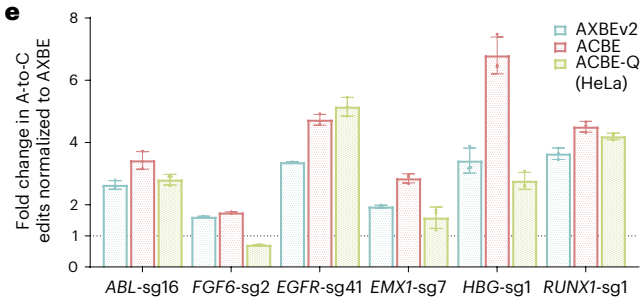












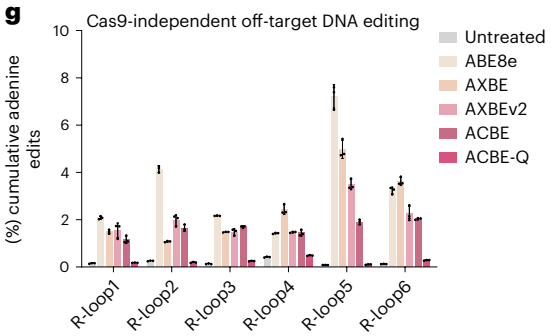


Fig. 5 | Enhanced editing properties of ACBEs via Cas embedding or TadA-8e engineering. a, Adenine conversion efficiency of AXBE, AXBEv2 and AXBEv2 variants at the *RUNX1*-sg1 site using ABE8e as a control 3 d after transfection of conducts. Preferred variants AXBEv2-Ce07 (ACBE) and AXBEv2-Ce07Q (ACBE-Q) are highlighted in orange. Data represent the mean ± s.d. of $n = 3$ independent replicates. **b**, A•T-to-C•G editing performance of AXBE, AXBEv2, ACBE and ACBE-Q at 13 genomic loci in HEK293T cells using ABE8e as a control. Editing positions of protospacers are annotated at the top of each diagram (PAM is at positions 21–23). A•T-to-C•G editing yield and product purity are shown on the *x* axis and *y* axis, respectively. Data represent the mean ± s.d. of $n = 3$ independent replicates. **c**,**e**, Summary of the fold changes for AXBEv2, ACBE or ACBE-Q in the highest A-to-C editing frequencies normalized to AXBE at indicated targets in HEK293T cells and HeLa cells, respectively. Data represent the mean ± s.d. of $n = 3$ independent replicates. **d**,**f**, The ratio defined as the A-to-C editing frequencies of the highest position / total rates of cumulative edits, including indels induced by ACBE and ACBE-Q at indicated targets (except for a site with only one adenine) in HEK293T cells and HeLa cells, respectively. Fold changes indicate the average of fold increases of ACBE-Q compared to ACBE from each target tested. Data represent the mean ± s.d. of $n = 3$ independent replicates. **g**, Cas9-independent off-target analysis of cumulative adenine edits induced by indicated editors using orthogonal R-loop assays (additional details in Methods). Data represent the mean ± s.d. of $n = 3$ independent replicates.

Premature stop codons cause numerous genetic diseases and can also be leveraged to terminate gene expression for multiple purposes. However, in AT-only codons, neither CBEs/CGBEs nor ABEs are able to generate premature stop codons. We converted TAA into TCA on the antisense stand to create a TGA stop codon in sense strand to terminate tyrosinase gene transcription in mouse embryos to generate an albinism disease model (Fig. 6g). ACBE-Q introduced desired A_6-to-C conversions in 13 of 15 F0 pups with an efficiency averaging 44% (up to 75%) (Supplementary Fig. 13e,f). Notably, the A-to-T byproducts on this TAA target also induced a stop codon to re-enforce the efficiency of transcriptional termination in the tyrosinase gene, leading to installation of stop codons in 45% of the alleles on average and an albino phenotype in founders (Fig. 6h,i). As previous reports showed that base transversion by prime editors was inefficient in primary cells and mouse embryos[4,5], our data demonstrated that ACBE-Q could generate highly efficient and accurate A-to-C transversion in vivo and expand the range of base editor for generation of premature stop codons at AT-only sites.

Installation and correction of SNVs with ACBEs

As A•T-to-C•G conversion can, in principle, reverse the second most common class of pathogenic SNVs (Fig. 1a), ACBE and ACBE-Q were evaluated to either mimic or correct pathogenic variants in human cells. To further examine ACBEs for generation of stop codons in an AT-rich sequence context, we tried to convert TAA into TCA on the antisense stand to create a TGA stop codon in the sense strand, which would terminate *SH3TC2* gene transcription to mimic Charcot–Marie–Tooth disease (accession: VCV001030852.1) in human cells. As A-to-T byproducts on this TAA target also create termination codons on the opposite strand, ACBEs generated premature stop codon in 63% of the alleles, among which 51% were A-to-C edits (Fig. 6j and Supplementary Fig. 14a). ACBE-Q was also able to precisely install a desired A-to-C mutation eliciting L618R missense substitution in *MYO7A* gene causing Usher syndrome type 2 (ref. 37) with 31% efficiency (Fig. 6k and Supplementary Fig. 14b). To investigate the therapeutic potential of ACBEs, a stable cell line containing a target sequence of c.1145G > T in *STAT3*, which causes recurrent infectious disease[38], was generated. Both ACBE and ACBE-Q introduced the desired A-to-C corrections with over 40% efficiency at the target site (Fig. 6l and Supplementary Fig. 14c). These data showed that ACBEs can generate efficient A-to-C conversions to either mimic or correct pathogenic SNVs. Due to its narrower editing window, ABCE-Q offers fewer bystander edits, which can be useful for diverse applications (Supplementary Fig. 14b,c), including potential correction of pathogenic mutations for therapeutic applications.

