nature chemical biology

Article

https://doi.org/10.1038/s41589-025-01902-7

Engineering eukaryotic transposon-encoded Fanzor2 system for genome editing in mammals

Received: 17 August 2024

Accepted: 3 April 2025

Published online: 20 May 2025

Check for updates

Yinghui Wei[1,2,7] ✉, Pengfei Gao[1,7], Deng Pan[3,7], Guoling Li[4,5,7], Yufei Chen[1,7], Shangpu Li[1], Henan Jiang[1], Yang Yue[1], Zhenmin Wu[1], Zujiang Liu[1], Min Zhou[6], Yulin Chen[1,2], Kun Xu[1,2] ✉, Zhaowei Wu[3] ✉ & Xiaolong Wang[1,2] ✉

Eukaryotic transposon-encoded Fanzor proteins hold great promise for genome-engineering applications as a result of their compact size and mechanistic resemblance to TnpB. However, the unmodified Fanzor systems show extremely low activity in mammalian cells. Guided by the predicted structure of a Fanzor2 complex using AlphaFold3, we engineered the NlovFz2 nuclease and its cognate ωRNA to create an evolved enNlovFz2 system, with an expanded target-adjacent motif (TAM) recognition scope (5′-NMYG) and a substantially improved genome-editing efficiency, achieving an 11.1-fold increase over the wild-type NlovFz2, comparable to two previously reported IS200 or IS605 transposon-encoded TnpBs and two CRISPR–Cas12f1 nucleases. Notably, enNlovFz2 efficiently mediated gene disruption in mouse embryos and restored dystrophin expression in a humanized Duchenne muscular dystrophy mouse model with single adeno-associated virus delivery. Our findings underscore the potential of eukaryotic RNA-guided Fanzor2 nucleases as a versatile toolbox for both biological research and therapeutic applications.

RNA-guided enzymatic systems play a pivotal role in biological processes by utilizing the complementarity between guide (g)RNA and target nucleic acid sequences to identify genetic elements in both prokaryotes and eukaryotes. CRISPR–Cas systems confer adaptive immunity in prokaryotes, protecting nearly all archaea and numerous bacterial species from invading nucleic acids[1–3]. Recently, a new class of RNA-guided systems, termed obligate mobile element-guided activity (OMEGA), has been reported, comprising TnpB[4–7], IscB[7–9], IsrB[10] and IshB[8] nucleases in prokaryotes and Fanzor[11–13] nucleases in eukaryotes. OMEGA systems encompass an RNA-guided DNA endonuclease and a single noncoding RNA (ncRNA), referred to as ωRNA, which is transcribed from the transposon end-region[8]. The OMEGA effector TnpB contains a RuvC-like nuclease domain (ribonuclease H fold) and shares structural similarity with the RuvC nuclease domains of various type V CRISPR–Cas12 effectors[14–16], suggesting a direct evolutionary path from TnpB to Cas12 (ref. 4,8,17). Recent research has provided evidence that the eukaryotic Fanzor acts as a CRISPR–Cas/OMEGA-like, programmable, RNA-guided endonuclease, suggesting that TnpB may be its evolutionary precursor[11,12].

Fanzor, a family of eukaryotic TnpB-IS200- or -IS605-like proteins encoded by transposable elements (TEs), was first identified in 2013 and is presumed to potentially regulate TE activity through

[1]International Joint Agriculture Research Center for Animal Bio-Breeding of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Northwest A&F University, Yangling, China. [2]Hainan Institute, Northwest A&F University, Sanya, China. [3]School of Physical Science and Technology, ShanghaiTech University, Shanghai, China. [4]Zhongshan Institute for Drug Discovery, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Zhongshan, China. [5]HuidaGene Therapeutics Co. Ltd, Shanghai, China. [6]Life Science Research Core Services, Northwest A&F University, Yangling, China. [7]These authors contributed equally: Yinghui Wei, Pengfei Gao, Deng Pan, Guoling Li, Yufei Chen. ✉e-mail: yinghuiwei@nwafu.edu.cn; xukunas@nwafu.edu.cn; wuzw1@shanghaitech.edu.cn; xiaolongwang@nwafu.edu.cn

methyltransferase activity[17]. Fanzors have been discovered across various eukaryotic lineages and can be primarily categorized into two major groups[11–13]: (1) Fanzor1, associated with eukaryotic transposons such as Mariners, IS4-like elements, Sola, Helitron and MuDr, predominantly found in fungi, protists, arthropods, plants and eukaryotic viruses, particularly within giant viruses; and (2) Fanzor2, found in IS607-like transposons and double-stranded (ds)DNA viral genomes. Two Fanzor1 proteins (SpuFz1 from *Spizellomyces punctatus* and KnFNuc from *Klebsormidium nitens*) and five Fanzor2 proteins (NlovFz2 from *Naegleria lovaniensis*, MmeFz2 or MmFNuc from *Mercenaria mercenaria*, ApmFNuc from *Acanthamoeba polyphaga mimivirus* and DpFNuc from *Dreissena polymorpha*) have been confirmed as programmable RNA-guided nucleases with potential application in mammalian genome editing[11,12]. Nevertheless, the low editing efficiency of unmodified Fanzor nucleases in mammalian cells has become a notable limitation. Furthermore, it remains unclear whether the compact Fanzor nucleases can be efficiently packaged into a single recombinant adeno-associated virus (rAAV) for in vivo gene therapy applications.

In the present study, we present a comprehensive effort to enhance the NlovFz2 system for genome editing. Using the predicted structure of the NlovFz2 complex generated by the state-of-the-art AlphaFold3 platform, we performed three rounds of ωRNA engineering and conducted a comprehensive screening of arginine substitutions in NlovFz2. This effort led to the development of an enhanced enNlovFz2 system, characterized by an expanded TAM recognition range (5′-NMYG) and significantly improved genome-editing efficiency in human cells, comparable to the previously reported enISDra2 TnpB[18], ISDra2 TnpBmax[19], ISAba30 TnpB[20], enCnCas12f1 (ref. 21) and enRhCas12f1 (ref. 22). Furthermore, enNlovFz2 successfully restored dystrophin expression in a humanized Duchenne muscular dystrophy (DMD) mouse model through single AAV delivery. Collectively, enFanzor2 expands the repertoire of miniature genome editors for therapeutic applications.

