

Engineered IscB- ω RNA system with expanded target range for base editing

Received: 2 December 2023

Accepted: 17 July 2024

Published online: 15 August 2024

Check for updates

Qingquan Xiao ^{1,2,7}, Guoling Li ^{1,7}, Dingyi Han ^{2,3,7}, Haoqiang Wang ^{1,7}, Mingyu Yao ^{4,5,6,7}, Tingting Ma¹, Jingxing Zhou¹, Yu Zhang¹, Xiumei Zhang¹, Bingbing He ², Yuan Yuan¹, Linyu Shi ¹, Tong Li ¹✉, Hui Yang ^{1,2}✉, Jinhai Huang ^{4,5,6}✉ & Hainan Zhang ¹✉

As the evolutionary ancestor of Cas9 nuclease, IscB proteins serve as compact RNA-guided DNA endonucleases and nickases, making them strong candidates for base editing. Nevertheless, the narrow targeting scope limits the application of IscB systems; thus, it is necessary to find more IscBs that recognize different target-adjacent motifs (TAMs). Here, we identified 10 of 19 uncharacterized IscB proteins from uncultured microbes with activity in mammalian cells. Through protein and ω RNA engineering, we further enhanced the activity of IscB ortholog IscB.m16 and expanded its TAM scope from MRNRAA to NNNGNA, resulting in a variant named IscB.m16*. By fusing the deaminase domains with IscB.m16* nickase, we generated IscB.m16*-derived base editors that exhibited robust base-editing efficiency in mammalian cells and effectively restored Duchenne muscular dystrophy proteins in diseased mice through single adeno-associated virus delivery. Thus, this study establishes a set of compact base-editing tools for basic research and therapeutic applications.

Clustered regularly interspaced short palindromic repeat (CRISPR) Cas systems, such as type II Cas9 and type V Cas12 systems, serving as the prokaryotic adaptive immunity system against viruses, have been developed into genome-editing tools in basic research and gene therapy^{1–3}. The engineered Cas9 nickase (nCas9) or deactivated Cas9 (dCas9) versions fused with various domains have been established as base-editing, prime-editing and epigenome-editing technologies^{4–6}. However, the large size of Cas9 and Cas12, particularly nCas9-based gene-editing tools, hinders the application of gene editing based on adeno-associated virus (AAV) vectors. Recently, compact Cas9 (refs. 7–9), Cas12f homologs^{10–14} (400–700 aa) and TnpB^{15,16} (~400 aa),

the ancestral branch of Cas12, have been reported. However, because of their poor editing activity or lack of an HNH domain, these proteins have limited base-editing activity.

IscB proteins are encoded in a distinct family of IS200/IS605 transposons possessing HNH and RuvC domains, such as Cas9, and are thought to be the ancestor of Cas9 (refs. 17,18). However, the size of IscB proteins is only two fifths of that of Cas9 (~400 aa). Recent studies have shown that the IscB system (IscB- ω RNA) is a programmable long noncoding RNA (referred to as ω RNA)-guided DNA endonuclease and engineered OgeulscB-based base editors (enOgeulscB-BEs) exhibit high base-editing efficiency in mammalian cells¹⁹. The IscB

¹HuidaGene Therapeutics Co. Ltd., Shanghai, China. ²Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Chinese Academy of Sciences, Shanghai, China. ³College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. ⁴Eye Institute and Department of Ophthalmology, Eye & ENT Hospital, Fudan University, Shanghai, China. ⁵NHC Key Laboratory of Myopia and Related Eye Diseases, Key Laboratory of Myopia and Related Eye Diseases, Chinese Academy of Medical Sciences, Shanghai, China. ⁶Shanghai Research Center of Ophthalmology and Optometry, Shanghai, China. ⁷These authors contributed equally: Qingquan Xiao, Guoling Li, Dingyi Han, Haoqiang Wang, Mingyu Yao. ✉e-mail: tongli@huidagene.com; huiyang@huidagene.com; jinhaihuang@fudan.edu.cn; zhanghainan@simmm.ac.cn

system requires a 3' terminal target-adjacent motif (TAM) to recognize the target DNA (usually 6 nt) and the recently reported enOgeulscB requires 4 nt (NWRRNA). Complex TAM sequences greatly reduce the number of sites that can be edited. The narrow TAM range of OgeulscB in mammalian cells has become an obvious limitation. Therefore, it is necessary to develop higher-efficiency miniature base editors with a broader TAM range. Here, we identified 19 natural IscB- ω RNA systems with various TAM scopes from metagenome datasets. By engineering both the ω RNA and the IscB.m16 protein, we generated the IscB.m16* system (IscB.m16 containing E326R;T459E;P460S;T462H substitutions (IscB.m16^{RESH}) and en ω RNA) with robust editing activity and expanded the TAM range to NNNGNA in mammalian cells. We further developed IscB.m16*-based adenine and cytosine base editors demonstrating robust base-editing efficiency and broad target recognition in mammalian cells and mouse models. Moreover, we provide a comprehensive dataset of IscB- ω RNA systems with diverse TAM scopes and a strategy to widen the TAM range.

Results

Functional identification of uncharacterized IscB orthologs

To identify additional IscB- ω RNA systems with diverse TAMs, we downloaded 200 Gb of rumen metagenome-assembled genomes²⁰. We used a computational pipeline to annotate IscB orthologs and their corresponding ω RNAs, which led to the discovery of 19 uncharacterized IscB systems. These systems were phylogenetically clustered into three subgroups on the basis of a sequence alignment of IscB effector proteins (Fig. 1a and Extended Data Fig. 1). Through the protein sequence alignment encompassing 500 aa, we identified the conserved residues within the RuvC domain, HNH domain, PID (P1 interaction domain) and TID (TAM interaction domain), suggesting the possibility of nuclease and nickase activity (Supplementary Fig. 1).

To detect whether these IscB proteins and predicted corresponding ω RNAs were capable of cleaving DNA and characterized by TAM recognition, we performed a bacterial depletion assay. We cotransformed *Escherichia coli* cells with plasmids carrying IscB and its cognate ω RNA with a spacer, as well as a TAM library plasmid carrying target sequences complementary to the spacer and 8-bp randomized sequences (Supplementary Fig. 2a). Through this assay, a series of specific depleted TAM sequences were enriched associated with each IscB system, indicating that these natural IscB orthologs have RNA-guided endonuclease activity in prokaryotes (Fig. 1b and Supplementary Fig. 2b). Subsequently, we analyzed the relationship between the divergence of IscB proteins and differences in TAMs and observed that most IscB proteins have notable distinctions in both their amino acid sequences and their respective optimal TAMs (Supplementary Fig. 2c).

To further assess the nuclease activity of these IscB orthologs in human cells, we used a fluorescence reporter system. This system involved cotransfecting a plasmid expressing the IscB protein and its corresponding target ω RNA, along with a reporter plasmid encoding GFxxFP, into cultured HEK293T cells. Green fluorescent protein (GFP) activation based on the GFxxFP reporter (GFxx-target site-xxFP) was carried out by endonuclease-mediated double-strand breaks (DSBs) to the target site and single-strand annealing (SSA)-mediated repair. We then measured the enhanced GFP (EGFP) signal intensity of the inactivated GFxxFP reporter, which was activated by IscB-mediated DSBs²¹ (Fig. 1c). Using the GFxxFP reporter with the experimentally determined TAM for each IscB, 10 of 19 IscBs exhibited a significant increase (>9-fold ratio of target and nontarget, with target recognition >3.0%) in EGFP signal intensity relative to nontarget sequences. Notably, IscB.m16 exhibited the highest signal intensity (Fig. 1d).

Engineering ω RNA to improve editing efficiency

Guide RNA (gRNA) engineering strategies have been widely applied to enhance the cleavage activity of RNA-guided nucleases^{11,12,19}. To enhance the activity of the natural IscB.m16 system, we engineered

its ω RNA by truncation or mutagenesis, generating ω RNA variants in five stem-loop regions: R1, R2, R3, R4 and R5 (Fig. 2a and Supplementary Fig. 3a). We performed the truncations by shrinking loops and truncating long stems. We screened the editing efficiency using the GFxxFP reporter and observed increased editing activity in the case of R1 (R1- Δ 13) and R5 (R5- Δ 10) truncated ω RNA (Fig. 2b). To increase the stability of the ω RNA, we replaced the A•U or mismatched base pairs in stem regions with thermodynamically stable G•C base pairs. We replaced the mismatched G•U and partial A•U base pairs in stem regions with thermodynamically stable G•C base pairs and five variants exhibited increased activity (Supplementary Fig. 4). For the R1- Δ 13 and R5- Δ 10 truncated ω RNA, we further combined five mutations and found that v2.27 (del15–20, del29–35, del171–180, 24-G, 25-C, 57-G, 79-C and 117-C) showed enhanced editing activity at the AAAGCA TAM reporter (Fig. 2c). Similarly, to improve the activity of the IscB.m17 system, we truncated six stem loops of the ω RNA on the basis of their secondary structure and found that R1 (R1- Δ 59) and R6 (R6- Δ 9) truncated ω RNA showed increased editing activity (Fig. 2d and Supplementary Fig. 3b). We then generated a variant with slightly improved editing activity by replacing A•U with C•G base pairs in the R1 stem loop of the truncated ω RNA (R1- Δ 59) (Fig. 2e). Notably, a truncation of the first (R1) or last (R5 or R6) stem loop of the ω RNA improved the IscB activity, while truncation of the intermediate (R2, R3, R4 or R5) stem loops markedly reduced activity. To further test this hypothesis and obtain more IscB systems with high activity, we trimmed the R1 and/or R5 stem loops of the ω RNA from four other IscB systems with different TAM ranges and activity in mammalian cells. We found that the truncation of R1 and/or R5 from IscB.m1, IscB.m15 and IscB.m18 markedly improved the editing activity (Supplementary Fig. 3c–e and Extended Data Fig. 2). Taken together, our extensive engineering of ω RNA resulted in numerous active IscB systems, particularly IscB.m16 and IscB.m17.

Engineering IscB to expand recognition and enhance activity

Substitutions of amino acid residues in the DNA-binding pocket or cleavage domains with positively charged arginine have been shown to enhance the editing activity of RNA-guided nucleases in eukaryotic cells^{21–23}. We performed a sequence alignment analysis of IscB.m16 and OgeulscB. According to conserved sequences, we divided the different structure domains of IscB.m16 and further performed an arginine scanning mutagenesis in the PID, TID and RuvC domain. According to the activated EGFP fluorescence intensity of cells with an AAAGAA TAM reporter, over 20 of 138 variants in the RuvC domain exhibited improved editing activity compared to wild-type (WT) IscB, with one variant (E326R) showing the highest editing activity (Fig. 3a).

Considering that the PID and TID domains are related to TAM recognition, we next screened 124 variants in these two domains using six GFxxFP reporters with different TAMs to broaden the TAM range. These reporters had the same target sequences but different 6-nt TAMs (AAAGAA, CAAGAA, ACAGAA, AACGAA, AAAGCA and AAAGAC). Compared to WT IscB.m16, seven variants (M424R, T462R, N463R, T465R, Q475R, K478R and I504R) showed improved activity and TAM recognition, as evidenced by the increase (>1.05-fold) in EGFP fluorescence intensity for all six reporters relative to the WT (Supplementary Fig. 5). Meanwhile, through predicted structure analysis of IscB.m16, we identified 11 potential sites associated with TAM recognition: H380, Q381, V433, T459, P460, I461, F467, Y468, R476, K478 and L481. In order to broaden the TAM range with improved editing efficiency, we conducted saturation mutagenesis at these 18 sites—the 7 sites from PID and TID screening and the 11 predicted sites. We then screened these mutants using a TAM pool characterized by low activity. TAM pool 1 consisted of ACAGAA, AATGAA, AAACAA, AAAGCA and AAAGAC, which had relatively low editing efficiency recognized by the IscB.m16 WT (Extended Data Fig. 3a,b). Using a similar fluorescence reporter system, we found that some variants, especially P460S, T462H, T462L and T465V, greatly enhanced the editing activity of TAM pool 1 (Fig. 3b

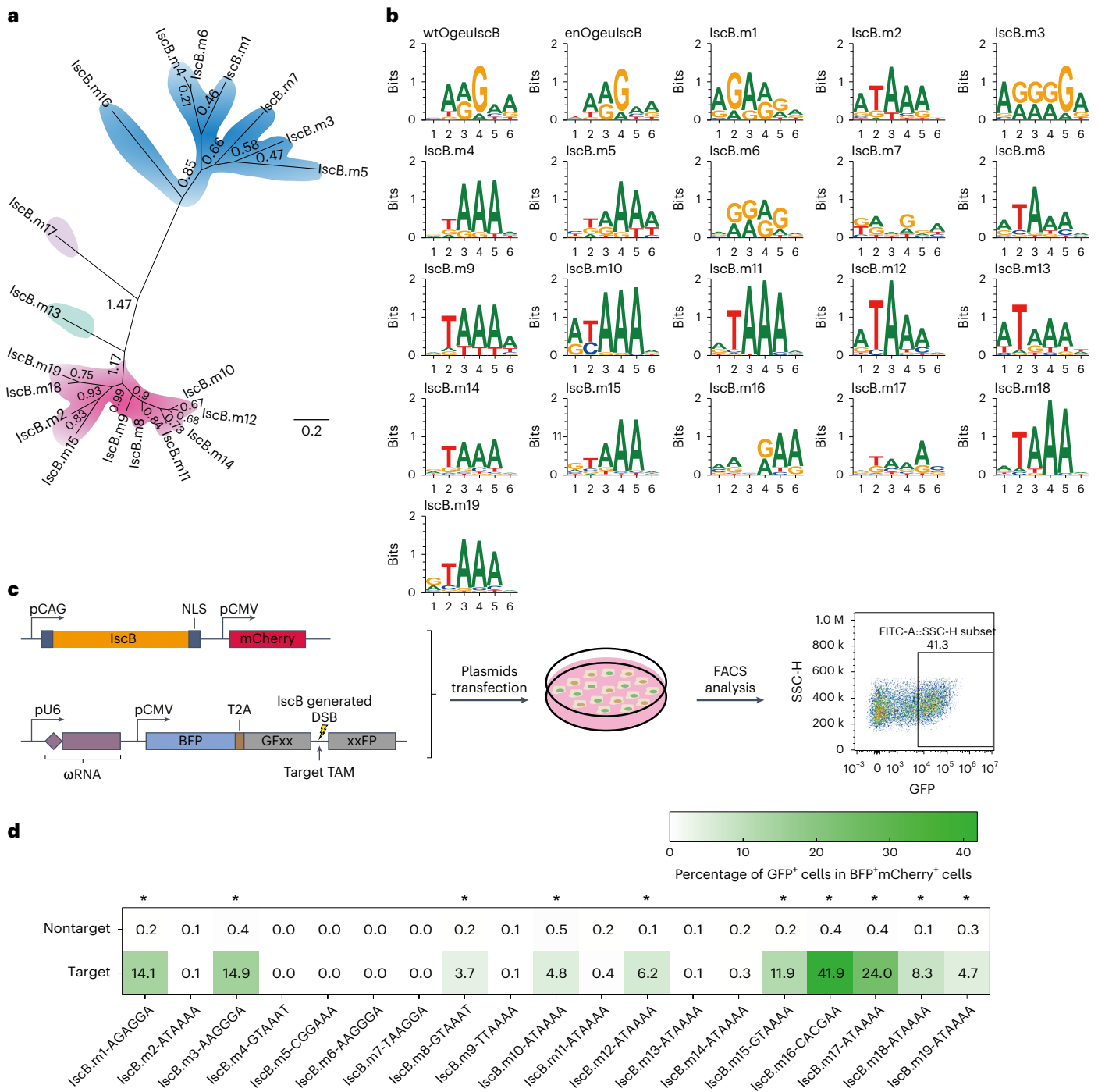


Fig. 1 | Identification and characterization of functional IscB orthologs. **a**, Phylogenetic tree of 19 uncharacterized IscB orthologs. **b**, TAMs of 19 active IscB proteins and two OgeulscB variants¹⁹ determined by bacterial depletion assay. **c**, Schematics describing the detection of editing activity based on the fluorescence signal of GFxxFP reporter activation in HEK293T cells. **d**, Fluorescence signal of