Discussion

Base editing technology enables highly efficient and programmable nucleotide conversions without requirement of donor templates and induction of DSBs. Previously reported base editors, including CBE, CGBE/GBE and ABE, can generate eight different types of base conversions, but the other four types of base substitutions (A-to-C, A-to-T, T-to-A and T-to-G) have not been achieved through base editing (Fig. 1a). This study describes the development and characterization of a series of base editors that mediate adenine transversions, including an ACBE architecture (ACBE-Q) that is capable of catalyzing efficient and selective A•T-to-C•G base transversions in distinct human cell lines and in mouse embryos, suggesting their broad application scenarios of variant cell types.

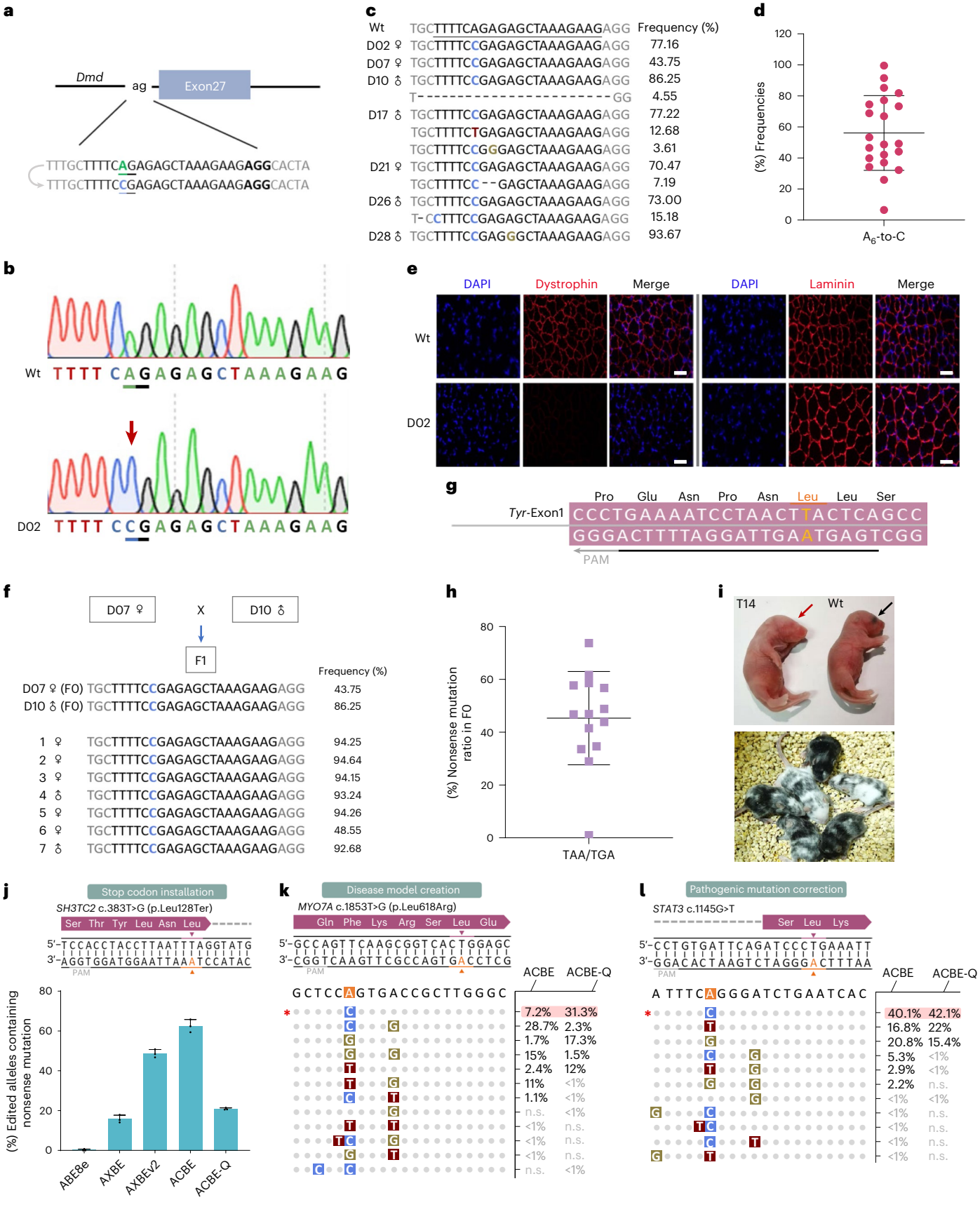
The critical steps to generate adenine transversions are adenine deamination and subsequent removal of hypoxanthine by DNA glycosylase. However, endogenous AAGs are inefficient or have difficulty accessing and catalyzing removal of target inosines, because their optimal substrates are alkyladenosines. For example, the AAG homologue in *E. coli* has an 84-fold weaker K_m and an even poorer (~70,000-fold) catalytic efficiency (k_{cat}/K_m) toward inosine versus alkyladenosine[20]. Several studies have attempted adenine transversion base editing via ectopic expression or fusion of hAAG with ABEmax or ABE8e in human embryonic stem cells or in plants but did not observe A-to-Y editing[39,40]. After fusing nine enzymes with hypoxanthine excision potential with ABE8e to increase their accessibility to substrate inosines, only DNA glycosylases from mouse, rat and *B. subtilis* supported detected adenine transversions, with hAAG possessing marginal activity to stimulate A-to-Y conversions (Fig. 1c). Low hypoxanthine glycosylase activity of hAAGs might be the reason why previous studies failed to detect adenine transversions. Through extensive molecular evolution of hAAG, Yang et al.[36] successfully developed an adenine transversion base editor, AYBE, indicating that variant AAGs can be applied to induce adenine transversion. Through direct comparation, our AXBEv2 exhibits similar activity with AYBEv3 (Supplementary Fig. 12a). Due to high bystander A-to-G editing, the major utility of adenosine transversion editors is hyper-mutagenesis-based applications. Further evolution of AAGs is very important to expand the targeting scope and efficiency of adenine transversion editors.