Results

Engineering the NlovFz2-ωRNA system to enhance its activity

Engineering ncRNA is essential for improving the genome-editing efficiency of various compact, programmable, RNA-guided nucleases[18,21–25]. To better understand the structure and interactions of NlovFz2, its cognate ωRNA and target DNA, we generated a predicted ternary complex structure using the state-of-the-art AlphaFold3 platform (Fig. 1a and Supplementary Fig. 1). As shown in Fig. 1b, the NlovFz2 ωRNA scaffold may form three central stem loops, S1, S2 and S3, along with a pseudo-knot (PK) and a triplex structure. Intriguingly, the ωRNA structural configurations differ between NlovFz2 (a member of Fanzor2) and SpuFz1 (a member of Fanzor1) (Extended Data Fig. 1a), because SpuFz1's ωRNA contains only two primary stem loops (S1 and S2) and lacks both a PK and a triplex structure[11]. In contrast, the NlovFz2 ωRNA may share structural similarities with ωRNAs found in TnpB systems[5] and single guide (sg)RNAs in certain miniaturized Cas12 systems[26]. The PK, triplex and adjacent ωRNA moieties probably interact primarily with NlovFz2 (Fig. 1a and Supplementary Fig. 1). Therefore, we attempted to truncate and mutate the distal ends of stem loops S1, S2 and S3 to enhance the structural stability of the ωRNA (Fig. 1b). The *B2M* site, identified as the most efficient in an earlier study[11], was selected as a benchmark for guiding the engineering efforts of the NlovFz2-ωRNA system in the present study. The genome-editing efficiency of ωRNA variant was evaluated by targeted amplicon sequencing (Extended Data Fig. 2a). To enhance the stability of ωRNA, we first replaced the A–U/G–U/U–U base-pairs with thermodynamically stable G–C/C–G base-pairs in S1, S2 and S3 (Fig. 1b). We found that seven variants (GU91GC, UA95GC, UA96CG, UA97CG, GU102GC, AU103CG and AU103GC) in S3 significantly enhanced the genome-editing activity compared with the wild-type (WT) ωRNA (Fig. 1c). Moreover, we discovered that none of the combinatorial mutations derived from these seven variants led to any further improvement in activity (Extended Data Fig. 2b). Next, we investigated whether the stem loops could be shortened. Three S3 loop truncations (S3Loop-Δ2U/Δ4U/Δ5U) and two S3 stem truncations (S3-Δ1bp/Δ2bp) exhibited improved activity (Fig. 1c). We concluded that S3 modifications may substantially contribute to the structural stability of the NlovFz2-ωRNA scaffold, thereby improving its genome-editing activity. Collectively, by combining the most efficient four variants (UA97CG, GU102GC, AU103CG and AU103GC) with four truncations (S3Loop-Δ4U/Δ5U and S3-Δ1bp/Δ2bp), we ultimately selected three combinatorial ωRNA mutants (S3Loop-Δ4U + UA97CG, S3Loop-Δ5U + AU103GC and S3Loop-Δ4U + S3-Δ1bp) for further evaluations (Fig. 1d). These variants exhibit an almost fivefold increase in genome-editing activity compared with the WT-ωRNA (Fig. 1d).

The NlovFz2 nuclease is predicted to consist of five domains: an amino-terminal structural unknown domain (UNK, ~1–109) with an expected local distance difference test value (pLDDT) of <70, a recognition domain (REC, ~118–211), a wedge domain (WED, ~110–117 and ~212–278), a RuvC nuclease domain (~279–440 and ~469–498) and a zinc-finger element (ZF, ~441–468) inserted within the RuvC domain (Supplementary Fig. 1a). It is interesting that NlovFz2 shares more structural similarities with TnpB[5] and certain miniaturized Cas12 nucleases[26], than with the SpuFz1 nuclease[11] (Extended Data Fig. 1b). A series of previous studies employed arginine substitution screening to enhance the interaction between RNA-guided nucleases and the nucleic acid molecules with which they are complexed, thereby improving genome-editing activity in eukaryotic cells[22,24,27–29]. A comprehensive arginine substitution screening was performed to identify NlovFz2 variants with significantly improved genome-editing efficiency. We employed a BFP-T2A-EGFxxFP fluorescence reporter system to detect the activity of NlovFz2 variants in mammalian cells. The reporter system contains two enhanced green fluorescent protein (EGFP) fragments (~1–561 bp and ~112–720 bp of the coding DNA sequence) separated by a short sequence that is identical to the endogenous *B2M* locus (5′-CCG TAM)[22,24,29] (Extended Data Fig. 2c). After targeted cleavage by RNA-guided DNA nucleases, EGFP can be activated through a repair pathway mediated by single-strand annealing[30]. In the present study, five mutations (Pro6Arg, Gln44Arg, Glu64Arg, Thr178Arg and Gln285Arg) were identified, showing more than a 1.2-fold improvement in genome-editing activity compared with WT-NlovFz2, as determined by the percentage of cells with activated EGFP fluorescence (Fig. 1e). Next, we tested the activity of combinatorial mutations and found that their activity did not improve further compared with individual mutants (Extended Data Fig. 2d). Therefore, we selected the top two single mutations with Pro6Arg and Gln285Arg for further optimization (Fig. 1e and Extended Data Fig. 2d). Finally, we evaluated the genome-editing activity by combining ωRNA variants (S3Loop-Δ4U + UA97CG, S3Loop-Δ5U + AU103GC and S3Loop-Δ4U + S3-Δ1bp) with NlovFz2 variants (Pro6Arg and Gln285Arg). We found that the combination of S3Loop-Δ4U + UA97CG ωRNA and the NlovFz2 Q285R variant, termed enNlovFz2, exhibits the highest genome-editing activity, showing approximately a sixfold increase compared with the WT-NlovFz2 system (Fig. 1f). We further compared the genome-editing activity and patterns of the WT and engineered versions of the NlovFz2-ωRNA system and found that enNlovFz2 showed the highest genome-editing activity and the broadest deletion size (Fig. 1g). Taken together, these results demonstrate that enNlovFz2 exhibits high genome-editing efficiency with its canonical 5′-NCCG TAM recognition.