EGFP activated by IscB-mediated DSBs quantified by flow cytometry. Nontarget denotes a spacer with a random sequence. Asterisks denote a > 9-fold ratio of target and nontarget, with target recognition > 3.0%, representing variants with activity in HEK293T cells. Values represent the mean of three independent biological replicates.

and Extended Data Fig. 3c). To validate the enhanced activity of these four variants, we performed TAM recognition with 16 reporters including NAAGAA, ANAGAA, AANGAA, AAAGNA and AAAGAN. The results demonstrated that four variants exhibited higher EGFP fluorescence intensity relative to the WT, suggesting their superior editing activity (Extended Data Fig. 3d). On the basis of the results of the 16 reporters described above, we combined the E326R, P460S, T462H, T462L and T465V substitutions and obtained the best-performing

combination variant with E326R, P460S and T462H, named IscB.m16-RSH (Fig. 3c and Extended Data Fig. 3e). To test the TAM preference of IscB.m16-RSH, we detected EGFP activation using 64 TAM reporters with the 5'-NNNGAA-3' TAM and IscB.m16-RSH showed high editing activity for most TAMs but remained low for others (Extended Data Fig. 3f). Considering the characteristics of TAM recognition, we designed three additional TAM pools, pool 2 (TTTGAA, TTGGAA, TCAGAA, CTAGAA and CTGAA), pool 3 (GTAGAA, GTTGAA, GTCGAA,

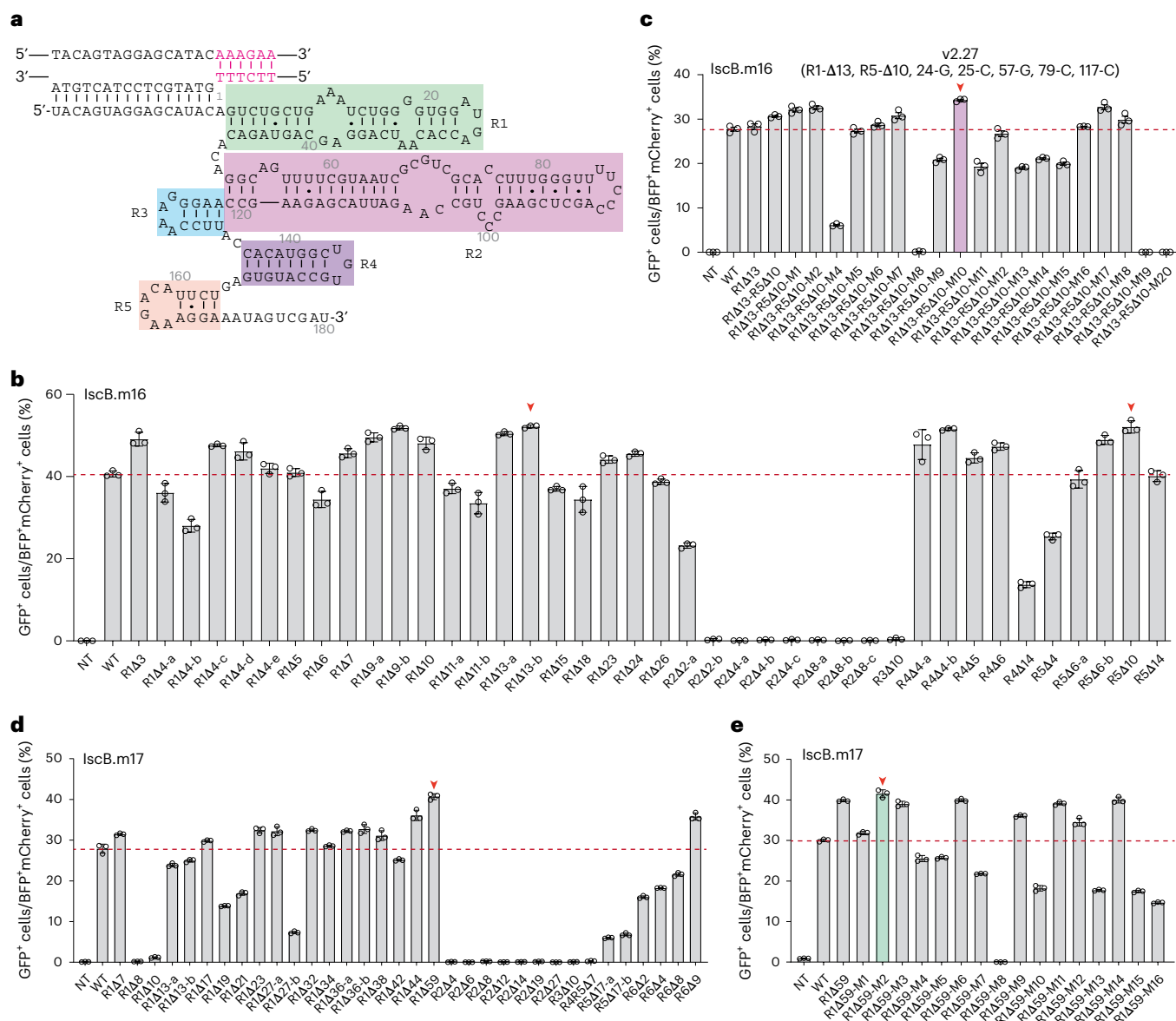


Fig. 2 | Engineering of various IscB ωRNAs to improve editing efficiency in mammalian cells. a, Secondary structure of IscB.m16 ωRNA predicted by RNAfold. Five regions are indicated as R1, R2, R3, R4 and R5. **b**, Increased EGFP signal induced by IscB.m16 caused by truncation of the stem loops in R1 (R1-Δ13b) and R5 (R5-Δ10) of the ωRNA. **c**, Substitutions of A•U to G•C based on the trimmed ωRNA (R1-Δ13b and R5-Δ10) enhanced the EGFP fluorescence signal mediated by IscB.m16. Here, v2.27 represents the IscB.m16-ωRNA variant with R1-Δ13, R5-Δ10,

24-G, 25-C, 57-G, 79-C and 117-C. **d, e**, ωRNA engineering for IscB.m17. Truncation of the stem loop and/or substitutions of A•U to G•C in the ωRNA improved the editing efficiency of IscB.m17. Nontarget (NT) denotes a spacer with a random sequence. Values represent the mean of three independent biological replicates. The red dashed lines represent the value of the WT. The red arrows represent the current optimal variants for each IscB ωRNA. Values and error bars represent the mean \pm s.d. ($n = 3$ independent biological replicates).

GTGGAA and GCAGAA) and pool 4 (ATAGAA, TGTGAA, CTCGAA and GAGGAA) as positive pools (Extended Data Fig. 3f). To further improve the activity at further TAMs, we selected sites that showed improved activity in TAM pool 1. We then separately combined the mutants at sites T459, N643, Q475, L481 or I504 with IscB.m16-RSH and evaluated the variants using reporters from pools 2 to 4 (Fig. 3d). We found that the combination of T459E with IscB.m16-RSH, named IscB.m16^{RESH}, exhibited increased editing efficiency of reporters from pool 2 relative to IscB.m16-RSH, with comparable editing efficiency of reporters from pools 3 and 4 (Fig. 3d). To assess the target range and editing activity of IscB.m16^{RESH}, we used 64 NNNGAA and 16 AAAGNN TAM reporters and found that IscB.m16^{RESH} exhibited significantly improved editing efficiency of these reporters compared to WT IscB.m16 (Supplementary Fig. 6).

To further optimize ωRNA-v2.27 (del15–20, del29–35, del171–180, 24-G, 25-C, 57-G, 79-C and 117-C) based on IscB.m16^{RESH}, we flipped the G•C base pairs of existing mutations (for example, converting 57G•117C to 57C•117G) or replaced the remaining mismatched G•U base pairs with G•C or C•G, guided by the ωRNA secondary structure. We found that v2.27-M21 (R1-Δ13, R5-Δ10, 24-G, 25-C, 57-C, 79-C, 117-C and 189-G) showed significantly enhanced editing efficiency, hereafter named enωRNA (Fig. 3e). Then, we examined the indel efficiency of IscB.m16^{RESH} with enωRNA at five endogenous loci in cultured HEK293T cells and found that enωRNA-IscB.m16^{RESH} (named IscB.m16*) showed the highest activity and the broadest range of deletion (Fig. 3f and Extended Data Fig. 4a). We also explored a range of spacer lengths for IscB.m16* using fluorescence reporters at two different targets and two endogenous loci in the human genome and found that IscB.m16* exhibited the

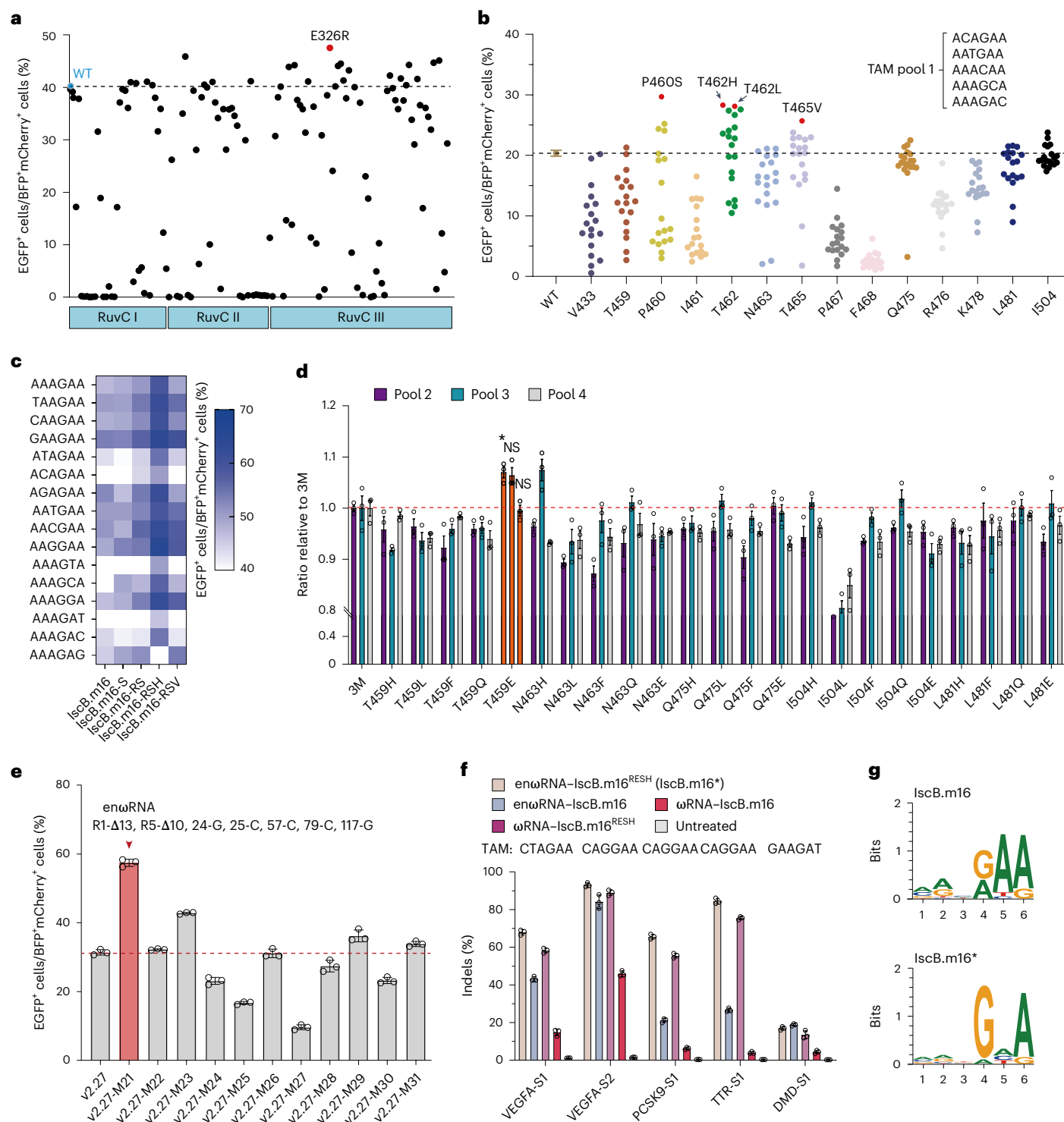


Fig. 3 | Protein engineering of IscB.m16 to improve editing efficiency and expand TAM range in mammalian cells. a, Screening for highly efficient variants by substitutions of amino acid residues in the RuvC domain of IscB.m16 protein with arginine. Each dot represents the editing activity for a single variant. The dashed line indicates the editing activity of the WT. **b**, Screening for highly efficient variants with saturation mutagenesis at selected sites using GfxxFP reporters containing a pooled TAM target. Each dot represents the editing activity for a single variant. The dashed line indicates the editing activity of the WT. **c**, Comparison of editing activity among WT IscB.m16 and its variants at 16 GfxxFP reporters with different TAMs. IscB.m16-S represents the variant P460S, IscB.m16-RS represents the variant with a combination of E326R and P460S, IscB.m16-RSH represents the variant with a combination of E326R, P460S and T462H and IscB.m16-RSV represents the variant with a combination

of E326R, P460S and T465V. Colored dots reflect the mean of three independent biological replicates. **d**, Screening for variants with improved editing frequency based on GfxxFP reporters containing three different TAM pools along with the same ωRNA-v2.27. The orange bar represents the IscB.m16^{RESH} variant with a combination of IscB.m16-RSH and T495E. *P* values were determined by Tukey's multiple comparisons test following ordinary one-way ANOVA. **P* < 0.05. NS, not significant. **e**, The second round of ωRNA engineering by substituting C-G base pairs on ωRNA-v2.27 (R1-Δ13, R5-Δ10, 24-G, 25-C, 57-G, 79-C and 117-C) based on IscB.m16^{RESH}. **f**, Comparison of indel frequency of WT IscB.m16 and its variants at five endogenous sites in HEK293T cells. **g**, TAM logos of IscB.m16 and IscB.m16* systems. Values and error bars represent the mean ± s.d. (*n* = 3 independent biological replicates).

highest activation with a guide spacer length of 14–21 nt (Extended Data Fig. 4b,c). Furthermore, TAM identification of IscB.m16* using bacterial depletion indicated that IscB.m16* recognized a 5'-NNNGNA-3' TAM, while IscB.m16 WT recognized a 5'-MRNRRA-3' TAM (Fig. 3g). To investigate off-target activity, we performed primer-extension-mediated sequencing (PEM-seq) experiments on IscB.m16*, enOgeulscB and SpG Cas9. IscB.m16* showed similar translocation events to enOgeulscB and SpG at the vascular endothelial growth factor A (VEGFA)-S6 site (Supplementary Fig. 7). Together, these results demonstrate that IscB.m16* exhibits high editing efficiency with highly flexible 5'-NNNGNA-3' TAM recognition.