Both AXBE and AYBE are not suitable to generate base conversions at specific positions due to bystander adenosine conversions, which severely limits their applications, such as generation of disease model and gene therapy. We found that introducing an N108Q mutation in TadA-8e greatly reduced bystander A•T-to-G•C editing[34], and the ACBE-Q variant showed much higher accuracy than ACBE (averaging 43-fold) to obtain desired single A-to-C yields efficiently (Fig. 5d). Additionally, introduction of the N108Q mutation markedly reduced Cas9-independent DNA off-target editing to near-background levels (Fig. 5g). Based on our observation that RNA off-target editing of AXBE reduced to 10% of that achieved with ABE8e (Fig. 3c), we speculate that AXBEv2 induces similar or fewer RNA off-target edits compared to AXBE because the higher glycosylase activity of mAAG-EF probably results in fewer RNA edits. Additionally, ACBE-Q was developed through two known strategies to reduce RNA off-targeting effects:

Fig. 6 | Applications of ACBEs in mouse embryos and human cells. a, The target sequence of the splicing acceptor site in intron 26 of the mouse *Dmd* gene (the 'ag' sequence and PAM are in black). The desired A6 (green)-to-C (blue) transversion is shown. **b**, Sanger sequencing chromatograms of DNA from wild-type (Wt) and representative F0 mice injected with ACBE-Q mRNA and sgRNA targeting the *Dmd* site. **c**, Genotyping of representative F0 pups treated with ACBE-Q mRNA. The frequencies of mutant alleles were determined by HTS. **d**, Desired A6-to-C transversion frequencies of A-to-C founders generated by ACBE-Q. Data represent the mean ± s.d., and each data point represents an individual mouse ($n = 21$). **e**, Immunofluorescence staining of tibialis anterior muscle tissues from 5-week-old Wt and founder (D02) mice. Dystrophin or laminin is indicated in red, with nuclei in blue. Histological analysis images are representative of three independent experiments. Scale bars, 50 μm. **f**, HTS alignments of mutant sequences from F1 generated by mating founder D07 (♀) with D10 (♂). **g**, The target DNA and corresponding amino acid sequence in exon 1 of the mouse *Tyr* gene. The target A•T pair and corresponding amino acid are highlighted in orange. **h**, Data represent the mean ± s.d., and each data point represents an individual mouse ($n = 14$). **i**, The albino phenotype of F0 pups by ACBE-Q injection. The photo on the top shows Wt and founder (T14) mice at 7 d old, whereas the bottom was taken at 14 d old. **j**, Ratio of HTS reads containing the stop codons introduced by the indicated editors for the disease-relevant cell model (*SH3TC2* c.383T > G). The desired A is marked in orange. Data represent the mean ± s.d. of $n = 3$ independent replicates. **k**, Generation of A-to-C conversion of the disease-relevant cell model by ACBE or ACBE-Q (*MYO7A* c.1853T > G). **l**, Correction of pathogenic mutations by ACBE or ACBE-Q in a target-integrated stable cell line (*STAT3* c.1145G > T). The desired A is marked in orange, and the first line marked with an asterisk indicates desired alleles. In **k** and **l**, the allele frequencies were determined by HTS. The target adenines in orange indicate the pathogenic mutations. The desired transversion patterns marked with asterisks are shown in the top, and corresponding values are highlighted with pink boxes. Data represent the mean of three biologically independent experiments.

https://doi.org/10.1038/s41587-023-01821-9

embedding the deaminase into Cas9n and introducing the N108Q mutation in TadA-8e[33,34]. Therefore, we think that ACBE-Q will induce background levels of RNA edits. We also demonstrated that ACBE-Q is highly efficient and accurate in mouse embryos with up to 100% A-to-C frequency (Fig. 6c,d,h and Supplementary Fig. 13b,e,f), suggesting its potential to support a wide range of in vivo applications.



Although reducing the deaminase editing window could increase the purity of CGBEs[11], ACBE-Q did not significantly increase in A•T-to-C•G product purity (Fig. 5a,b). This difference between CGBEs and ACBEs may arise from the difference in DNA repair mechanisms between adenine and cytosine deamination outcomes, especially the large difference in endogenous DNA glycosylase activity on uridine and inosine. Additional dissection of the DNA damage repair mechanisms involved in inosine excision, similar to our previous study on CGBEs[13], may suggest strategies to further increase the efficiency and product purity of adenine transversion editors.

Adenine base transversion editors could be used for a wide range of applications. As AXBEs can access up to 558 codon and 147 amino acid alterations (with 436 codon and 115 amino acid alterations that are unique compared to other editors) (Fig. 3d,e and Supplementary Table 4), they may serve as a useful tool for laboratory evolution, genetic screening and lineage tracing, especially if combined with a dual base editing strategy[41]. As TadA-8e and mAAG are compatible with PAM-relaxed Cas9 variants to expand the targeting scope of A-to-C editing, generation of ACBEs based on other CRISPR–Cas systems would further broaden the targeting range and potentially correct a wide range of C•G-to-A•T pathogenic SNVs, which represent the second most common category of known human pathogenic SNVs (Fig. 1a). As a prototype of base editor to induce A-to-C conversion, the product purity and efficiency of ACBEs are not perfect, but they are potentially used to correct SNVs to ameliorate many genetic diseases. For example, we and others have shown that about 10% of correction in hepatocytes can ameliorate certain liver metabolic diseases and hemophilia (such as phenylketonuria[42,43], ornithine transcarbamylase (OTC) deficiency-caused hyperammonemia[44] and hemophilia B[45,46]). All these diseases have reported C•G-to-A•T pathogenic SNVs (obtained from the ClinVar database[1]) that can potentially be corrected by ACBEs. Because natural adenosine deamination and hypoxanthine excision in DNA are very inefficient in mammalian cells, our ACBEs also provide unique tools for probing the mechanism of how specific abasic sites opposite thymine are repaired, which could lead to further improvements of adenosine transversion editors. The development of AXBEs and ACBEs thus substantially expands the capabilities of base editing for both basic research and potential therapeutic applications.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-023-01821-9.