EnNlovFz2 recognizes 5′-NMYG TAMs for dsDNA targeting

To determine the TAM preference of enNlovFz2 during DNA targeting, we conducted a TAM depletion assay by in vitro cleavage of a plasmid library containing 6-bp randomized TAM sequences, using purified enNlovFz2 ribonucleoprotein (RNP) expressed in *Escherichia coli* cells (Fig. 2a and Supplementary Fig. 2). Throughout this process, plasmids containing effective TAMs for enNlovFz2 were depleted,

https://doi.org/10.1038/s41589-025-01902-7

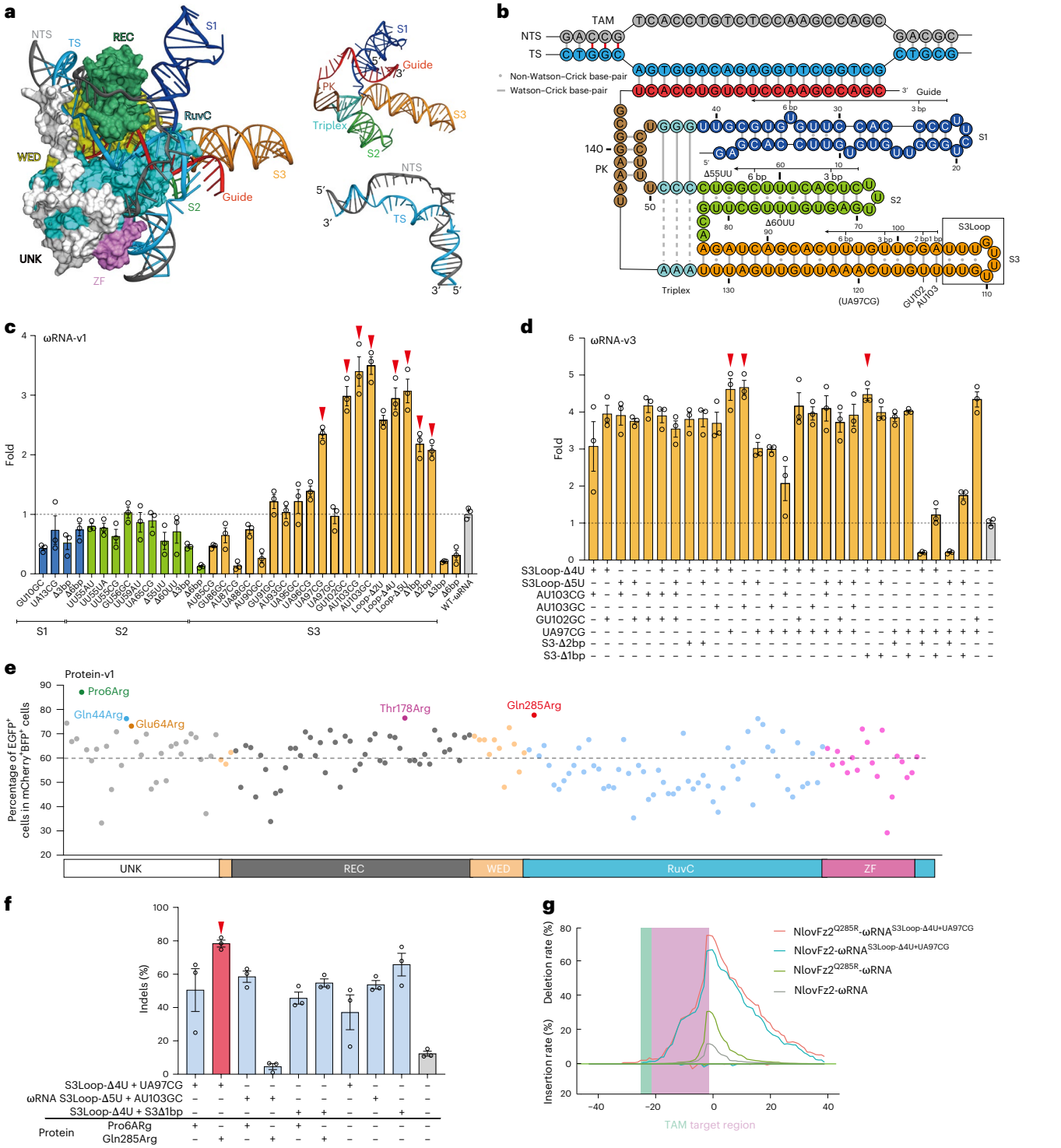


Fig. 1 | Structure-guided ωRNA and protein engineering for NlovFz2 in human cells. a, Predicted structure of the NlovFz2 RNP in complex with a *B2M* target dsDNA, generated by AlphaFold3. **b**, Secondary structure of NlovFz2 ωRNA complexed with a *B2M* target dsDNA. Stem loops, S1, S2 and S3 for ωRNA optimization are indicated. NTS, non-target strand; TS, target strand. **c**, First round of ωRNA optimization by replacing the A–U/G–U/U–U base-pairs and truncating the stems. The optimal ωRNAs (UA97CG, GU102GC, AU103CG, AU103GC, S3Loop-Δ4U, S3Loop-Δ5U, S3-Δ1bp and S3-Δ2bp), chosen for further optimization, are marked with red triangles. **d**, Third round of ωRNA optimization by combining high-efficiency variants of A–U and G–U base-pair substitutions with stem truncations in S3. The optimal ωRNA combinations (UA97CG + S3Loop-Δ4U, AU103GC + S3Loop-Δ5U and S3-Δ1bp + S3Loop-Δ4U), selected for further optimization, are marked with red triangles. **e**, Comprehensive screening for highly efficient variants by substituting amino acid residues in the functional domains of the NlovFz2 protein with arginine. The dashed line indicates the cleavage activity of WT-NlovFz2. **f**, Combination of three ωRNA variants and two protein mutants (Pro6Arg and Gln285Arg) further increases genome-editing efficiency at the endogenous *B2M* locus in HEK293T cells. **g**, Metaplot showing small indel patterns generated by NlovFz2-ωRNA, NlovFz2^{Q285R}-ωRNA, NlovFz2-ωRNA$^{S3Loop-\Delta4U+UA97CG}$ and NlovFz2^{Q285R}-ωRNA$^{S3Loop-\Delta4U+UA97CG}$ (enNlovFz2) in HEK293T cells at the endogenous *B2M* locus. Values and error bars are expressed as the mean ± s.e.m. ($n = 3$ independent biological replicates), except for **e** (two replicates). Fold-change represents the ratio of the editing efficiency of ωRNA variants to that of WT-ωRNA. The top 30% of mCherry$^+$ cells were sorted by FACS to evaluate the genome-editing efficiency of the NlovFz2-ωRNA system at the endogenous *B2M* locus.