IscB.m16*-mediated base editing in mammalian cells

Using prior information of the catalytic residues for IscB¹⁹ or SpCas9 (refs. 5,6), we constructed the inactive mutant D61A in the RuvC-I domain, H248A in the HNH domain and D61A;H248A on the basis of IscB.m16 and IscB.m16*. We tested nickase activity using the dual target reporter according to a previous study¹⁹. Consistently, IscB.m16^{D61A} showed the highest nickase activity and IscB.m16*-D61A;H248A showed no activity (Supplementary Fig. 8). In view of the compact size of IscB (Fig. 4a), we next fused IscB.m16^{D61A} with TadA8e-V106W to generate IscB.m16*-ABE (adenosine base editor) or with human APOBEC3A-W104A to generate IscB.m16*-CBE (cytosine base editor)^{24,25}.

To comprehensively evaluate the editing performance of IscB.m16*-ABE, we designed dozens of TAM-matched and protospacer-adjacent motif (PAM)-matched endogenous loci for IscB.m16-ABE, IscB.m16*-ABE, enOgeulscB-ABE¹⁹ and SpG-ABE²² (Fig. 4b). We found that the editing window of IscB.m16*-ABE ranged from positions 1 to 10 (counting the TAM as positions 15–20), while the optimal editing window occurred within positions 2–5 (Fig. 4c). At these matched G-containing TAM and PAM sites in HEK293T cells, IscB.m16*-ABE showed significantly higher A-to-G base-editing efficiency ($46.15\% \pm 4.08\%$) than IscB.m16-ABE ($9.19\% \pm 2.34\%$) and enOgeulscB-ABE ($31.34\% \pm 4.90\%$) and comparable base-editing efficiency to SpG-ABE ($50.77\% \pm 4.13\%$) (Fig. 4d, Extended Data Fig. 5 and Supplementary Fig. 9). In addition, the indel activity of IscB.m16*-ABE was similar to that of enOgeulscB-ABE but lower than that of SpG-ABE (Supplementary Fig. 10a,b). To characterize the TAM compatibility of IscB.m16*-ABE, we further analyzed the base-editing results and found that it showed A-to-G base editing at all TAM sites, while enOgeulscB-ABE showed no activity at some TAM sites such as N3GCA, N3GGA and N3GTA (Fig. 4e and Supplementary Fig. 9). Among the 33 designed TAM sequences, 19 TAM sequences were NWRGNA, which conformed to the canonical TAM and PAM sequences for each nuclease (IscB.m16*, NNNGNA; enOgeulscB, NWRRNA; SpG, NGN). For the 19 NWRGNA TAM sites, IscB.m16*-ABE, enOgeulscB-ABE and SpG-ABE exhibited comparable A-to-G efficiency (Extended Data Fig. 6a). For the 14 non-NWRGNA TAM sequences, IscB.m16*-ABE showed significantly higher A-to-G base editing than enOgeulscB-ABE and comparable base editing to SpG-ABE (Extended Data Fig. 6b). At some sites such as EMX1-S1 (GAAGAA) and VEGFA-S4 (AAAGCA), enOgeulscB-ABE exhibited higher editing efficiency than IscB.m16*-ABE and SpG-ABE (Extended Data Fig. 5 and Supplementary Fig. 9). This result also indicated the different preferred recognition TAM of each nuclease. To further evaluate the specificity of IscB.m16*-ABE in HEK293T cells, we conducted gRNA-dependent off-target DNA editing at predictive sites using Cas-OFFinder²⁶ and gRNA-independent off-target DNA editing using the orthogonal R-loop assay²⁷ at the ALDH1A3-S1, VEGFA-S1 and EMX1-S2 target sites. Targeted deep sequencing analysis revealed that IscB.m16*-ABE exhibited similar gRNA-dependent off-target effects to enOgeulscB-ABE and SpG-ABE at predicted off-target sites (Extended Data Fig. 7 and Supplementary Fig. 11). Using five previously reported SaCas9 target sites, we observed that IscB.m16*-ABE showed comparable low gRNA-independent off-target events to enOgeulscB-ABE and SpG-ABE (Extended Data Fig. 8). In addition,

IscB.m16*-CBE exhibited comparable base-editing activity and indels to enOgeulscB-CBE and SpG-CBE, with base-editing efficiencies of $60.01\% \pm 8.08\%$, $63.72\% \pm 5.33\%$ and $75.42\% \pm 8.12\%$, respectively (Fig. 4f, Extended Data Fig. 6c and Supplementary Fig. 10c,d). In addition, we detected IscB.m16*-ABE, enOgeulscB-ABE and SpG-ABE at five endogenous sites (ALDH1A3-S1, EMX1-S1, EMX1-S2, PCSK9-S1 and VEGFA-S5) in the U-2OS and HeLa cell lines. Consistent with its editing efficiency in the HEK293T cell line, IscB.m16*-ABE exhibited high editing efficiency in the U-2OS and HeLa cell lines (Extended Data Fig. 9). Collectively, these results indicate that IscB*-based base editors exhibit highly active editing, a broad target range and low off-target effects in mammalian cells.

Considering the bystander editing of base editors, we used high-fidelity TadA8e variants and different linkers combined with IscB to narrow the editing window (Supplementary Fig. 12). To address this concept, we tested high-fidelity deaminases and different linkers. We replaced TadA8e-V106W of IscB.m16*-ABE with TadA8e-N108Q or TadA8e-N108Q;L145T (ref. 28) and found a narrower editing window but lower editing efficiency compared to IscB.m16*-ABE (Extended Data Fig. 10a). The replacement of linkers between IscB and TadA8e-V106W showed no significant improvement with respect to narrowing the editing window (Extended Data Fig. 10b).

IscB-derived CBE restores dystrophin expression in mice

Taking advantage of its small size, the IscB*-derived base editor can be packaged with its ω RNA into a single rAAV vector, making it a greatly promising candidate for the treatment of certain genetic diseases, such as Duchenne muscular dystrophy (DMD)^{29,30}. Previous studies have shown that exon 50 skipping of the dystrophin gene can restore dystrophin expression in a mouse model with an exon 51 deletion, a mutation occurring in nearly 8% of patients with DMD^{31,32}. To access IscB.m16*-based base editing in DMD therapy, we devised a strategy whereby IscB.m16*-CBE disrupted the splicing signal by converting the G (in the paired chain of C) within the splicing acceptor site ('AG') to other bases (A, C or T), resulting in exon skipping (Fig. 5a). We first tested IscB.m16*-CBE with the ω RNA targeting the AG site adjacent to exon 50 in HEK293T cells. We observed that IscB.m16*-CBE displayed approximately 25% activity at position 10, which is the splicing acceptor site, while enOgeulscB-CBE and SpG-CBE showed almost no base-editing activity (Fig. 5b). To conveniently package tools into a single AAV, we chose IscB.m16*-CBE (4.0 kb) without the uracil DNA glycosylase inhibitor (UGI) domain (two UGI domains and a linker, 196 aa) to be packaged into AAV9 and detected the base-editing activity in mice. Two versions of IscB.m16*-CBE carrying different nuclear localization signals (NLSs) were designed and delivered to the muscle of mice with humanized exon 50 knock-in and exon 51 deletion (Fig. 5c). Then, 4 weeks after injection, we performed an editing efficiency evaluation, western blot analysis and histological staining for dystrophin expression. Targeted deep sequencing analysis showed that IscB.m16*-CBE-v2 achieved an approximate 7% conversion of G-to-H (G-to-A, G-to-T and G-to-C) and up to 30% level of exon 51 skipping (Fig. 5d,e). Western blotting and histological staining quantitative analysis of the tibialis anterior (TA) muscle and immunostaining results indicated that IscB.m16*-CBE-v2 restored the dystrophin protein levels in myofibers to 40% of the WT control (Fig. 5f–h). Together, these results indicate that the IscB.m16*-derived base editor, as a highly effective and broad-TAM miniature base-editing tool, provides a promising approach for basic research and therapeutic applications.

Discussion

In summary, through computational mining of metagenomic sequence datasets, we identified 19 natural IscB orthologs with various TAM recognition sites and 10 of the IscBs showed activity in mammalian cells, highlighting the diversity of the IscB family. By examining the

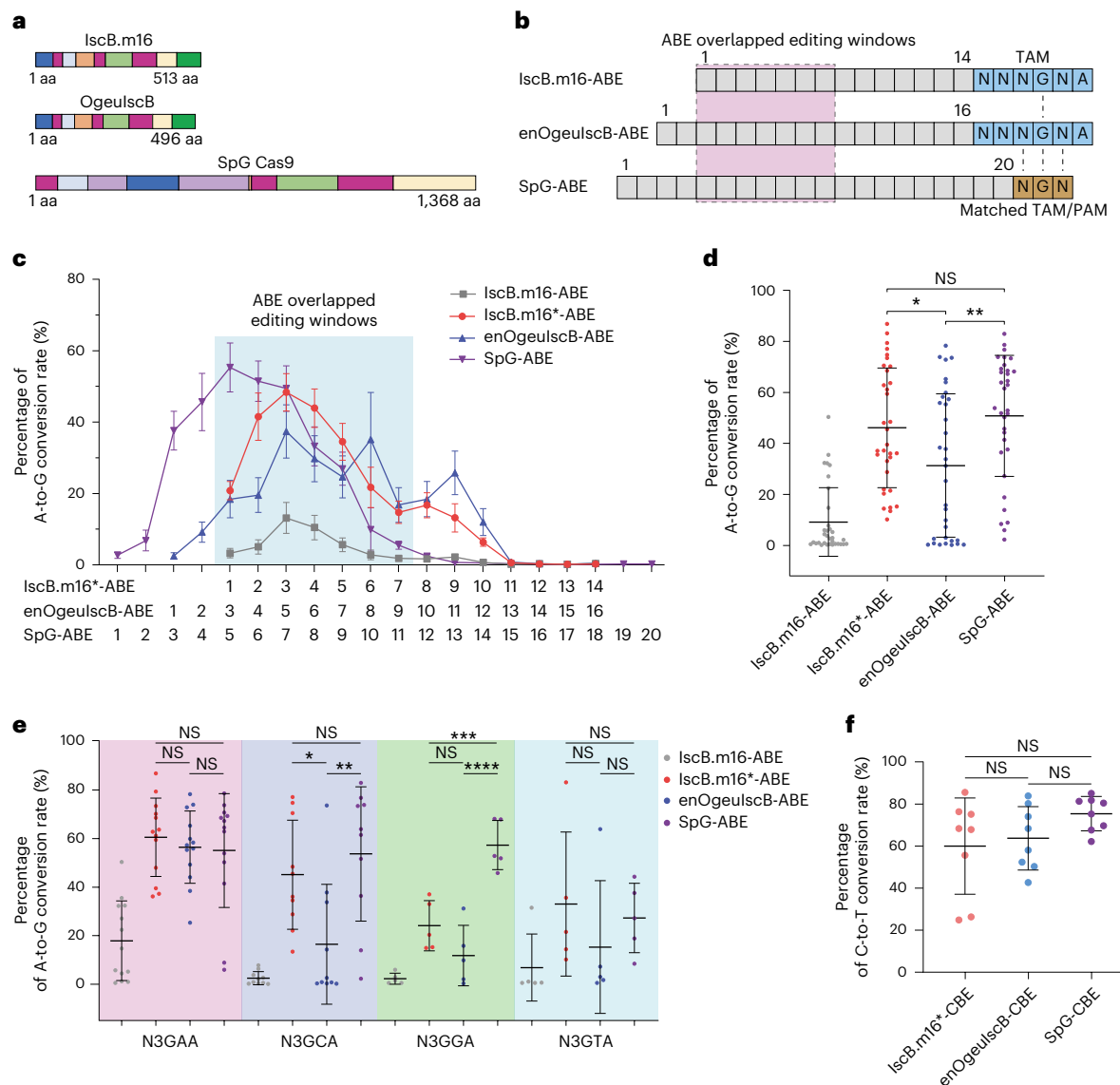


Fig. 4 | Characterization of IscB-derived and SpG-derived base editors in mammalian cells. **a**, Schematic of IscB, OgeulscB and SpG with different sizes. **b**, Overview of TAM-matched and PAM-matched sites used to compare IscB.m16-derived ABE to enOgeulscB-ABE and SpG-ABE. **c**, Editing window and base-editing activity of IscB.m16-ABE, IscB.m16*-ABE, enOgeulscB-ABE and SpG-ABE at all protospacer positions. Data are presented as the mean \pm s.e.m. Values are average editing efficiencies at each position of A within the target of three independent biological replicates from 33 endogenous sites. **d**, Comparison of the A-to-G conversion efficiency of IscB.m16-ABE, IscB.m16*-ABE, enOgeulscB-ABE and SpG-ABE at 33 endogenous loci. Data were collected from 33 endogenous sites and are presented as the mean \pm s.d. Each dot represents the average highest base-editing activity at each endogenous target site of three independent biological replicates. Adjusted P (P_{adj}) values are 0.8453, 0.046 and 0.0041, respectively. **e**, Comparison of the A-to-G conversion efficiency of IscB.m16-ABE, IscB.m16*-ABE, enOgeulscB-ABE and SpG-ABE

grouped by TAM at 33 target sites. Data are presented as the mean \pm s.d. The number of values from left to right is 13, 10, 5 and 5, respectively, and the values represent optimal editing efficiencies within the target as the mean of three independent biological replicates from endogenous sites. The P_{adj} values of N3GAA sites are 0.8698, 0.9404 and 0.9974, respectively. The P_{adj} values of N3GCA sites are 0.8154, 0.0271 and 0.0027, respectively. The P_{adj} values of N3GGA sites are 0.0003, 0.2172 and 0.000007, respectively. The P_{adj} values of N3GTA sites are 0.9744, 0.6102 and 0.8342, respectively. **f**, Comparison of the C-to-T conversion efficiency of IscB.m16*-CBE, enOgeulscB-CBE and SpG-CBE at eight target sites. Data were collected from eight endogenous sites and are exhibited as the mean \pm s.d. Each dot represents the average highest base-editing activity within the target at each endogenous target site of three independent biological replicates. All P values were determined by Tukey's multiple comparisons test following ordinary ANOVA. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$. NS, not significant.

results of six engineered ω RNAs, we found that the truncation of the first (R1) and last (R5 or R6) stem loops of the ω RNA usually enhanced the editing activity of IscBs. Through structure-guided design and protein engineering of the PID, TID and RuvC domain of IscB, we developed the IscB.m16* system that exhibited improved editing activity and extended the TAM scope to 5'-NNNGNA-3'. This is a notably broader recognition range than the previously reported enOgeulscB with a 5'-NWRRNA-3' TAM¹⁹, although we found that enOgeulscB showed

efficient activity with a broader TAM (not only NWRRNA) in mammalian cells. Furthermore, IscB.m16*-derived base editors showed editing activity comparable to SpG-BE and even higher editing activity than SpG-BE and enOgeulscB-BE at some disease-related loci, such as DMD. Therefore, considering their compact size and extended editing scope, IscB.m16*-derived base editors have the potential to be alternatives to enOgeulscB-derived and Cas9-derived base editors for AAV-based therapeutic applications.