References

1. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
2. Liu, G., Lin, Q., Jin, S. & Gao, C. The CRISPR–Cas toolbox and gene editing technologies. *Mol. Cell* **82**, 333–347 (2022).
3. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
4. Petri, K. et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2022).
5. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell Discov.* **6**, 27 (2020).
6. Bock, D. et al. In vivo prime editing of a metabolic liver disease in mice. *Sci. Transl. Med.* **14**, eabl9238 (2022).
7. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
8. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652 (2021).
9. 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).
10. 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).
11. Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).
12. Zhao, D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021).
13. Koblan, L. W. et al. Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* **39**, 1414–1425 (2021).
14. 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).
15. Arbab, M. et al. Determinants of base editing outcomes from target library analysis and machine learning. *Cell* **182**, 463–480 (2020).
16. Caldecott, K. W. Single-strand break repair and genetic disease. *Nat. Rev. Genet.* **9**, 619–631 (2008).
17. Thompson, P. S. & Cortez, D. New insights into abasic site repair and tolerance. *DNA Repair (Amst).* **90**, 102866 (2020).
18. Alseth, I., Dalhus, B. & Bjoras, M. Inosine in DNA and RNA. *Curr. Opin. Genet. Dev.* **26**, 116–123 (2014).
19. Schouten, K. A. & Weiss, B. Endonuclease V protects *Escherichia coli* against specific mutations caused by nitrous acid. *Mutat. Res.* **435**, 245–254 (1999).
20. Saparbaev, M. & Laval, J. Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. USA.* **91**, 5873–5877 (1994).
21. Saparbaev, M., Mani, J. C. & Laval, J. Interactions of the human, rat, *Saccharomyces cerevisiae* and *Escherichia coli* 3-methyladenine-DNA glycosylases with DNA containing dIMP residues. *Nucleic Acids Res.* **28**, 1332–1339 (2000).
22. Wyatt, M. D. & Samson, L. D. Influence of DNA structure on hypoxanthine and 1,N^6-ethenoadenine removal by murine 3-methyladenine DNA glycosylase. *Carcinogenesis* **21**, 901–908 (2000).
23. Lee, H. W., Dominy, B. N. & Cao, W. New family of deamination repair enzymes in uracil-DNA glycosylase superfamily. *J. Biol. Chem.* **286**, 31282–31287 (2011).
24. Ayala-Garcia, V. M., Valenzuela-Garcia, L. I., Setlow, P. & Pedraza-Reyes, M. Aag hypoxanthine-DNA glycosylase is synthesized in the forespore compartment and involved in counteracting the genotoxic and mutagenic effects of hypoxanthine and alkylated bases in DNA during *Bacillus subtilis* sporulation. *J. Bacteriol.* **198**, 3345–3354 (2016).
25. Sidorkina, O., Saparbaev, M. & Laval, J. Effects of nitrous acid treatment on the survival and mutagenesis of *Escherichia coli* cells lacking base excision repair (hypoxanthine–DNA glycosylase–ALK A protein) and/or nucleotide excision repair. *Mutagenesis* **12**, 23–28 (1997).
26. 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).
27. 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).

28. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).
29. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl Acad. Sci. USA.* **97**, 13573–13578 (2000).
30. Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).
31. Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
32. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR–Cas9 variants. *Science* **368**, 290–296 (2020).
33. Li, S. et al. Docking sites inside Cas9 for adenine base editing diversification and RNA off-target elimination. *Nat. Commun.* **11**, 5827 (2020).
34. Chen, L. et al. Engineering a precise adenine base editor with minimal bystander editing. *Nat. Chem. Biol.* **19**, 101–110 (2023).
35. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
36. Tong, H. et al. Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat. Biotechnol.* 1–5 (2023).
37. Kabahuma, R. I. et al. Spectrum of *MYO7A* mutations in an indigenous South African population further elucidates the nonsyndromic autosomal recessive phenotype of DFNB2 to include both homozygous and compound heterozygous mutations. *Genes* **12**, 274 (2021).
38. Holland, S. M. et al. *STAT3* mutations in the hyper-IgE syndrome. *N. Engl. J. Med.* **357**, 1608–1619 (2007).
39. Park, J. C. et al. High expression of uracil DNA glycosylase determines C to T substitution in human pluripotent stem cells. *Mol. Ther. Nucleic Acids* **27**, 175–183 (2022).
40. Zeng, D. et al. Exploring C-to-G and A-to-Y base editing in rice by using new vector tools. *Int. J. Mol. Sci.* **23**, 7990 (2022).
41. Zhang, X. et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* **38**, 856–860 (2020).
42. Hamman, K. et al. Low therapeutic threshold for hepatocyte replacement in murine phenylketonuria. *Mol. Ther.* **12**, 337–344 (2005).
43. Yin, S. et al. Enhanced genome editing to ameliorate a genetic metabolic liver disease through co-delivery of adeno-associated virus receptor. *Sci. China Life Sci.* **65**, 718–730 (2022).
44. Yang, Y. et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).
45. Li, H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217–221 (2011).
46. Guan, Y. et al. CRISPR/Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse. *EMBO Mol. Med.* **8**, 477–488 (2016).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Plasmid construction