https://doi.org/10.1038/s41589-025-01902-7

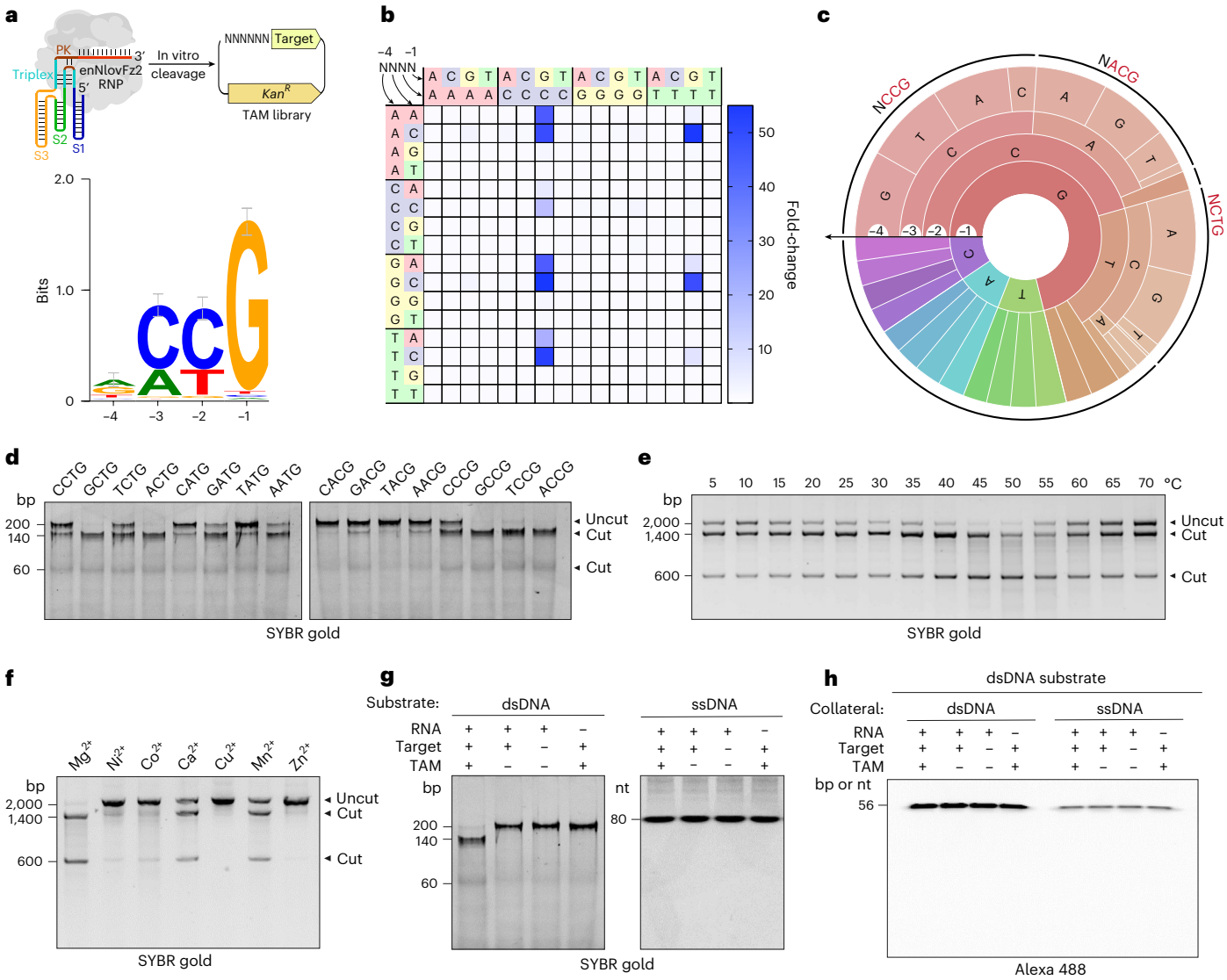


Fig. 2 | Preferences in TAM and biochemical characterization of enNlovFz2. **a**, Schematic representation of the TAM depletion assay and WebLogo depicting the TAM preference of enNlovFz2. **b**, Comprehensive TAM preference analysis of all 256 4-nt TAMs by enNlovFz2. **c**, TAM wheel plotting the TAM preference of enNlovFz2. Sequences from the inner to the outer circles correspond to the PAM read moving away from the protospacer. Colors indicate the relative frequency of the innermost nucleotide, whereas the sector area is directly proportional to the sequence's relative enrichment in the library. **d**, In vitro dsDNA cleavage by ωRNA-guided enNlovFz2 with 16 different TAMs. **e**, EnNlovFz2-mediated cleavage activity of target dsDNA across temperatures ranging from 5 °C to 70 °C. **f**, EnNlovFz2-mediated target dsDNA cleavage dependence on divalent metal ions. After proteinase treatment, the substrates were analyzed by agarose gel electrophoresis. **g**, EnNlovFz2 exclusively cleaving the target dsDNA. After treatment with proteinase, the target nucleic acid species were analyzed by tris base, boric acid and EDTA–polyacrylamide gel electrophoresis to detect dsDNA and ssDNA. The gels were imaged using SYBR Gold channels. **h**, enNlovFz2 not exhibiting collateral activity on Alexa Fluor-488-labeled collateral dsDNA and ssDNA.