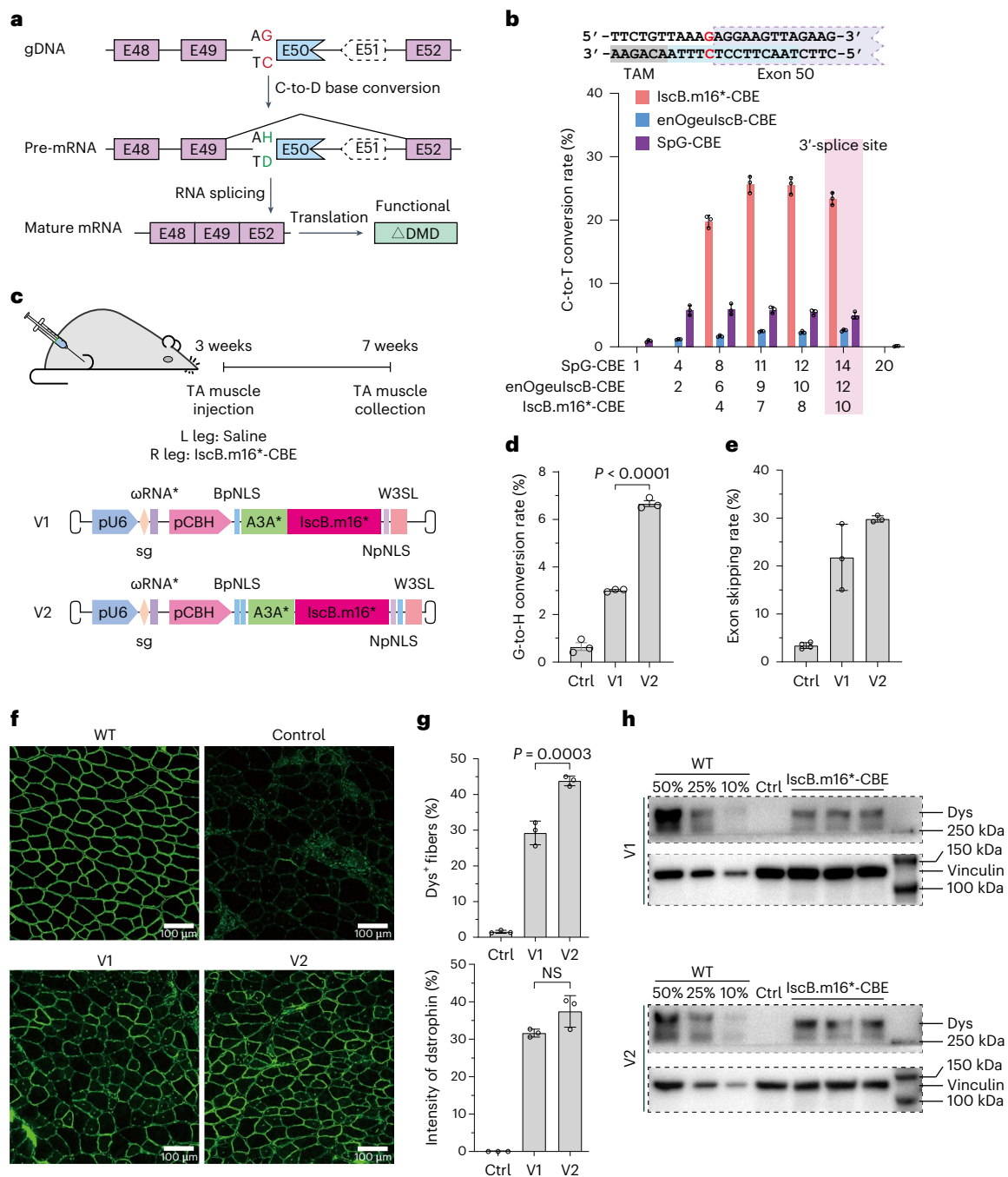


Fig. 5 | IscB.m16*-based cytosine base editor mediates effective base editing and restores dystrophin expression in humanized DMD^{E51del} mice.

a, Schematic of the strategy of IscB.m16*-CBE DMD treatment. IscB.m16*-CBE disrupts the conserved guanine within the splice acceptor site for programmable exon 50 skipping, leading to the restoration of dystrophin expression. **b**, The C-to-T conversion efficiency of IscB.m16*-CBE, enOgeulscB-CBE and SpG-CBE at the splice acceptor site of the DMD intron between exon 49 and exon 50 in HEK293T cells. Data are shown as the mean \pm s.d. ($n = 3$ independent biological replicates). **c**, Schematics of single AAV9 carrying two versions of IscB.m16*-CBE delivered to the muscles in mice by TA muscle injection. Saline was injected in the left leg, while AAV9 cargo IscB.m16*-CBE was injected in the right.

d, e, The in vivo G-to-H (C-to-D, including C-to-T, C-to-A and C-to-G) editing efficiencies (**d**) and RNA level of exon 50 skipping (**e**) of AAV9-IscB.m16*-CBE were detected by targeted deep sequencing. Data are presented as the mean \pm s.d. ($n = 3$ independent biological replicates; $n = 4$ for the control in **e**). The P_{adj} value is 0.000002 in **d, f**. Dystrophin immunohistochemistry showing the restoration of dystrophin expression 4 weeks after TA injection of IscB.m16*-CBE. Dystrophin is shown in green. Scale bars, 100 μ m. **g**, Quantification of Dys⁺ fibers and dystrophin in cross sections of TA muscles from **f**. Data are presented as the mean \pm s.d. ($n = 3$ independent biological replicates). **h**, Western blot analysis of dystrophin and vinculin expression in TA muscles 4 weeks after injection with AAV9-IscB.m16*-CBE or saline.

Additionally, we found that both IscB.m16* and enOgeulscB showed indel activity with a guide containing a spacer length of 14–21 nt (Extended Data Fig. 4c). This is similar to a previous study showing that OgeulscB exhibited indel activity with a guide containing a spacer length of 14–26 nt¹⁷. Thus, most IscB- ω RNA systems have less

stringent requirements with regard to spacer length. Given the potential for off-target effects with short spacer lengths, screening variants to specifically bind long spacers (for example, 20 nt), increasing the mismatch of spacers to extend the spacer length or designing a specific stable lock-and-key structure³³ may minimize the off-target effects.

Considering the 14-nt spacer length, a 2-nt TAM may achieve the best balance between targeting range and specificity. Therefore, developing more highly active IscB orthologs with different TAM recognition sites may be more important for future applications than developing a near-TAMless IscB. Given the experience gained in engineering the IscB nuclease and ω RNA, the activity and TAM editing range of other IscB orthologs such as OgeulIscB could be improved and expanded using similar strategies. Together with IscB.m16*, a set of engineered IscB orthologs may constitute a miniature genome base-editing toolbox.

Overall, the engineered compact IscB-derived base editors were proven to be a platform with highly efficient, specific and broad-TAM-scope DNA base editing in mammalian cells and in mouse models of diseases, highlighting their potential in gene therapy.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-024-01706-1>.

References

- Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
- Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
- 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).
- 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).
- 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).
- Chen, S. Y. et al. Compact Cje3Cas9 for efficient genome editing and adenine base editing. *CRISPR J.* **5**, 472–486 (2022).
- Davis, J. R. et al. Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. *Nat. Biomed. Eng.* **6**, 1317 (2022).
- Zhang, H. et al. Adenine base editing in vivo with a single adeno-associated virus vector. *GEN Biotechnol.* **1**, 285–299 (2022).
- Wu, Z. et al. Programmed genome editing by a miniature CRISPR–Cas12f nuclease. *Nat. Chem. Biol.* **17**, 1132–1138 (2021).
- Kim, D. Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2021).
- Kong, X. et al. Engineered CRISPR–OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. *Nat. Commun.* **14**, 2046 (2023).
- Hino, T. et al. An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* **186**, 4920–4935 (2023).
- Xu, X. et al. Engineered miniature CRISPR–Cas system for mammalian genome regulation and editing. *Mol. Cell* **81**, 4333–4345 (2021).
- Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
- Kim, D. Y. et al. Hypercompact adenine base editors based on a Cas12f variant guided by engineered RNA. *Nat. Chem. Biol.* **18**, 1005–1013 (2022).
- Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
- Schuler, G., Hu, C. Y. & Ke, A. L. Structural basis for RNA-guided DNA cleavage by IscB- ω RNA and mechanistic comparison with Cas9. *Science* **376**, 1476–1481 (2022).
- Han, D. et al. Development of miniature base editors using engineered IscB nickase. *Nat. Methods* **20**, 1029–1036 (2023).
- Stewart, R. D. et al. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953–961 (2019).
- Zhang, H. et al. An engineered xCas12i with high activity, high specificity and broad PAM range. *Protein Cell* **14**, 538–543 (2023).
- 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).
- Nishimasu, H. et al. Engineered CRISPR–Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
- Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
- Wang, X. et al. Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* **31**, 107723 (2020).
- 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).
- 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).
- Chen, L. et al. Engineering a precise adenine base editor with minimal bystander editing. *Nat. Chem. Biol.* **19**, 101–110 (2022).
- Min, Y. L., Bassel-Duby, R. & Olson, E. N. CRISPR correction of Duchenne muscular dystrophy. *Annu. Rev. Med.* **70**, 239–255 (2019).
- Olson, E. N. Toward the correction of muscular dystrophy by gene editing. *Proc. Natl Acad. Sci. USA* **118**, e2004840117 (2021).
- Chemello, F. et al. Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Sci. Adv.* **7**, eabg4910 (2021).
- Yuan, J. J. et al. Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell* **72**, 380–394 (2018).
- Wang, L. et al. Eliminating base-editor-induced genome-wide and transcriptome-wide off-target mutations. *Nat. Cell Biol.* **23**, 552–563 (2021).

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Methods

Computational analysis of IscB systems

More than 200 Gb of metagenome assemblies were downloaded from the European Nucleotide Archive database (accession number [PRJEB31266](https://doi.org/10.1038/s41589-024-01706-1)). Firstly, we used TBLASTN and the OgeulscB protein to identify IscB-containing sequences of metagenomes with an E value $< 1 \times 10^{-50}$ (ref. 20). Then, Prodigal was used to annotate the proteins of the IscB-containing sequences³⁴. We further searched for previously trained ω RNA models to annotate the ω RNA sequences with an E value $< 1 \times 10^{-10}$. RNAfold was used to predict the secondary structure of the ω RNA^{35,36}. MEGAX was used to construct the phylogenetic tree³⁷. All newly IscB protein sequences and ω RNA scaffolds (DNA sequences) are provided in Supplementary Table 1.

Plasmid construction

All *E. coli* codon-optimized IscB-encoding genes and their associated ω RNA scaffolds were synthesized by Shanghai Huagene Biotechnology and assembled into a pUC19-derived vector (EcoNI + XbaI) under the *lac* and J23119 promoters using a 2 \times pEASY basic seamless cloning and assembly kit (TransGen Biotech). All human codon-optimized IscB-encoding sequences were synthesized by GenScript and incorporated into a mammalian expression vector under the CBh promoter. For endogenous genome-editing experiments in HEK293T cells, the gRNA oligos were synthesized and cloned into a Bpil-digested backbone of the U6 promoter using T4 ligase (Thermo Fisher Scientific). All colonies were sequence-verified from promoter to poly(A) using Sanger sequencing (Genewiz). Information on all IscB expression sequences is provided in Supplementary Tables 2–4.

Generation of the TAM library and TAM depletion assay

A randomized TAM library containing a target sequence followed by eight randomized bases downstream was constructed. The synthesized single-stranded DNA (ssDNA) (HuaGene) was converted into double-stranded DNA (dsDNA) by annealing with a short ssDNA and second-strand synthesis using the Large (Klenow) fragment (New England Biolabs). The resulting dsDNA was then assembled into pACYC184 vectors using Gibson assembly (New England Biolabs). The products were purified using isopropanol, electroporated into TransforMax EC100 electrocompetent *E. coli* according to the manufacturer's instructions and plated on chloramphenicol plates. After 13 h of growth at 37 °C, *E. coli* cells were scraped from the plates and extracted using a NucleoBond Xtra Midiprep kit (Machery Nagel).

For the bacterial TAM depletion assay, we cotransformed 200 ng of TAM library plasmids and 300 ng of plasmids expressing *E. coli* codon-optimized IscB and ω RNA into TransforMax EC100 electrocompetent *E. coli* cells by electroporation. Then, the transformed cells were recovered for 1 h at 37 °C with antibiotic-free medium and plated on 250 mm \times 250 mm carbenicillin and chloramphenicol plates. After 13 h of growth, cells were harvested and plasmid DNA was extracted using a NucleoBond Xtra Midiprep kit (Machery Nagel). The TAM-containing region was amplified by Phanta Max super-fidelity DNA polymerase (Vazyme Biotech) for 12 cycles and Illumina adaptors and unique barcodes were added by a second round of PCR for 18 cycles. The resulting PCR products were purified with a gel extraction kit (Omega) and sequenced by a Illumina NovaSeq 6000 platform with 150-bp paired-end reads (Genewiz).

TAM regions were extracted, counted and then normalized to the total TAM counts for each sample. For each specific TAM, TAMs that appeared more than once were filtered and the log fold change (logFC) of its frequency was measured as the log ratio compared to nontarget control. Depletions with a logFC $< -3\sigma$ (s.d.) were considered statistically significant. A position weight matrix (PWM) was built from all significantly depleted sequences, with $-\log$ FC values serving as the

corresponding weight. A sequence logo was generated on the basis of this PWM using WebLogo (version 3.7.12)¹⁷.