The primers and oligonucleotides used in this study are listed in Supplementary Table 2. Plasmid DNA sequences are listed in Supplementary Table 1. ABE8e (138489), lentiCRISPR v2 (52961) and pCMV-PE2 (132775) plasmids were purchased from Addgene. New base editor plasmids were constructed using a previously published method[41]. In brief, DNA fragments were amplified using PrimeSTAR Max DNA Polymerase (TaKaRa) and assembled with a ClonExpress MultiS One Step Cloning Kit (Vazyme) according to the manufacturer's protocol. Codon-optimized AAGs, endonuclease V and MLH1dn sequences for human cell expression were synthesized (GENEWIZ). PEmax architecture employing a human codon-optimized RT and an additional C-terminal c-Myc NLS with R221K/N394K mutations in SpCas9 was constructed. The codon-optimized MLH1dn sequence was co-expressed with pCMV-PEmax using a P2A sequence for the PE4max construct. AH candidates were constructed by attaching human codon-optimized AAGs and endonuclease V to the C-terminal of ABE8e vector backbone. A series of AXBE variants introduced individual or combinational mutations and were constructed by site-directed mutagenesis using a PCR-based method. The SpCas9 variants nCas9-NG (D10A) and nCas9-RY (D10A) were synthesized (GENEWIZ) and constructed to obtain ABE8e-SpNG, ABE8e-SpRY, AXBEv2-SpNG and AXBEv2-SpRY. Engineered ACBE constructs were generated by inserting TadA-8e/TadA-8e-N108Q and mAAG (R165E•Y179F) inside SpCas9 with different combinations. AYBEv3 was constructed by integrating the synthesized fragment including corresponding mutations at the hAAG (GENEWIZ) based on ABE8e. For construction of sgRNAs, a pair of oligonucleotides were annealed from 95 °C down to room temperature and ligated into BbsI-linearized U6-sgRNA(sp)-EF1α-GFP (Thermo Fisher Scientific). Plasmids for mammalian expression of epegRNAs containing the spacer sequence, 'flip and extension' (F+E) sgRNA scaffold, 3′ extension and tevoPreQ$_1$ motif were cloned using Golden Gate assembly as previously reported[7]. A lentiviral transfer plasmid was constructed by cloning a 150-bp fragment associated with a C•G-to-A•T disease-associated gene from the ClinVar database into a modified lenti-vector (lentiCRISPR v2). Antibiotics were used at the following working concentrations: ampicillin, 100 μg ml^{-1}, and kanamycin, 50 μg ml^{-1}. Plasmids used for further transfection were isolated using the TIANprep Mini Plasmid Kit according to the manufacturer's instructions.

Cell culture and transfection

Human HEK293T cells (American Type Culture Collection (ATCC), CRL-3216) and HeLa cells (ATCC, CCL-2) were cultured in DMEM (Gibco) supplemented with 10% (v/v) FBS (Gibco) and 1% (v/v) penicillin–streptomycin. All cell types were maintained at 37 °C with 5% CO_2 and passaged every 2 or 3 d. HEK293T cells or HeLa cells were seeded in 24-well plates (Corning) and transfected at 70–80% confluence. For base editing, 750 ng of base editor plasmid and 250 ng of sgRNA plasmid were co-transfected into HEK293T cells or HeLa cells using polyethyleneimine (PEI, Polysciences) following the manufacturer's protocol. For prime editing, 750 ng of PE4max plasmid, 250 ng of epegRNA plasmid and 83 ng of nicking sgRNA plasmid (for PE5max) were co-transfected into HEK293T cells using PEI. For Cas9-independent off-target evaluation, the orthogonal R-loop assay was modified by dSaCas9-sgRNA plasmid. Similarly, 300 ng of SpCas9 sgRNA plasmid, 400 ng of base editor plasmid (ABE8e, AXBE, AXBEv2, ACBE or ACBE-Q) and 300 ng of dSaCas9-sgRNA plasmid were co-transfected into HEK293T cells using PEI. For the RNA sequencing experiment, HEK293T cells were seeded in 10-cm dishes and transfected with 25 μg of Cas9n-P2A-GFP, ABE8e-P2A-GFP and AXBE-P2A-GFP plasmids using PEI.

Genomic DNA extraction and amplification

HEK293T cells or HeLa cells transfected after 72 h were washed with 1× PBS and digested with 0.25% trypsin (Gibco) for fluorescence-activated cell sorting (FACS). Positive cells with green fluorescent protein (GFP) were collected, and the genomic DNA was extracted using QuickExtract DNA Extraction Solution (Epicentre) according to the manufacturer's recommended protocol. The extraction solution was incubated at 65 °C for 6 min and then at 98 °C for 2 min. To obtain genotype of modified mice, genomic DNA for PCR was extracted from tail tips of newborn mice using the One Step Mouse Genotyping Kit (Vazyme) based on the manufacturer's instructions. The extraction solution was incubated at 55 °C for 30 min and then at 95 °C for 5 min. To assess gene editing efficiency, genome loci of interest were amplified with site-specific primers (Supplementary Table 2) using KOD-Plus-Neo DNA Polymerase (Toyobo) or DNA Polymerase in the One Step Mouse Genotyping Kit (Vazyme).

Western blot assay

The western blot assays were performed as previously described[47]. In brief, HEK293T cells were cultured in six-well plates (Corning) and transfected with 3.75 μg of Flag-tagged base editor plasmids. Three days after transfection, cells were lysed using RIPA buffer with proteinase and phosphatase inhibitors (Calbiochem). The total protein samples were boiled and quantified using the BCA Protein Assay Kit (Thermo Fisher Scientific). Then, 20 μg of total protein was loaded per well into a 15-well 8% SDS-polyacrylamide gel for electrophoresis and transferred to nitrocellulose membranes (Millipore). Subsequently, the membranes were blocked with 5% BSA for 1 h at room temperature and incubated with different primary antibodies separately overnight, including anti-tubulin (1:5,000 dilution, Abcam, ab210797) and the anti-Flag (1:10,000 dilution, Sigma-Aldrich, F1804). Then, the membranes were incubated with goat anti-rabbit IgG H&L (IRDye 800CW, 1:10,000 dilution, Abcam, ab216773) or anti-mouse IgG H&L (Alexa Fluor 488, 1:10,000 dilution, Abcam, ab150113) for 1 h and then visualized using an Odyssey imager. Uncropped blots are shown in Supplementary Fig. 15.