and the remaining TAMs were determined through targeted amplicon sequencing. The position–weight matrix indicated that enNlovFz2 typically recognizes 5′-NMYG TAMs (where M = C or A and Y = C or T) (Fig. 2a). The full profile of the TAM preference for enNlovFz2 is shown in Fig. 2b, where the depletion fold-change of 256 4-nt TAMs is plotted in a heatmap. This is consistent with the TAM wheel analysis[31] shown in Fig. 2c. In addition, the in vitro DNA cleavage assay using different TAM-containing substrates demonstrated that enNlovFz2 prefers 5′-NCCG > 5′-NCTG > 5′-NATG = 5′-NACG (Fig. 2d). Given that the highest genome-editing activity was observed with 5′-NCCG TAMs, we concluded that the TAM preferences of enNlovFz2 follow this ranking: 5′-GCCG > 5′-TCCG > 5′-ACCG > 5′-CCCG (Fig. 2b–d). The cleavage activity of enNlovFz2 is supported by magnesium, calcium, manganese, cobalt and nickel, and it adapts to a broad temperature range from 5 °C to 70 °C (Fig. 2e, f). We also demonstrated that enNlovFz2 performs ωRNA-guided, TAM- and target-dependent dsDNA cleavage, but does not cleave the targeted single-stranded (ss)DNA (Fig. 2g). We further confirmed that, on recognizing dsDNA targets, enNlovFz2 does not exhibit cleavage activity on collateral dsDNA and ssDNA substrates (Fig. 2h).

Efficient genome editing in mammalian cells using enNlovFz2

The TAM of enNlovFz2 was characterized as 5′-NMYG using biochemical assay (Fig. 2). However, the optimal TAM for enNlovFz2-mediated genome editing at endogenous loci in mammalian cells remains unclear. In the present study, we investigated the genome-editing efficiencies of WT-NlovFz2 and enNlovFz2 at 124 endogenous loci with various TAMs in HEK293T cells. Plasmids expressing WT-NlovFz2 or enNlovFz2 were transfected into HEK293T cells and all mCherry$^+$ cells were sorted for targeted amplicon sequencing to determine the genome-editing

efficiency (Extended Data Fig. 3a). Our results confirmed that enNlovFz2 showed higher activity at endogenous loci with a TAM as 5′-NCYG compared with 5′-NAYG (Fig. 3a,b). Over 10% of genome-editing efficiency was observed at endogenous loci with TAMs as 5′-ACCG (8 out of 15 sites), 5′-ACTG (5 out of 9 sites), 5′-GCCG (5 out of 14 sites), 5′-GCTG (3 out of 13 sites) and 5′-TCCG (1 out of 6 sites) (Fig. 3a and Extended Data Fig. 3b–e). This indicates that the TAM preferences of enNlovFz2 follow this ranking: 5′-ACCG > 5′-ACTG > 5′-GCCG > 5′-GCTG > 5′-TCCG > 5′-TCTG > 5′-CCTG > 5′-CCCG (Fig. 3a and Extended Data Fig. 3b–e). Furthermore, the preference of enNlovFz2 was also analyzed by summarizing the insertion or deletion (indel) efficiencies at 124 endogenous loci, revealing the editing efficiencies ranked as follows—5′-NCCG > 5′-NCTG > 5′-NATG > 5′-NACG—consistent with the results of biochemical assays (Fig. 3b and Extended Data Fig. 3b–e).

To further assess the robustness of enNlovFz2 for genome editing, we systematically compared its activity with that of other previously reported miniaturized Cas12 nucleases, including enCnCas12f1 (ref. 21) and enRhCas12f1 (ref. 22), which share similar TAM or protospacer adjacent motif (PAM) preferences. The comparison was performed across a panel of 26 endogenous loci spanning ten genes. The results revealed a significant improvement in genome-editing efficiency across all 26 target loci with enNlovFz2, compared with WT-NlovFz2 and enRhCas12f1 (Fig. 3c). Notably, with the exception of a few sites among the 26 endogenous loci, the efficiency improvement achieved by enNlovFz2 was more pronounced than that of enCnCas12f1 (Fig. 3c). On average, enNlovFz2 (14.41 ± 11.74%) exhibited an 11.1-fold increase in efficiency compared with WT-NlovFz2 (1.30 ± 1.89%) (Fig. 3d). In addition, we compared the genome-editing activity of three previously reported IS200 or IS605 transposon-encoded TnpB nucleases[18–20] (enISDra2 TnpB, TnpBmax and ISAba30 TnpB) with that of enNlovFz2. Notably, WT-NlovFz2, enNlovFz2, enCnCas12f1 and enRhCas12f1 recognize the same 5′-CCG TAM, whereas ISDra2 TnpB and TnpBmax recognize a 5′-TTGAT TAM and ISAba30 TnpB a 5′-TGAC TAM. To enable direct comparison, we selected a subpanel of endogenous loci featuring additional 5′-TTGAT or 5′-TGAC TAMs, which are compatible with all the types of nucleases. We observed that the genome-editing activity of enNlovFz2 was slightly higher than that of ISAba30 TnpB and comparable to that of enISDra2 TnpB and TnpBmax nucleases (Extended Data Fig. 4a,b).

Next, we also compared the genome-editing activity of enNlovFz2 and WT-NlovFz2 at 12 endogenous loci within two genes (*Tyr* and *Hpd*) in the mouse N2a cells. Consistent with the observations in HEK293T cells, enNlovFz2 exhibited significantly higher genome-editing efficiency than WT-NlovFz2 across all 12 loci (Fig. 5a and Supplementary Fig. 3). Moreover, we systematically compared the editing efficiency of enNlovFz2 at the *B2M* and *DYRK1A* loci in HEK293T cells using two different delivery strategies: messenger RNA electroporation and plasmid lipofection. We found that delivery by mRNA electroporation significantly improved the genome-editing efficiency compared with plasmid lipofection (79.89% versus 35.14% at the *B2M* site; 72.67% versus 38.70% at the *DYRK1A* site) (Supplementary Fig. 4). Collectively, these results demonstrate that the enNlovFz2 system could be used for efficient genome editing in mammalian cells.