Cell culture and transfection

HEK293T cells were cultivated in DMEM (Sigma) supplemented with 10% FBS (Gibco), 1% penicillin–streptomycin–glutamine (Gibco) and 1% minimum essential medium nonessential amino acids (Gibco) in a humidified incubator at 37 °C with 5% CO₂. For the detection of IscB nuclease activities and screening of its variants, HEK293T cells were seeded in 24-well plates with 70–80% confluence. After a 12-h incubation, 1.6 μ g of plasmids were cotransfected into HEK293T cells using polyethylenimine (PEI) following the manufacturer's manual. The plasmids included those encoding the BFP–T2A–GFxxFP and IscB systems, with a molar ratio of 1:1. For genome or base editing at endogenous loci, 1.6 μ g of all-in-one plasmids were transfected to express gRNA and the nuclease-editing or base-editing system. After 48 h, cells were sorted by fluorescence-activated cell sorting (FACS) analysis.

FACS analysis

Before FACS analysis, cells were subjected to treatment with 0.25% trypsin-EDTA (Gibco) for dissociation and suspended in FBS-containing DMEM. For the assessment of IscB nuclease activity and screening of variants using the fluorescence reporter system, cells were analyzed for EGFP, mCherry and BFP fluorescence. A total of 25,000 single cells were recorded to analyze efficiency using a Beckman CytoFlex flow cytometer 48 h after transfection. Data analysis was performed by FlowJo X (version 10.0.7). For genome-editing analysis, approximately 15,000 transfection-positive cells (defined as those with a fluorescence intensity $\geq 10^3$ among fluorescence-positive cells) were sorted 48 h after transfection using a BD FACS Aria III flow cytometer. Following FACS sorting, genomic DNA from the collected cells was extracted by cell lysis with 25 μ l of proteinase K-added lysis buffer (Vazyme Biotech) per sample, as described previously. The cell lysates were stored at -20 °C until further use.

Targeted deep sequencing and analysis

To detect the editing efficiency at endogenous loci, the target genome regions of interest were amplified from cell lysates by PCR using Phanta Max super-fidelity DNA polymerase (Vazyme Biotech). For targeted deep sequencing analysis, PCR reactions were performed using primers with unique barcodes. The amplified products were purified using a gel extraction kit (Omega) and sequenced by an Illumina NovaSeq 6000 platform with 150-bp paired-end reads (Genewiz). The deep sequencing data were first demultiplexed by a custom script based on sample barcodes. The demultiplexed reads were then analyzed by CRISPResso2 (version 2.0.20b)³⁸ for the quantification of the editing efficiency, including indels and base conversions at each target locus. All targeting sites and primers used are provided in Supplementary Table 5.

PEM-seq assay

PEM-seq in HEK293T cells was performed as previously described^{39,40}. Specifically, all-in-one plasmids containing IscB.m16*, enOgeulscB and SpCas9-SpG with targeting VEGFA-S6 ω RNA were transfected into HEK293T cells using PEI; after 48 h, positive cells were harvested for DNA extraction. A total of 10 μ g of genomic DNA was fragmented with a peak length of 300–700 bp by Covaris sonication. Those DNA fragments were firstly tagged with biotin through a one-round biotinylated primer extension at the 5' end and then primers were removed by AMPure XP beads and purified by streptavidin beads. Then, the ssDNA attached to the streptavidin beads was ligated with a 14-bp random molecular barcode bridge adapter and a nested PCR was performed for enriching the DNA fragment containing the bait DSB and tagging the DNA fragment with Illumina adaptor sequences. The prepared sequencing library was subjected to high-throughput sequencing on a Hi-seq 2500 with 2 \times 150 bp reads.

gRNA-dependent off-target analysis

To examine the gRNA-dependent off-target effects of IscB.m16*-ABE, enOgeulscB-ABE and SpG-ABE, CRISPR RGEN Tools (Cas-OFFinder, <http://www.rgenome.net/cas-offinder/>) was used to predict potential off-target sites as described previously²⁶. For the ABE based on IscB.m16*, the search queries covered both the 14-nt target spacer sequences and 'NNNGNA'. The PAM of the search was set as 'NNN' and the number of mismatches was set to three. The search queries of enOgeulscB-ABE were set similarly but with a 16-nt spacer sequence and a 6-nt TAM sequence containing 'NWRRNA'. For the ABE based on SpG, search queries covered 20-nt target spacer sequences, the PAM type was set to 'NG' and the number of mismatches was set to four. All other parameters were default. Off-target sites for each gRNA in each group were manually selected in order of the number of mismatches from low to high. Sites with a 5'-NNNGNA-3' TAM were retained for IscB.m16*-ABE and sites with a 5'-NWRRNA-3' TAM were retained for enOgeulscB-ABE. All potential sites and primers are provided in Supplementary Tables 6–15.

Orthogonal R-loop assay

An orthogonal R-loop assay was performed to detect the gRNA-independent off-target editing as described previously²⁷. First, 0.8 µg of plasmids that encode IscB.m16*-ABE, enOgeulscB-ABE or SpG-ABE with their respective ωRNA or single-guide RNA (sgRNA) and 0.8 µg of dSaCas9 plasmids with their corresponding sgRNA targeting five previously reported R-loop sites were cotransfected into HEK293T cells using PEI. After a 48-h cultivation, transfected cells were analyzed by FACS followed by genomic DNA extraction with 25 µl of freshly prepared lysis buffer (Vazyme) containing proteinase K. Amplification and targeted deep sequencing were performed at the ABE target sites and dSaCas9 R-loop off-target sites. All targeting sequences and primers are provided in Supplementary Table 16.

Animals

All animal experiments in this study were performed following approved protocols and guidelines set by the Animal Care and Use Committee of HuidaGene Therapeutics. Mice were housed in a controlled barrier facility with a 12-h light–dark cycle at 18–23 °C with 40–60% humidity. Diet and water were accessible at all times. DMD^{AmE5051, K1hE50/Y} mice were generated in the C57BL/6J background using the CRISPR–Cas9 system. DMD is the most common sex-linked lethal disease in humans; thus, male mice were selected for this study.

Production and delivery of AAV9 to DMD^{AmE5051, K1hE50/Y} mice

AAVs were manufactured by HuidaGene Therapeutics. Briefly, cells were grown in culture until they reached a confluency of 70–90%. Before transfection, the growth medium was replaced with prewarmed growth medium. For each 15-cm dish, a mixture of 20 µg of pHelper, 10 µg of pRepCap and 10 µg of the gene-of-interest plasmid was prepared and added dropwise to the cell medium. After a 3-day incubation period, AAVs were harvested and purified using iodixanol density gradient centrifugation. For intramuscular injection, 3-week-old DMD^{AmE5051, K1hE50/Y} mice were anesthetized and their TA muscle was injected with either 30 µl of AAV9 (2.5 × 10¹¹ vg) preparations or an equivalent volume of saline solution. Tissue samples were collected for genomic DNA, RNA, immunoblotting and immunofluorescence analyses 4 weeks after treatment.

Western blot analysis

Tissue samples were homogenized using radioimmunoprecipitation assay buffer supplemented with a protease inhibitor cocktail. The supernatants of the lysates were quantified using a Pierce BCA protein assay kit (Thermo Fisher Scientific, 23225) and adjusted to a uniform concentration using H₂O. Equal volumes of the samples were mixed with NuPAGE LDS sample buffer (Invitrogen, NP0007) and 10%

β-mercaptoethanol and then subjected to boiling at 70 °C for 10 min. A total of 10 µg of protein per lane was loaded into 3–8% Tris-acetate gels (Invitrogen, EA03752BOX) and underwent electrophoresis for 1 h at 200 V. Proteins were then transferred onto a PVDF membrane under wet conditions at 350 mA for 3.5 h. The membrane was then blocked in 5% nonfat milk in TBST buffer and incubated with the primary antibody to mark the target protein. After three washes with TBST, the membrane was incubated with a horseradish peroxidase-conjugated secondary antibody specific to the IgG of the species from which the primary antibody against dystrophin (Sigma, D8168) or vinculin (CST, 13901S) was derived. The target proteins were visualized using chemiluminescent substrates (Invitrogen, WP20005).

Immunofluorescence

Tissues were encased in optimal cutting temperature compound and rapidly frozen in liquid nitrogen. Serial frozen cryosections, each measuring 10 µm in thickness, were fixed for 2 h at 37 °C, followed by permeabilization with PBS containing 0.4% Triton-X for 30 min. After washing with PBS, samples were blocked with 10% goat serum for 1 h at room temperature. Following this, the slides were incubated overnight at 4 °C with primary antibodies against dystrophin (Abcam, ab15277) and spectrin (Millipore, MAB1622). The following day, samples were thoroughly washed with PBS and incubated with compatible secondary antibodies (Alexa Fluor 488 AffiniPure donkey anti-rabbit IgG (Jackson ImmunoResearch labs, 711-545-152) or Alexa Fluor 647 AffiniPure donkey anti-mouse IgG (Jackson ImmunoResearch labs, 715-605-151)) and DAPI for 3 h at room temperature. After a 15-min PBS wash, slides were sealed with fluoromount-G mounting medium. All images were captured using a Nikon C2 camera. The number of Dys⁺ muscle fibers was represented as a percentage of the total spectrin-positive muscle fibers.

Statistical analysis

All values are shown as the mean ± s.d. except for values of the editing window from base editors, which were shown as the mean ± s.e.m. A one-way analysis of variance (ANOVA) was used for statistical comparisons and a *P* value < 0.05 was considered statistically significant. Details of statistical values are provided in corresponding figure legends. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. GraphPad Prism (version 8.2.1) was used for statistical analysis (www.graphpad.com/).

Reporting summary

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

Data availability

All sequencing data were deposited to the National Center for Biotechnology Information Sequence Read Archive under accession number [PRJNA1043847](https://www.ncbi.nlm.nih.gov/sra/PRJNA1043847). All materials are available upon reasonable request. Source data are provided with this paper.

References

- Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
- Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
- Lorenz, R. et al. ViennaRNA package 2.0. *Algorithm Mol. Biol.* **6**, 26 (2011).
- Kumar, S., Stecher, G., Li, M., Niyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
- Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

39. Yin, J. et al. Optimizing genome editing strategy by primer-extension-mediated sequencing. *Cell Discov.* **5**, 18 (2019).
40. Liu, Y. et al. PEM-seq comprehensively quantifies DNA repair outcomes during gene-editing and DSB repair. *STAR Protoc.* **3**, 101088 (2022).

Acknowledgements

This work was supported by HUIDAGENE Therapeutics Inc. We thank Y. Yuan for assistance with intellectual property. This work was supported by the National Natural Science Foundation of China (31925016, 82021001 and 82271048), National Science and Technology Innovation 2030 Major Program (2021ZD0200900), Basic Frontier Scientific Research Program of the Chinese Academy of Sciences From 0 to 1 Original Innovation Project (ZDBS-LY-SM001), Project of Shanghai Municipal Science and Technology Commission (23HC1401100), Science and Technology Commission of Shanghai Municipality (22S11900200 and 23XD1420500), Project of Shanghai Xuhui District Science and Technology (2020-015), Program for Professor of Special Appointment (Eastern Scholar, TP2022046) at the Shanghai Institutions of Higher Learning and Shanghai Sailing Program (21YF1453100).

Author contributions

Q.X. and H.Z. jointly conceptualized the project and designed the experiments. D.H. conducted the bacterial depletion assay and performed the orthogonal R-loop assay. Q.X., D.H., H.W. and G.L. performed the ω RNA and protein engineering and reporter assay. Q.X., M.Y. and T.M. conducted the experiments related to genome editing and base editing and the off-target assays. G.L. performed the animal

experiments. Q.X. performed the bioinformatics analysis. M.Y. and T.M. constructed plasmids and performed the FACS analysis. J.Z., Y.Z., X.Z., B.H., Y.Y., J.H. and L.S. assisted with experiments. J.H., H.Y. and H.Z. supervised the project. T.L., J.H., H.Y. and H.Z. wrote and revised the manuscript with data contributed by all authors who participated in the project.

Competing interests

H.Z. and Q.X. disclose a patent application (PCT/CN2024/071744) related to the proteins described in this manuscript. H.Y. and L.S. are cofounders of HUIDAGENE Therapeutics. The other authors declare no competing interests.

Additional information

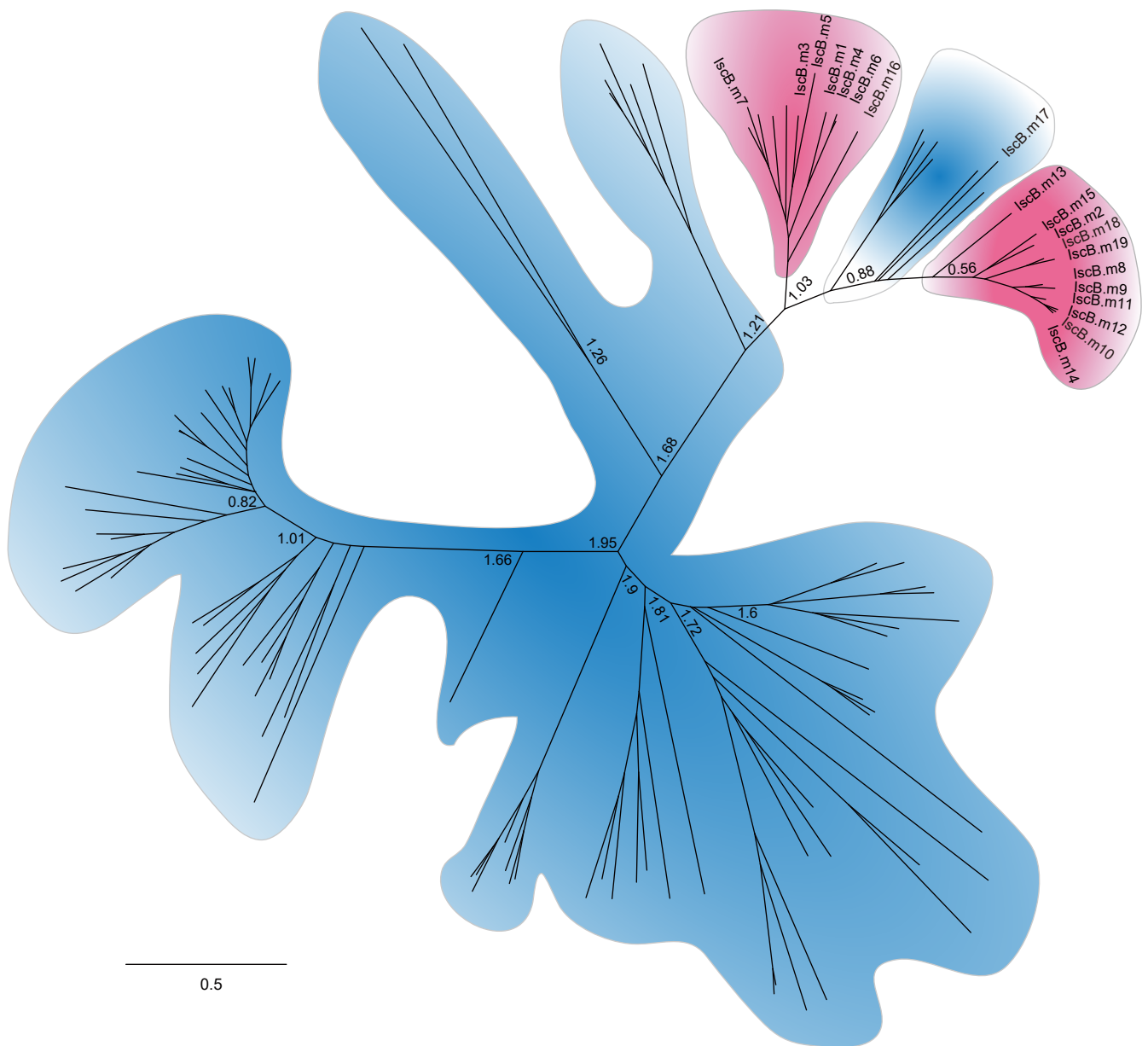
Extended data is available for this paper at <https://doi.org/10.1038/s41589-024-01706-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-024-01706-1>.