RNA sequencing experiments

Seventy-two hours after transfection, HEK293T cells were washed with 1× PBS and digested with 0.25% trypsin (Gibco). Around 400,000 cells with the top 15% GFP-positive signal were collected through on a FACSAria III (BD Biosciences). FACS gating data were collected using FACSDiva version 8.0.2 (BD Biosciences). mRNA was prepared according to the standard protocol. A total of 3 μg of RNA per sample was adopted as input material to get preparations for the sample. Sequencing libraries were generated with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs) following the manufacturer's instructions. Sequence library was purified (Agencourt AMPure XP Beads), and the quality was evaluated on a Bioanalyzer (Agilent High Sensitivity Chip). According to the manufacturer's instructions, index-coded samples were clustered by a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina). After cluster generation, the library was sequenced on an Illumina NovaSeq 6000 platform, and 125–150-bp paired-end reads were obtained.

Transcriptome-wide RNA analysis

The analysis of RNA sequencing data was performed as previously reported[48]. For the RNA sequencing analysis, clean data were obtained from raw data of FASTQ format by removing adapter and trimming low-quality base with Trimmomatic. At the same time, the Q20, Q30 and GC contents of the clean data were calculated. All the downstream analysis was based on the clean data. Paired-end clean reads were aligned to the reference genome (Ensembl GRCh38) with index added using Hisat2 version 2.0.5. 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 (version 4.0) software was used to perform single-nucleotide polymorphism (SNP) calling.

Variant loci of ABE8e or AXBE overexpression were filtered to exclude sites without high-confidence reference genotype calls of the control experiment. The read coverage for a given SNV in the control group should be more than 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 control were filtered to include only loci with more than 10 reads and with more than 0% of reads containing an alternate allele. Base edits labeled as A-to-I comprise A-to-I edits on the positive strand as well as T-to-C edits on the negative strand.

Protein preparation

Wild-type and mutant mAAG (enzymatic domain) were cloned into pMCSG7, separately. The protein was then expressed in *E. coli* BL21 strain. Cells were grown at 37 °C until culture density OD_{600} reached 1.2. The culture was cooled down to 20 °C and induced with 0.4 mM IPTG for 16 h. The cells were harvested by centrifugation at 1,500*g* per minute for 50 min and lysed by a high-pressure homogenizer. Cell debris was removed by centrifugation, and the soluble fraction was applied onto an Ni-affinity column. The column was washed thoroughly with 20 ml of wash buffer1 (20 mM Tris pH 7.5, 1 M NaCl, 25 mM imidazole) and 20 ml of wash buffer2 (20 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole), which lately eluted with 15 ml of elution buffer (20 mM Tris pH 7.5, 500 mM NaCl, 300 mM imidazole). Finally, the protein was purified by size-exclusion chromatography using a Superdex 75 column equilibrated with 20 mM HEPES pH 7.5, 150 mM NaCl and 0.1 mM TCEP. The protein was pure when analyzed by SDS-PAGE. The protein was concentrated to 4 mg ml^{-1}.

In vitro DNA glycosylase activity assay

The DNA oligonucleotides containing hypoxanthine were 3′-end labeled with FAM (Supplementary Table 3). Each reaction contained 0.5 nM substrate DNA and 4 μg of purified glycosylase. The incubations were carried out in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 1 mM DTT and 2 mM EDTA at 37 °C for the indicated time. After incubation, NaOH was added to 0.15 M, and each reaction was raised to 95 °C for 15 min to break abasic sites. Later, HCl was added to the samples until the pH was adjusted to 6–10. Lastly, proteinase K was used; the reaction was treated at 55 °C for 1 h to separate DNA glycosylase and DNA; and 20% denaturing polyacryamide gel was used to separate the product from the substrate. Relative catalytic rate was analyzed by images taken on a Chemiluminescence Imaging System (Bio-Rad).

Preparation of sgRNA and mRNA

Chemically modified sgRNA was synthesized by GenScript. mRNA preparation was performed as previously described[47,49]. In brief, ACBE-Q coding region with T7 promoter was amplified by PCR using primer T7-mRNA (ACBE-Q)-F and -R as follows. T7-ACBE-Q PCR product was transcribed with an in vitro RNA transcription kit (mMACHINE T7 ULTRA Kit, Ambion) following the manufacturer's instructions. ACBE-Q mRNA was eluted in nuclease-free water and stored at −80 °C.

Primers	Primer sequence (5′–3′)
T7-mRNA (ACBE-Q)-F	GCCGCGATCACTAATACGACTCACTATAGGGAGAGCCGC
T7-mRNA (ACBE-Q)-R	CTAGACTTTCCTCTTCTTCTTGGGCTCGAATTCG

Animals and microinjection of zygotes

Animal manipulation was consistent with a previous report[50]. All animal experiments met the regulations drafted by the Association for Assessment and Accreditation of Laboratory Animal Care in Shanghai and were approved by the Experimental Animal Welfare Ethics Committee of East China Normal University (license number ARXM2022042). C57BL/6J and ICR mice used as embryo donors and foster mothers were housed in standard cages in a specific pathogen-free (SPF) facility under a 12-h dark/light cycle. For microinjection, the preparations, including superovulation, and the embryo collection were described in keeping with a previous report[49]. The mixture of ACBE-Q mRNA (200 ng μl^{-1}) and sgRNA (100 ng μl^{-1}) was 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 or after overnight culture in KSOM medium at 37 °C under 5% CO_2 in air.