Genome-editing specificity of enNlovFz2

To further evaluate the specificity of enNlovFz2, we first employed CasOFFinder[32] to predict potential off-target sites for six endogenous loci (*B2M*, *CXCR4*, *DMD*-guide1, *DYRK1A*, *HPRT1* and *DMD*-guide4), by setting the DNA or RNA bulge size from 0 to 2 and the base-pair mismatch size from 2 to 5. Targeted amplicon sequencing revealed that the on-target editing efficiency of enNlovFz2 was significantly higher than that of enCnCas12f1 and enRhCas12f1 across nearly all sites (Fig. 4a and Extended Data Fig. 5). Furthermore, no off-target editing events were detected for all three editors across the six tested loci (Fig. 4a and Extended Data Fig. 5). To comprehensively analyze the genome-wide editing specificity, primer-extension-mediated sequencing (PEM–seq)[33] was conducted to detect potential off-target events of enNlovFz2, enCnCas12f1 and enRhCas12f1 at the *B2M*, *CXCR4* and *DMD*-guide1 loci (Fig. 4b). The data showed that enNlovFz2 exhibited slightly higher translocation rates than enCnCas12f1 and enRhCas12f1 (Fig. 4c,d), probably attributable to its significantly higher on-target editing efficiency compared with the other two nucleases (Fig. 4e). Together, these results indicate that enNlovFz2 exhibits high genome-editing efficiency with minimal off-target effect in mammalian cells.

In vivo genome-editing applications of enNlovFz2

To showcase the in vivo applicability and therapeutic potential of enNlovFz2, we selected the mouse *Tyr* (*Tyrosinase*) gene, associated with coat color, as a proof-of-concept target. By screening candidate targets on the *Tyr* gene in mouse N2a cells, we identified two genomic loci, site1 (13.17%) and site2 (16.69%), which exhibited relatively high editing efficiencies, and were selected for further experiments (Fig. 5a and Supplementary Figs. 5a and 6a). Next, we microinjected enNlovFz2 mRNA and *Tyr*-targeting ωRNA into one-cell embryos of D2B6F1 mice, generated from a cross between DBA2 males and C57BL/6 females (black coat color). Given that the *Tyr* gene determines black coat color in D2B6F1 mice, the genome-editing efficiency of enNlovFz2 can be estimated by directly examining coat color changes in treated mice. No abnormalities in blastocyst development rate was observed in the treated groups targeting site1 and site2, respectively, compared with the control group (Fig. 5b). We observed high editing efficiencies on both targets with averages of 38.26% and 48.72%, respectively (Fig. 5c and Supplementary Figs. 5b,c and 6b). Gene disruption induced by enNlovFz2 at two *Tyr* targets resulted in frameshift mutations, leading to albino or mosaic coat color phenotypes in the F0 mice (Fig. 5d and Supplementary Fig. 5c), indicating efficient disruption of the *Tyr* gene. Thus, our findings indicate that the enNlovFz2 system is an efficient miniature genome editor derived from eukaryotic species, which functions in cell lines and mouse embryos.

Given its hypercompact size, enNlovFz2 can be packaged into a single rAAV vector, making it a promising candidate for treating specific genetic diseases, such as DMD[22,34]. Previous studies have shown that skipping exon 43 of the dystrophin gene can restore the dystrophin expression in a mouse model with an exon 44 deletion[35], a mutation occurring in nearly 4% of patients with DMD[36]. We utilized enNlovFz2 to disrupt the GU site adjacent to the end of exon 43 in HEK293T cells and found that enNlovFz2 exhibited an editing efficiency of 22.82% at the *DMD*-guide4 locus (Fig. 3c). To further assess the potential of enNlovFz2 for in vivo therapeutic applications, we generated a genetically humanized DMD mouse model ($DMD^{\Delta E44\,mdx}$ mice) (Fig. 5e and Extended Data Fig. 6a). Immunostaining and western blotting revealed a complete loss of dystrophin expression in $DMD^{\Delta E44\,mdx}$ mice (Extended Data Fig. 6b,c). In addition, muscular histology, creatine kinase activity and motor function also suggested that $DMD^{\Delta E44\,mdx}$ mice presented severe DMD symptoms (Extended Data Fig. 6d–f). Then, we employed AAV serotype 9 (AAV9) for local delivery of enNlovFz2 and *DMD*-guide4-targeting ωRNA into skeletal muscle (Fig. 5f). Then 3 weeks postinjection, we evaluated the editing efficiency through targeted amplicon sequencing and analyzed the dystrophin expression using western blotting and histological staining (Fig. 5f). PCR detection across the genomic locus confirmed the anticipated exon 43 skipping (Fig. 5g). Genomic editing analysis revealed that the incidence of total indels ranged from 4.03% to 5.66%, with productive editing events ($3n + 2$ indels in exon 44) occurring at rates between 1.56% and 2.37%, resulting in therapeutic-level expression of functional dystrophin (Fig. 5h). Reverse transcription (RT)–PCR analysis of mRNA extracted from whole muscle revealed an out-of-frame efficiency of 9.93 ± 0.32%, an in-frame efficiency of 23.55 ± 0.94% and a skipping efficiency of 17.52 ± 0.75% for the enNlovFz2-treated group (Fig. 5i). Western blotting and immunostaining results further confirmed that dystrophin protein production was