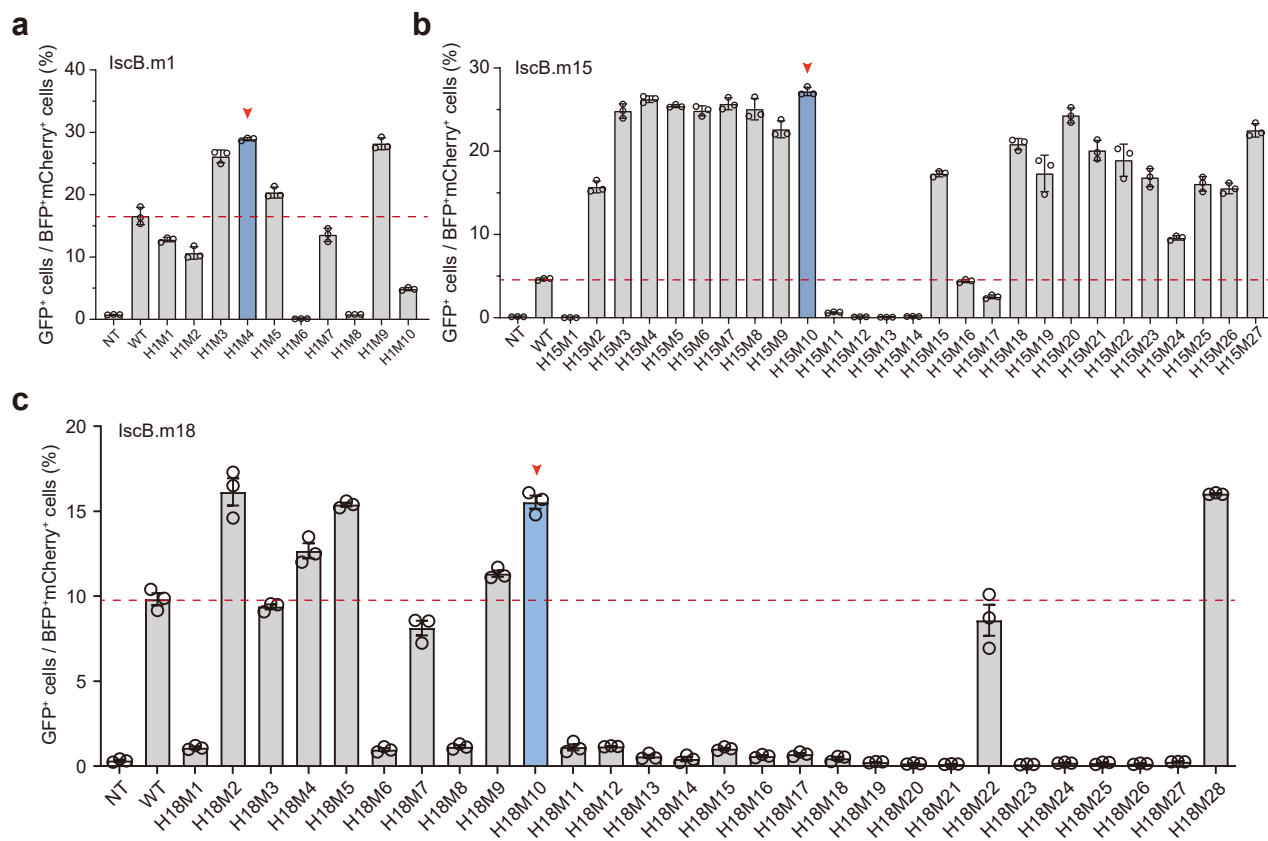
Correspondence and requests for materials should be addressed to Tong Li, Hui Yang, Jinhai Huang or Hainan Zhang.

Peer review information *Nature Chemical Biology* 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.

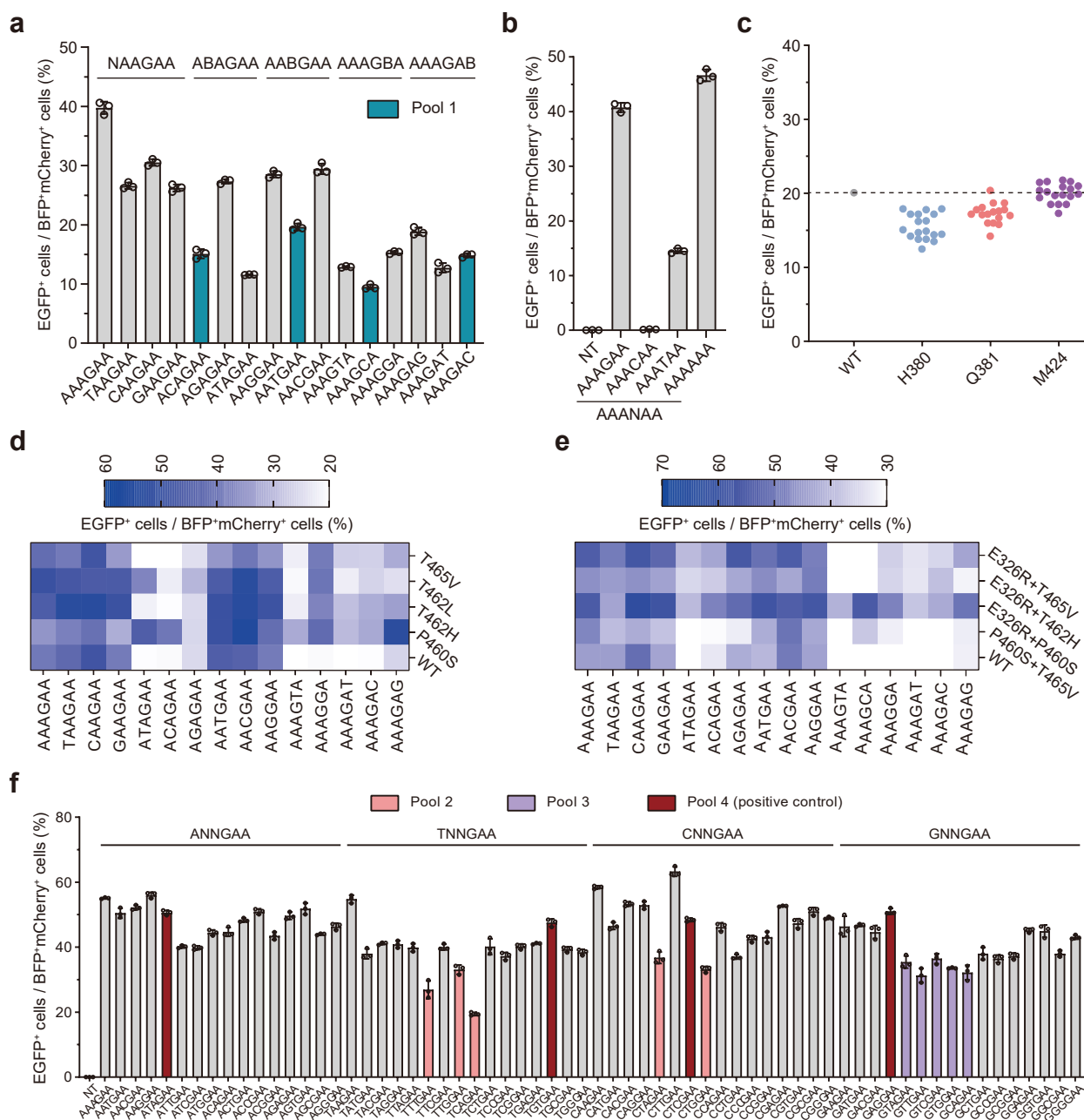


Extended Data Fig. 1 | Maximum-likelihood tree of identified IscB orthologs and previously reported IscBs. The name of 19 identified IscB orthologs are shown. The evolutionary distance scale of 0.5 is shown.



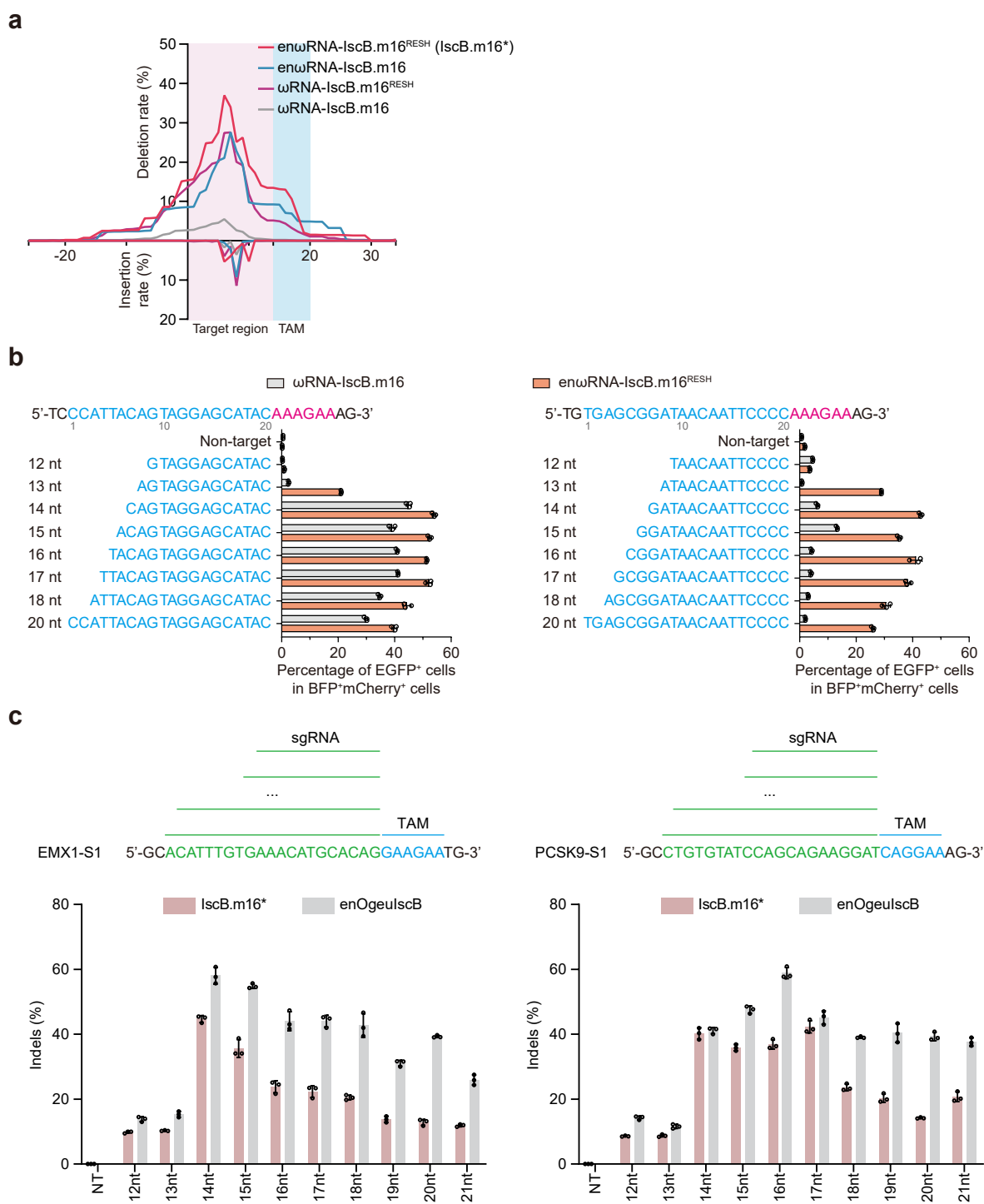
Extended Data Fig. 2 | Engineering various IscB ω RNAs to improve their editing efficiency in mammalian cells. a–c, IscB.m1 (a) improved EGFP signal by truncation of stem loop in R1 of ω RNA, and IscB.m15 (b) and IscB.m18 (c) improved EGFP signal by truncation of stem loop in R1 and R5 of ω RNA. The red

dashed lines represent the value of WT. The red arrows represent the current optimal variants for each IscB ω RNA. Data are shown as mean \pm s.d., n = 3 independent biological replicates.