Immunofluorescence staining

Tibialis anterior muscle of about 5-week-old wild-type or *Dmd* mutant mice was frozen in liquid-nitrogen-cooled isopentane and then embedded in OCT compound and kept frozen in liquid nitrogen. For immunofluorescence staining, the transversal sections (5–10 μm) were washed twice with PBST for 3–5 min. After blocking with 5% BSA for 40 min, the tissue sections were incubated with specific antibodies, respectively, overnight at 4 °C. Dystrophin or laminin proteins were detected with a 1:500 dilution of anti-dystrophin antibody (Abcam) or a 1:500 dilution of anti-laminin antibody (Abcam), respectively, and the nuclei were stained with DAPI (Sigma-Aldrich). After washing twice with PBST for 3–5 min, antibodies were detected by 1:1,000 dilution of Alexa Fluor 594 donkey anti-rabbit secondary antibody. Images were obtained by a Leica DMI4000B fluorescence microscope.

Creating stable cell line of disease model

To package lentivirus for generating a stable cell line, HEK293T cells were seeded into a 24-well plate. After 12–16 h, 300 ng of transfer plasmid (Lenti *STAT3* (c.1145G > T)-EF1α-DsRed-P2A-puro), 300 ng of pMD2.G and 300 ng of psPAX2 were co-transfected using PEI at 80–90% confluency according to the manufacturer's instructions. Eight hours after transfection, the medium was changed. Viral supernatant was harvested and centrifuged at 12,000 r.p.m. for 5 min to remove cellular debris 48 h after transfection. For construction of a stable cell line, 50 μl of filtered viral supernatant was added into a 12-well plate cultured with HEK293T cells of approximately 40–50% confluency. Twelve hours after transfection, the medium was changed. After 1 μg μl^{-1} puromycin (Sangon) selection for 3 d, cells were harvested from the well with the fewest surviving colonies to ensure single-copy integration and were further cultured for transfection.

Next-generation sequencing and data analysis

The second PCR amplifications were performed with primers containing an adaptor sequence (forward 5′-GGAGTGAGTACGGTGTGC-3′; backward 5′-GAGTTGGATGCTGGATGG-3′) and diverse barcode sequences at the 5′ end. The resulting HTS libraries were pooled and purified by electrophoresis with a 1.5% agarose gel using HiPure Gel Pure DNA Mini Kit (Magen) eluting with 60 μl of water and then sequenced on an Illumina HiSeq X Ten platform. To assess base editing efficiencies, the A-to-C, A-to-T, A-to-G efficiencies and indels in the HTS data were analyzed using BE-Analyzer[51]. To assess prime editing efficiencies, A-to-C efficiencies and indels in the HTS data were analyzed using CRISPResso2 (ref. 52). For base editing, base editing efficiencies were calculated as: base substitution reads divided by total reads. Purities were calculated as: percentage of (the reads of A-to-C edits) / (the reads of adenine edits without indels). Indel frequencies were calculated as: percentage of (the reads of indels) / (total reads). For prime editing, A-to-C substitution frequencies were calculated as: percentage of (the reads of intended edits without indels) / (total reads). Purities and indels were calculated the same as base editing.

Statistical analysis and reproducibility

Data are presented as mean ± s.d. from independent experiments. All statistical analyses were performed on $n = 3$ biologically independent experiments unless otherwise noted in the figure captions, using GraphPad Prism version 9.3.1 software.

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

Data availability
HTS data have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession codes PRJNA954164, PRJNA954271 and PRJNA954456 (refs. 53–55). RNA sequencing data have been deposited in the NCBI Sequence Read Archive under accession code PRJNA954055 (ref. 56). Source data for Figs. 1–6 and Supplementary Figs. 1–15 are presented with the paper. There are no restrictions on data availability. Source data are provided with this paper.

References

47. Chen, L. et al. Re-engineering the adenine deaminase TadA-8e for efficient and specific CRISPR-based cytosine base editing. *Nat. Biotechnol.* **41**, 1–10 (2022).
48. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
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. Zhang, X. et al. Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. *Nat. Cell Biol.* **22**, 740–750 (2020).
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).
53. Chen, L. et al. Adenine transversion editors enable precise and efficient A•T-to-C•G base editing in mammalian cells and embryos. National Center for Biotechnology Information Sequence Read Archive, BioProject PRJNA954164 (2023).
54. Chen, L. et al. Adenine transversion editors enable precise and efficient A•T-to-C•G base editing in mammalian cells and embryos. National Center for Biotechnology Information Sequence Read Archive, BioProject PRJNA954271 (2023).
55. Chen, L. et al. Adenine transversion editors enable precise and efficient A•T-to-C•G base editing in mammalian cells and embryos. National Center for Biotechnology Information Sequence Read Archive, BioProject PRJNA954456 (2023).
56. Chen, L. et al. Adenine transversion editors enable precise and efficient A•T-to-C•G base editing in mammalian cells and embryos. National Center for Biotechnology Information Sequence Read Archive, BioProject PRJNA954055 (2023).

Acknowledgements
We are grateful to S. Siwko (Texas A&M University Health Science Center) for proofreading the manuscript and to support from the East China Normal University Public Platform for Innovation (011). We thank Y. Zhang from the Flow Cytometry Core Facility of the School of Life Sciences at East China Normal University and H. Jiang from the Core Facility and Technical Service Center of the School of Life Sciences and Biotechnology at Shanghai Jiao Tong University. We thank L. Ji (MedSci) for designing schematic diagrams. This work was partially supported by grants from the National Key R&D Program of China (2019YFA0110802 to D.L. and 2022YFC3400203 to Y.G.); the National Natural Science Foundation of China (32025023 and 32230064 to D.L. and 82100773 to Y.G.); the Shanghai Municipal Commission for Science and Technology (21CJ1402200, 20140900200 and 20MC1920400 to D.L.); the Innovation Program of the Shanghai Municipal Education Commission (2019-01-07-00-05-E00054 to D.L. and NK2022010207 to D.L.); the Innovative Research Team of High-Level Local Universities in Shanghai (SHSMU-ZDCX20212200 to D.L.); Fundamental Research Funds for the Central Universities; and the East China Normal University Outstanding Doctoral Students Academic Innovation Ability Improvement Project (YBNLTS2021-026 to L.C.). P.B.R., A.A.S. and D.R.L acknowledge support from US National Institutes of Health grants (U01AI142756, R35GM118062 and RM1HG009490 to D.R.L) and the Howard Hughes Medical Institute.