https://doi.org/10.1038/s41589-025-01902-7

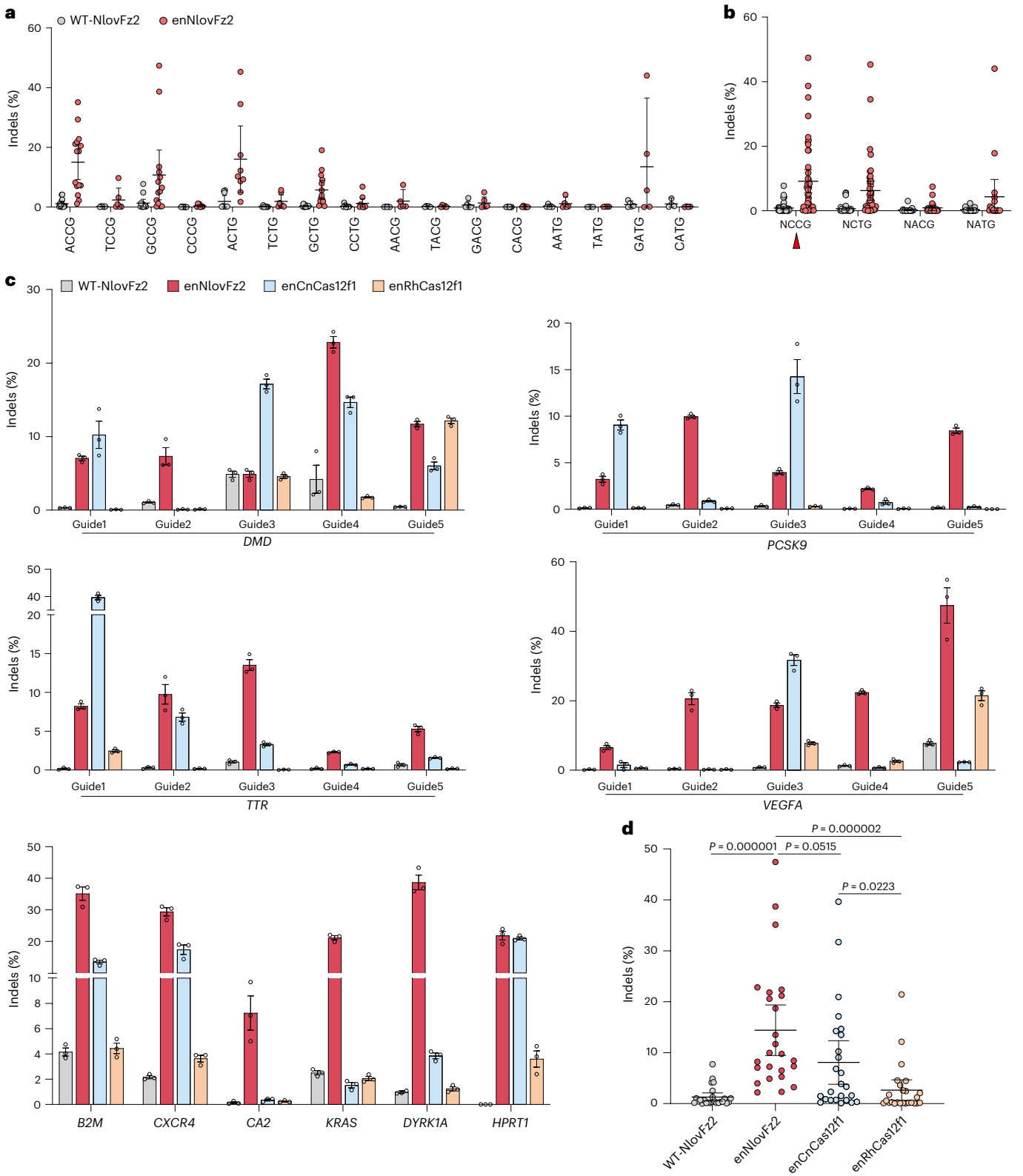


Fig. 3 | Thorough validation of TAM preferences and genome-editing efficiency of enNlovFz2 at endogenous loci in human cells. a, Validation of the TAM preferences for 16 TAM sequences at endogenous loci in WT-NlovFz2 and enNlovFz2. **b**, Summarizing the TAM preferences for 4 TAM sequences at 124 endogenous loci in WT-NlovFz2 and enNlovFz2. **c**, Comparison of genome-editing efficiency at 26 endogenous loci by WT-NlovFz2, enNlovFz2, enCnCas12f1 and enRhCas12f1 with 5′-NCCG TAM in HEK293T cells. Values and error bars are expressed as the mean ± s.e.m. ($n = 3$ independent biological replicates). **d**, Dot plot showing the genome-editing activity of WT-NlovFz2, enNlovFz2, enCnCas12f1 and enRhCas12f1. *P* values were derived using a two-sided Student's *t*-test. All mCherry$^+$ cells were sorted by FACS to assess the genome-editing efficiency of the enNlovFz2-ωRNA system at endogenous genomic loci. Adjusted P (P_{adj}) values are 0.000001, 0.0515, 0.000002 and 0.0223, respectively. For **a**, **b** and **d**, data were collected from endogenous sites and are exhibited as the mean ± s.e.m. Each data point represents the average editing efficiency measured at a given endogenous locus across three independent biological replicates.

https://doi.org/10.1038/s41589-025-01902-7

a

DNA and RNA bulge size = 0, mismatch number = 4 | DNA and RNA bulge size = 1, mismatch number = 2 or 3

	B2M	OT1	OT2	OT3	OT4	OT5	OT6	OT7	OT8	OT9	OT10	OT11	OT12	OT13	OT14	OT15	OT16	OT17	OT18	OT19	OT20
No. 1	35.14	0.04	0.02	0.02	0.06	0.03	0.02	0.02	0.13	0.02	0.02	0.02	0.02	0.03	0.04	0.01	0.03	0.01	0.01	0.03	0.01
No. 2	13.50	0.66	0.01	0.06		0.02	0.01	0.18	0.20	0.02	0.17	0.02	0.02	0.05	0.04	0.01	0.03	0.01	0.01	0.03	0.01
No. 3	4.45	0.04	0.01	0.03		0.01	0.02	0.08	0.12	0.02	0.06	0.02	0.02	0.02	0.04	0.01	0.03	0.01	0.02	0.02	0.01
No. 4	0.44	0.03	0.02	0.22	0.01	0.02	0.02	0.02	0.12	0.02	0.02	0.01	0.02	0.02	0.04	0.01	0.03	0.01	0.02	0.03	0.01

Indels (%)

DNA and RNA bulge size = 0, mismatch number = 4 or 5 | DNA and RNA bulge size = 1, mismatch number = 2 or 3

	CXCR4	OT1	OT2	OT3	OT4	OT5	OT6	OT7	OT8	OT9	OT10	OT11	OT12	OT13	OT14	OT15	OT16	OT17	OT18	OT19	OT20
No. 1	29.40	0.01	0.09	0.05	0.01	0.02		0.04	0.05	0.04	0.02	0.03	0.03	0.02	0.01	0.01	0.02	0.02	0.03	0.02	0.02
No. 2	17.40	0.01	0.07	0.10	0.01	0.01	0.02	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.03	0.01	0.02	0.03	0.01	0.01
No. 3	3.63		0.05	0.15	0.01	0.02		0.05	0.04	0.04	0.03	0.02	0.02	0.38	0.01	0.03	0.07	0.02	0.03	0.02	0.02
No. 4	0.44		0.06	0.05	0.02	0.01	0.01	0.03	0.04	0.04	0.03	0.02	0.01	0.21	0.01	0.03	0.08	0.02	0.03	0.02	0.01