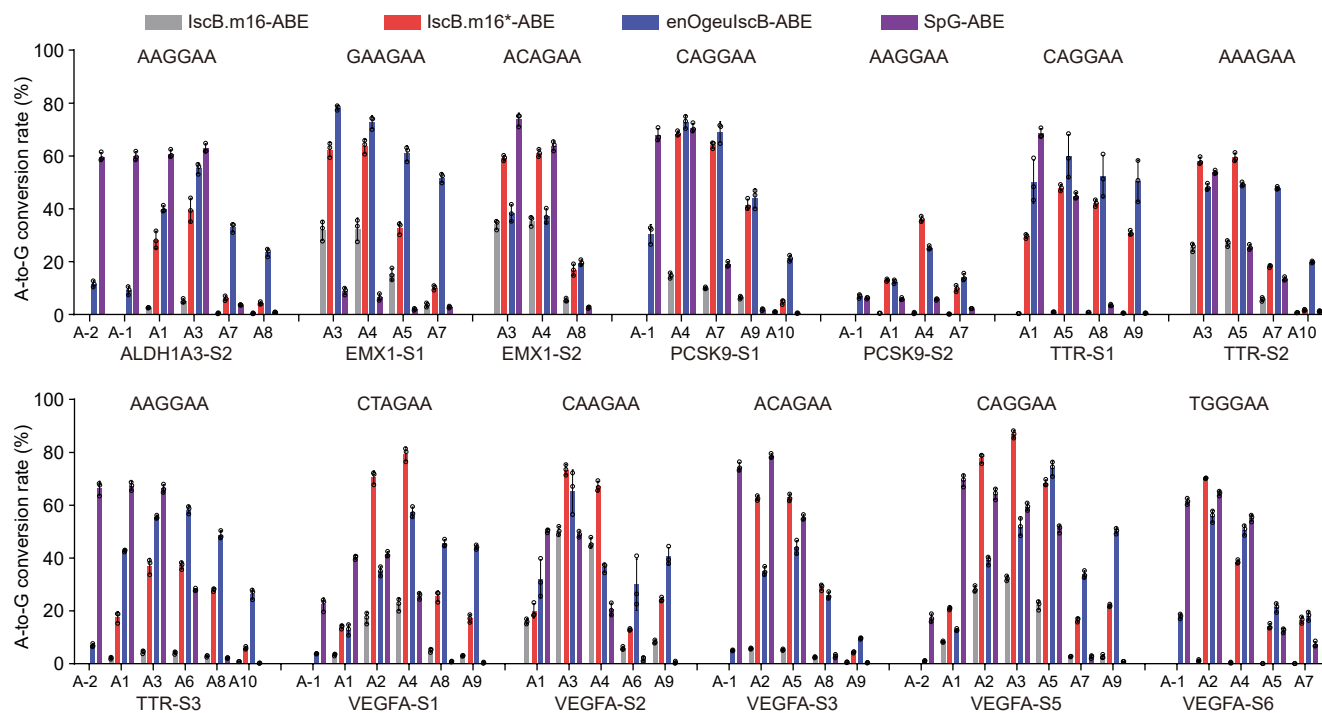
Extended Data Fig. 3 | TAM profiling of IscB.m16 WT and its variants using GFxxFP fluorescence reporter system in HEK293T cells. a, b, TAM profiling of IscB.m16 at 19 reporters containing different TAMs. **c,** Screening for highly active variants from saturated mutants at H380, Q381 and M424 of IscB.m16 proteins based on GFxxFP reporters containing a TAM pool1 target. Each dot represents the editing activity for a single variant. The dashed line indicates the editing

activity of WT. Colored dots reflect the mean of three independent biological replicates. **d, e,** Comparison of editing activity among IscB.m16 WT and its variants at 19 GFxxFP reporters, which contain different TAMs. Data are shown as mean of 3 independent biological replicates. **f,** TAM profiling of IscB.m16^{BSH} at 64 reporters containing the same spacer and 5'-NNNGAA-3' TAM sequences. Values and error bars represent mean and s.d., n = 3 independent biological replicates.



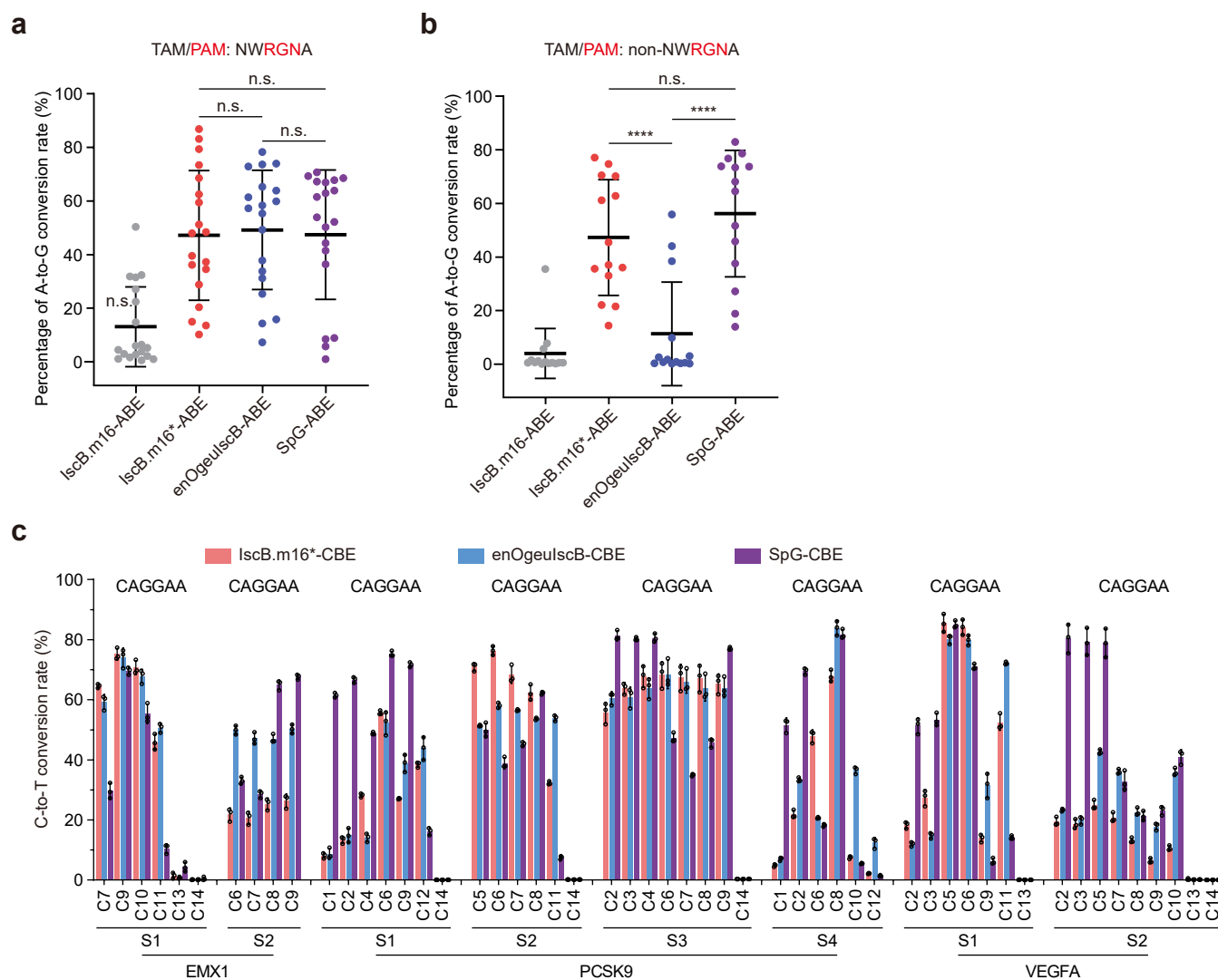
Extended Data Fig. 4 | Characterization of LscB.m16 and LscB.m16^{*} editing activities in HEK293T cells. a, The indels patterns generated by wild-type LscB.m16 and its variants at VEGFA-S1 site. **b,** Cleavage activity of wild-type LscB.m16 and its variants are affected by spacer length at two target sites on GFxxFP

reporters. **c,** Spacer length-dependent effects on cleavage activity in LscB.m16^{*} and enOgeulscB systems targeting two sites in human genome. Values and error bars were shown as mean and s.d., n = 3 independent biological replicates.



Extended Data Fig. 5 | A-to-G conversion efficiency of IscB- and SpG-derived adenine base editors at endogenous loci in HEK293T cells. Comparisons of A-to-G conversion efficiency of IscB.m16-ABE, IscB.m16*-ABE, enOgeulscB-ABE

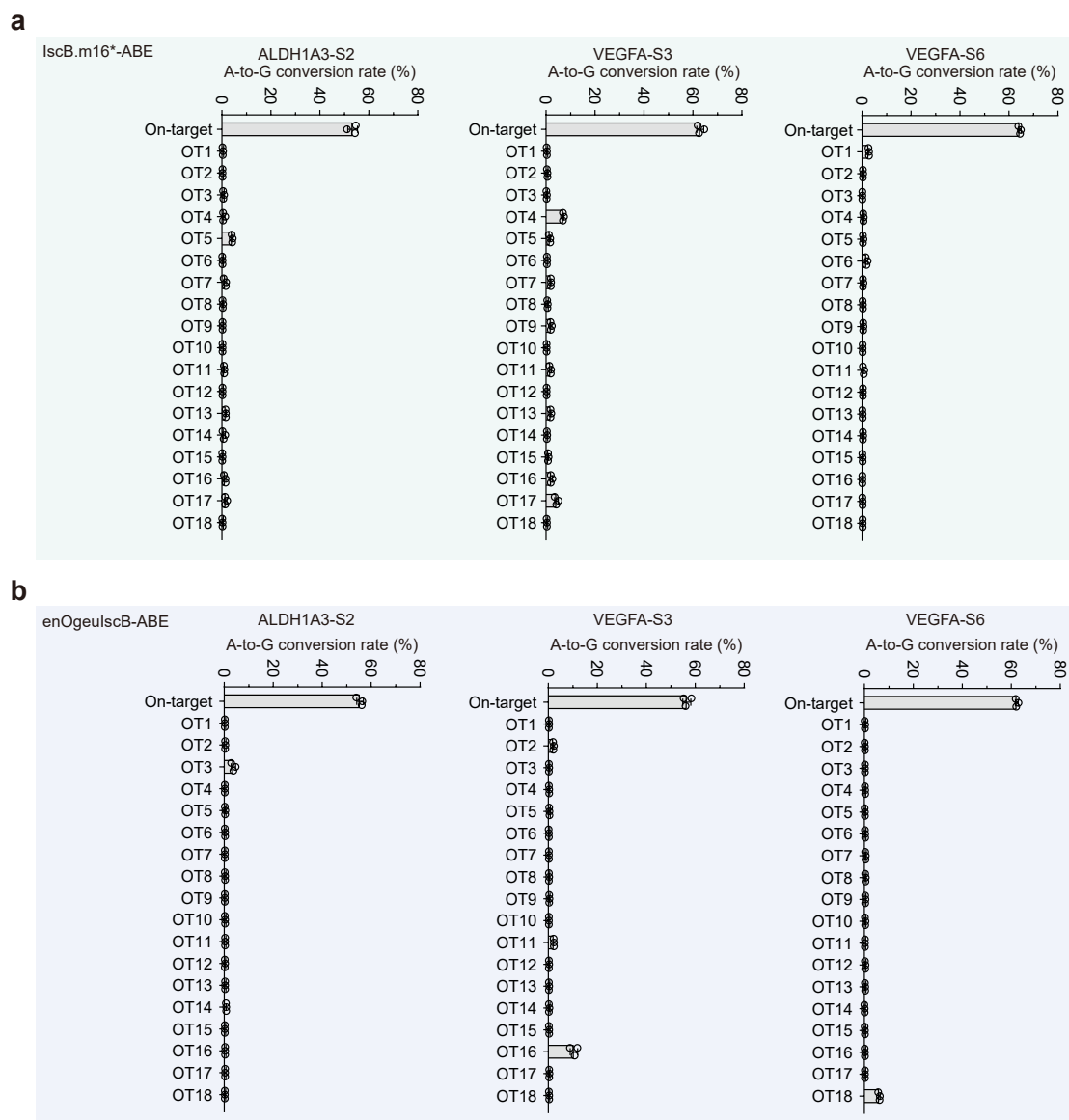
and SpG-ABE at 13 target sites containing various TAMs in HEK293T cells. Value and error bars are presented as means \pm s.d., $n = 3$ independent biological replicates.



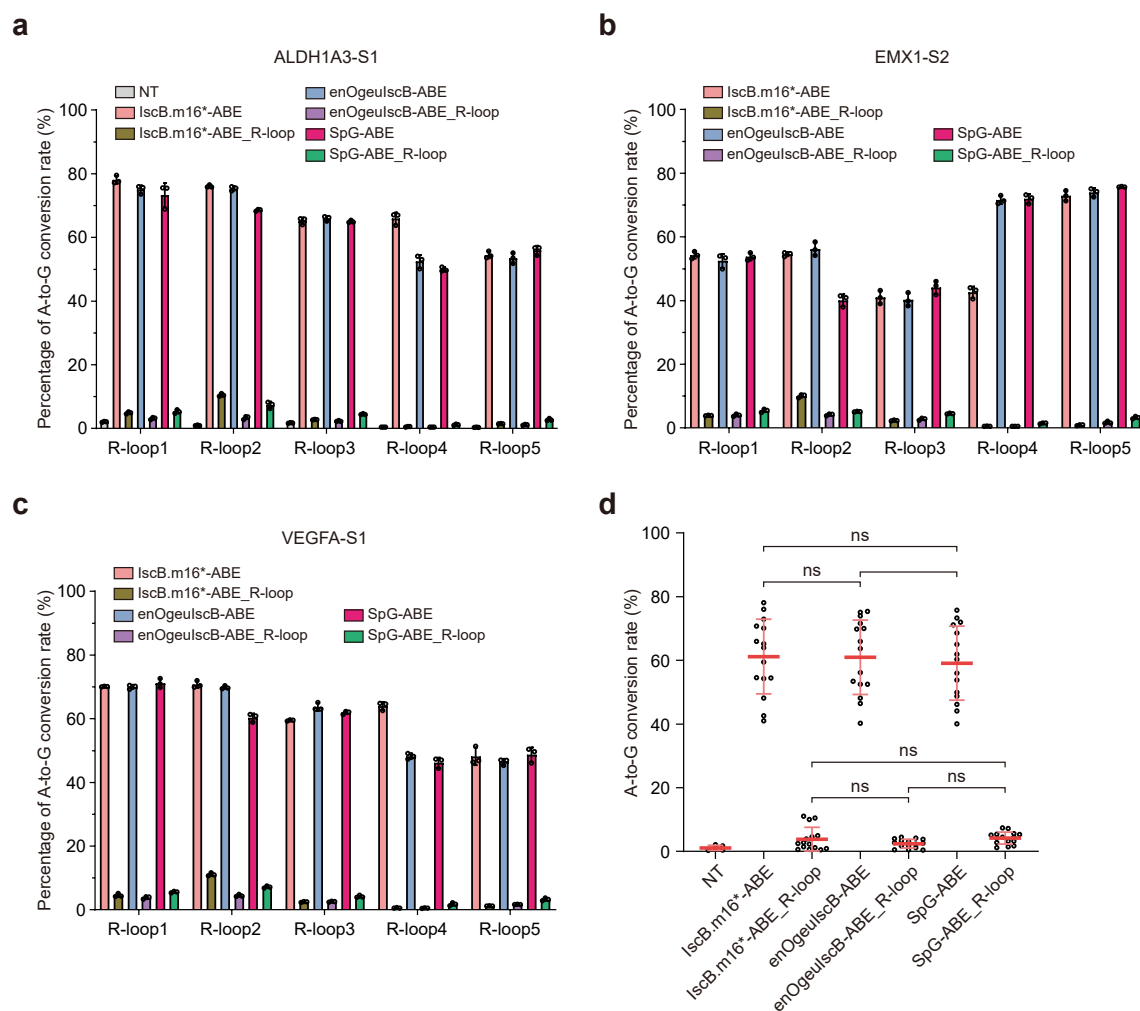
Extended Data Fig. 6 | Base editing efficiency of IscB- and SpG-based adenine and cytosine base editors at endogenous loci in HEK293T cells.

a, b, Comparison of the A-to-G conversion efficiency of IscB.m16-ABE, IscB.m16*-ABE, enOgeulscB-ABE, and SpG-ABE at 19 TAM sites of NWRGNA (**a**) and 14 TAM sites of NNNGNA, but non-NWRGNA (**b**). Data are collected from 19 and 14 endogenous sites, respectively, and presented as means \pm s.d. Each dot represents the average highest base editing activity at each endogenous target

site of three independent biological replicates. *P* values determined by Tukey's multiple comparisons test following ordinary one-way analysis of variance. The *P*_{adj.} are 0.613, 0.00005 and 0.0000007, respectively. **c**, Comparisons of C-to-T conversion efficiency of IscB.m16-CBE, IscB.m16*-CBE, enOgeulscB-CBE, SpG-CBE at 8 target sites containing various TAMs in HEK293T cells. Value and error bars are presented as means \pm s.d., *n* = 3 independent biological replicates.