Author contributions
L.C., D.R.L. and D.L. designed the experiments. P.R. and A.S. designed prime editing agents. L.C., M.H., C.L., H.G., G.R., X.G., D.Z., S.Z. and C.Q. performed the experiments. L.C., M.H., C.L., H.G., G.R., S.Z., D.Z., J.W., Y.Z, P.R., A.S., C.L., M.L., B.F., G.S., D.R.L. and D.L. analyzed the data. L.C. and D.L. wrote the manuscript, with input from all authors. D.L. supervised the research.

Competing interests
The authors have submitted patent applications based on the results reported in this study (L.C., D.L., M.H. and C.L.). P.R., A.S. and D.R.L. are co-inventors on prime editing patent applications. D.R.L. is a consultant for Prime Medicine, Beam Therapeutics, Pairwise Plants, Chroma Medicine and Nvelop Therapeutics, companies that use or deliver genome editing or genome engineering agents, and owns equity in these companies. The remaining authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-023-01821-9.

Correspondence and requests for materials should be addressed to Dali Li.

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

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

nature portfolio

Corresponding author(s): Dali Li

Last updated by author(s): May 4, 2023

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Targeted amplicons sequencing data were collected and demultiplexed by an Illumina HiSeq X Ten instrument. RNA-seq data were collected and demultiplexed by an Illumina NovaSeq 6000 instrument. FACS gating data were collected on a FACSAria III (BD Biosciences) using FACSDiva version 8.0.2 (BD Biosciences).
Data analysis	High-throughput sequencing data were analyzed by BE-Analyzer (http://www.rgenome.net/be-analyzer/#!) (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 (A>C, A>T and A>G), prime editing (A>C) and indel efficiencies. Potential DNA off-target sites for ABE8e and AXBE were predicated using Cas-OFFinder web (http://www.rgenome.net/cas-offinder/). RNA-seq data were analyzed using Hisat2 v2.0.5 and GATK (v4.0) software. GraphPad Prism 9.3 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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

HTS data have been deposited in the NCBI Sequence Read Archive database under accession codes PRJNA954164, PRJNA954271 and PRJNA954456. RNA-seq data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA954055. 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. Sample sizes were chosen based on standards and precedents in the field/ literature that have been carried out in the past, to obtain statistically significant data.
Data exclusions	No data were excluded from the analysis.
Replication	Three independent biological replicates were performed on different days. All replications were successful.
Randomization	Samples were randomly distributed into groups.
Blinding	Investigators were not blinded to group allocation in this research since experimental conditions were evident and all samples of treatment were consistent throughout experiments.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	Anti-laminin (Abcam, ab11575) (1:500 for Immunofluorescent staining). Anti-dystrophin (Abcam, ab15277) (1:500 for Immunofluorescent staining). Donkey Anti-Rabbit IgG H&L (Abcam, ab150076) (1:1000 for Immunofluorescent staining). Anti-tubulin(Abcam, ab210797) (1:5,000 for western blotting). Anti-Flag(Sigma, F1804)(1:10,000 for western blotting). goat anti-rabbit IgG H&L (Abcam, ab216773) (1:10,000 dilution for western blotting). goat anti-mouse IgG H&L (Abcam, ab150113) (1:10,000 dilution for western blotting).
Validation	The laminin antibody has been validated by mouse anterior tibialis (https://www.abcam.cn/laminin-antibody-ab11575.html).

Validation	The dystrophin antibody has been validated by mouse triceps muscle (https://www.abcam.cn/dystrophin-antibody-ab15277.html). The tubulin antibody has been validated by western blot in HEK293T cells (https://www.abcam.cn/tubulin-antibody-epr13796-loading-control-ab210797.html). The Flag antibody has been validated by western blot in HEK293T cells (https://www.sigmaaldrich.cn/sigma/f1804).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells (purchased from ATCC CRL-3216) and HeLa cells (purchased from ATCC CCL-2).
Authentication	Cells from ATCC were undergone authenticated manufacturers by STR profiling.
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, ICR mouse strain purchased from Shanghai Laboratory Animal Center were housed in standard cages at 20-22°C with 40-60% humidity in specific pathogen-free facility (SPF) under a 12h dark-light 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. Dmd mutant founders (#D02, #D03, #D04, #D07, #D11, #D15, #D18, #D20, #D21, #D23, #D24, #D29) were 5 week-old female. Dmd mutant founders (#D01, #D05, #D06, #D09, #D10, #D12, #D13, #D14, #D16, #D17, #D22, #D25, #D26, #D27, #D28, #D30, #D31, #D32) were 5 week-old male.
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 Experimental Animal Welfare Ethics Committee of East China Normal University (IACUC). License number is ARXM2022042.

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 methods. Cells were washed and filtered through a 45μm cell strainer cap before sorting (72h after transfection).
Instrument	FACSAria III (BD Biosciences)
Software	BD FACSDiva Software Diva 8.0.2
Cell population abundance	HEK293T cell numbers gated for target populations were similar in different biology replicates.
Gating strategy	For HEK293T cells, gates were established using uninfected control cells and GFP positive control. Gates were drawn to collect subsets of GFP-expressing cells. And for specified transcriptome profiling, cells with top 15% of GFP signal were sorted. Detailed gating strategy is provided in the Supplementary Note.

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