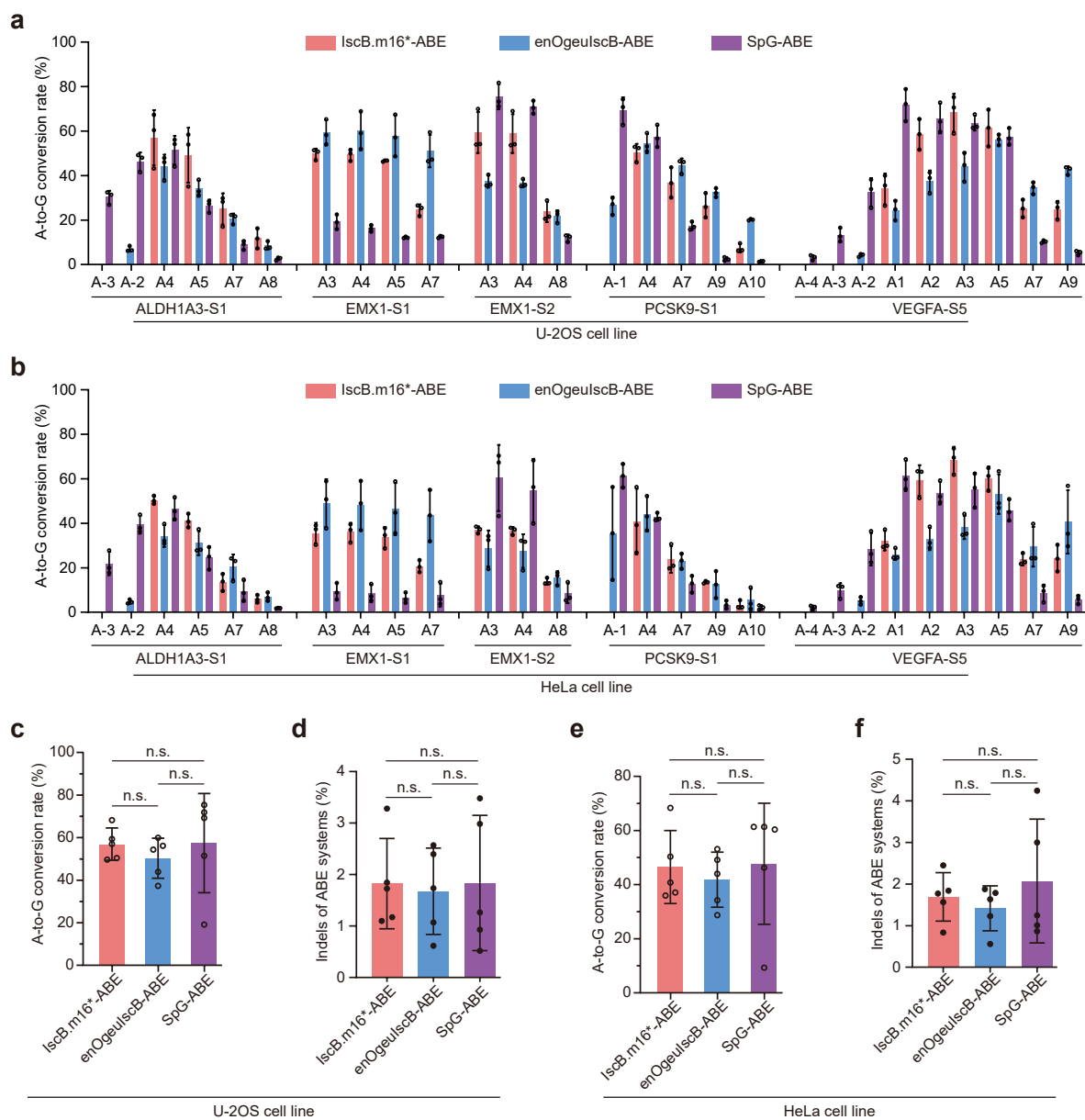


Extended Data Fig. 7 | The gRNA-dependent off-target levels of ABEs at *in-silico* predicted off-target sites, determined by targeted deep sequencing. **a, **b** A-to-G conversion rate of IscB.m16*-ABE (**a**) and enOgeulscB-ABE (**b**) targeting ALDH1A3-S2, VEGFA-S3 and VEGFA-S6 sites respectively. Value and error bars are presented as means \pm s.d., n = 3 independent biological replicates.**



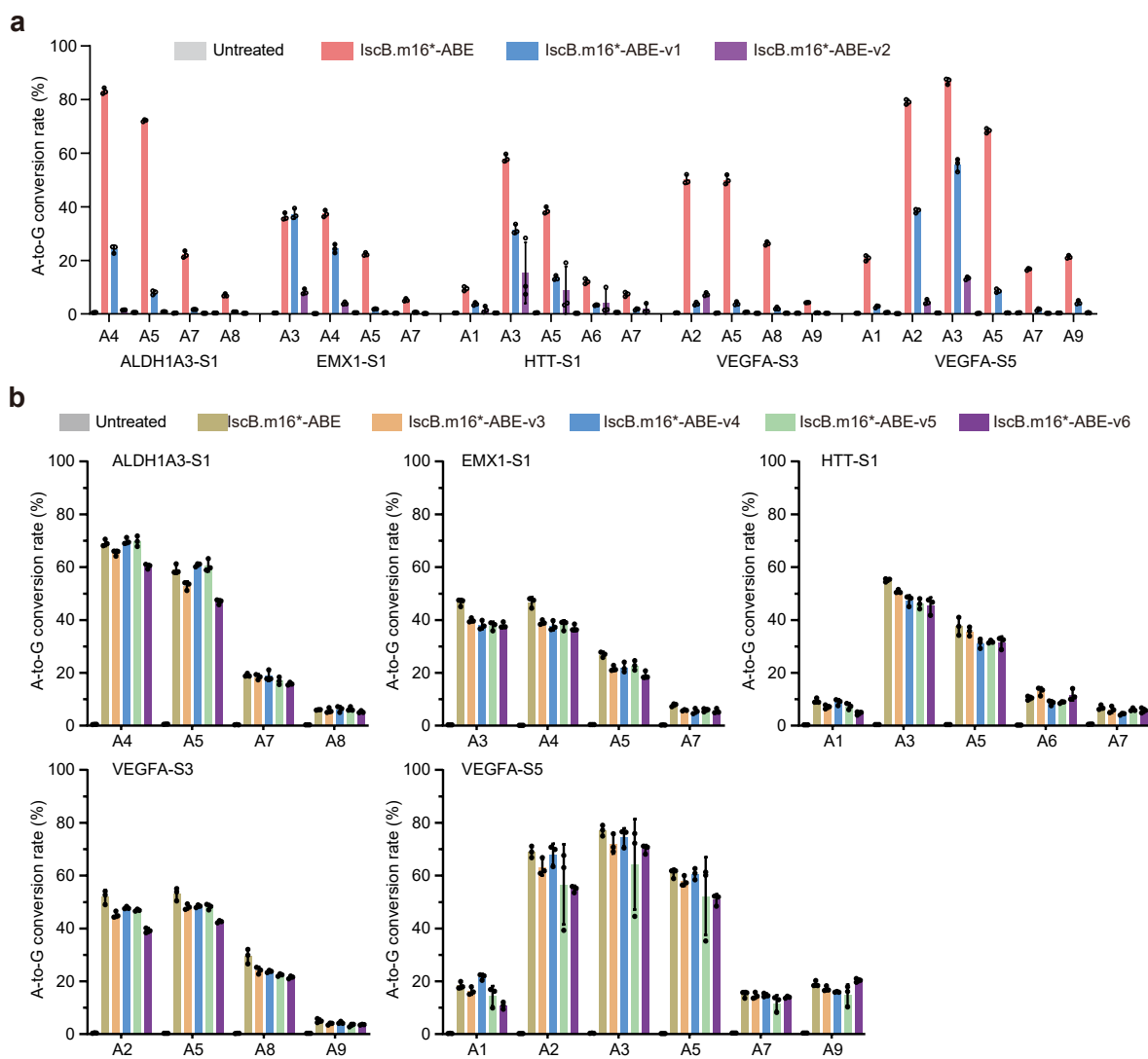
Extended Data Fig. 8 | The gRNA-independent off-target levels of IscB- and SpG-derived adenine base editors at five R-loops formed by dSaCas9 in HEK293T cells. a–c. The gRNA-independent off-target levels of base editors at ALDH1A3-S1 (**a**), EMX1-S2 (**b**), and VEGFA-S1 (**c**). **d.** Statistical analysis of gRNA-independent off-target levels at 3 endogenous sites. Data collected from

3 endogenous sites and each site with 5 R-loops of each editor respectively. The single dot presents mean of the highest edits at one site and one of the R-loop. Value and error bars are presented as means \pm s.d., *P* values determined by Tukey's multiple comparisons test following ordinary one-way analysis of variance. NS, not significant.



Extended Data Fig. 9 | A-to-G conversion efficiency of LscB- and SpG-based adenine base editors at endogenous loci in U-2OS and HeLa cells. **a, b**, Comparisons of A-to-G conversion efficiency of LscB.m16*-ABE, enOgeulscB-ABE, SpG-ABE at five target sites containing various TAMs in U-2OS (**a**) and HeLa (**b**) cells. Value and error bars are presented as means \pm s.d., $n = 3$ independent

biological replicates. **c-f**, Statistical of A-to-G editing efficiency (**c, e**) and indels (**d, f**) at five endogenous sites in U-2OS (**c, d**) and HeLa (**e, f**) cells, respectively. Value and error bars are presented as means \pm s.d., $n = 3$ independent biological replicates. P values determined by Tukey's multiple comparisons test following ordinary one-way analysis of variance. NS, not significant.



Extended Data Fig. 10 | A-to-G conversion efficiency of IscB.m16*-ABE with different TadA8e variants and different linkers. a. Comparison of editing window and base editing activity of IscB.m16*-ABEs with different TadA8e

variants at 5 endogenous loci. b. Comparison of editing window and base editing activity of IscB.m16*-ABEs with different linkers at 5 endogenous sites. Value and error bars are presented as means \pm s.d., $n = 3$ independent biological replicates.

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

Data analysis

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](#)

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	Except for the mutant screening experiment, which had only one biological experiment, all the other experiments used at least three independent biological replicates. Sample sizes number were listed in the corresponding figure legends. Sample sizes for three experiments were chosen based upon field standards and prior knowledge of experimental variation. (e.g. Gaudelli et al., Nature 2017; Koblan et al., Nat Biotechnol 2021)
Data exclusions	No data were excluded.
Replication	All main text figure experiments were repeated at least once and attempts at replication were successful.
Randomization	Human cells were grown at identical condition, after seeding cells into 24-wells plates. Wells were randomly assigned for cell transfection of all groups.
Blinding	Blinding was not relevant to our study because it is not a subjective trial, NGS and other experiments used for quantification in our study are not influenced by human interpretation and/or bias.

Behavioural & social sciences study design

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

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Non-participation

State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.

Randomization

If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.

Ecological, evolutionary & environmental sciences study design

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

Study description

Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample

Describe the research sample (e.g. a group of tagged *Passer domesticus*, all *Stenocereus thurberi* within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy

Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection

Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale

Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken

Data exclusions

If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility

Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization

Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding

Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Did the study involve field work? Yes No

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 | Included in the study |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | | |
|-------------------------------------|--|
| n/a | Included in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

Antibodies used

Antibodies for western blot antibodies: anti-dystrophin (Sigma, D8168, 1:1000 dilution), anti-vinculin (CST, 13901S, 1:1000 dilution), HRP-conjugated Goat Anti-Rabbit IgG (Beyotime, A0216, 1:1000 dilution); Antibodies for immunofluorescence: anti-dystrophin

(Abcam, ab15277, 1:100 dilution), anti-spectrin (Millipore, MAB1622, 1:500 dilution), Alexa Fluor® 488 AffiniPure donkey anti-rabbit IgG (Jackson ImmunoResearch labs, 711-545-152, 1:1000 dilution) and Alexa Fluor 647 AffiniPure donkey anti-mouse IgG (Jackson ImmunoResearch labs, 715-605-151, 1:1000 dilution) were used in this study.

Validation

Validation information is available at the following link: <https://www.sigmaaldrich.com/HK/zh/product/sigma/d8168>; <https://www.cellsignal.com/products/primary-antibodies/vinculin-e1e9v-xp-rabbit-mab/13901>; <https://www.abcam.com/products/primaryantibodies/dystrophin-antibody-ab15277.html>; https://www.merckmillipore.com/HK/en/product/msds/MM_NF-MAB1622; <https://dshb.biology.uiowa.edu/PAX7>; <https://www.beyotime.com/product/AF7686.htm>; ReferrerURL=<https://www.bing.com/>

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HEK293T (Catalog # BNCC353535), U-2OS (Catalog # BNCC352039) and HeLa (Catalog # BNCC340396) cell lines were purchased from BNCC.

Authentication

Cell lines were authenticated using STR profiling by the supplier.

Mycoplasma contamination

All cell lines tested negative for mycoplasma contamination.

Commonly misidentified lines (See [ICLAC](#) register)

None misidentified lines were used.

Animals and other research organisms

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

Laboratory animals

Humanized DMDΔmE5051 and K1hE50/Y C57BL/6 J mice at the age of 3-weeks were used for recipients.

Wild animals

This study did not involve wild animals.

Reporting on sex

Male mice were used in this study.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

All animal experiments in this study were performed following approved protocols and guidelines set by the Animal Care and Use Committee of Huidagene Therapeutics Co., Ltd.

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

Post-transfected cells were culture medium, and analyzed by flow cytometry.

Instrument

For EGFP activation assay, Beckman CytoFlex was used, and BD FACSAria III was used for cell sorting.

Software

FACS data were analyzed with FlowJo X (v10.0.7)

Cell population abundance

EGFP activation rates = EGFP positive cell numbers / mCherry and BFP double positive cell numbers.

Gating strategy

Gating strategy: 1) in FSC-A/SSC-A or FSC-H/SSC-H gate for living cells; 2) using the non-transfected and no-target cells to define the gate for BFP+ and GFP+ cells for indel analysis.

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