Indels (%)

DNA and RNA bulge size = 0, mismatch number = 4 or 5 | DNA and RNA bulge size = 1 or 2, mismatch number = 2 or 3

	DMD-guide1	OT1	OT2	OT3	OT4	OT5	OT6	OT7	OT8	OT9	OT10	OT11	OT12	OT13	OT14	OT15	OT16	OT17	OT18	OT19	OT20
No. 1	7.06	0.01	0.08	0.03	0.01	0.01	0.04	0.04	0.02	0.02	0.17	0.01	0.02	0.02	0.03	0.02	0.02	0.02	0.05	0.07	0.02
No. 2	10.25	0.01	0.01	0.06	0.01	0.03	0.02	0.04	0.04	0.02	0.03	0.01	0.02	0.02	0.05	0.01	0.02	0.02	0.05	0.07	0.02
No. 3	0.08	0.01	0.10	0.02	0.01	0.15	0.09	0.03	0.01	0.03	0.04	0.01	0.02	0.03	0.06	0.03	0.02	0.02	0.06	0.10	0.03
No. 4	0.68	0.01	0.09	0.02	0.01	0.01	0.04	0.02	0.01	0.02	0.02	0.01	0.02	0.02	0.07	0.03	0.02	0.02	0.06	0.12	0.02

Indels (%)

No. 1: enNlovFz2 No. 2: enCnCas12f1 No. 3: enRhCas12f1 No. 4: 293T control

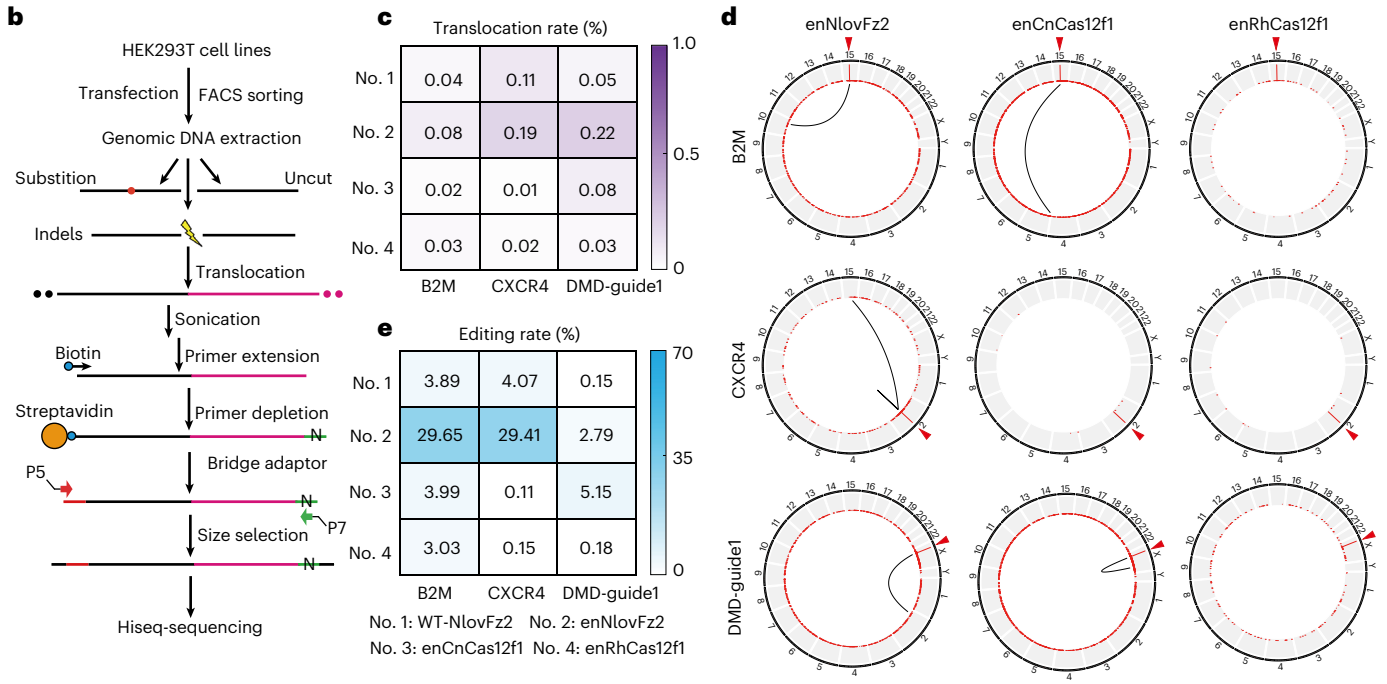


Fig. 4 | Specificities of enNlovFz2-mediated genome editing in human cells. **a**, Off-target analysis of enNlovFz2, enCnCas12f1 and enRhCas12f1 performed at the top 20 predicted off-target sites, with DNA or RNA bulge sizes ranging from 0 to 2 and mismatch numbers from 2 to 5, identified by Cas-OFFinder. Values are expressed as the mean ($n = 3$ independent biological replicates). **b**, Schematic diagram of PEM–seq (refer to Methods for details). **c**, Translocation rate of WT-NlovFz2, enNlovFz2, enCnCas12f1 and enRhCas12f1 targeting *B2M*, *CXCR4* and *DMD*-guide1 sites assessed using the PEM–seq pipeline, dividing the number of translocation events by the total number of editing events. **d**, PEM–seq quantifying genome-wide translocation efficiencies induced by off-target indels at *B2M*, *CXCR4* and *DMD*-guide1 sites with enNlovFz2, enCnCas12f1 and enRhCas12f1 nucleases, respectively. The circos plots depict off-target sites linked to the bait DSB (red triangle). **e**, Editing efficiency of WT-NlovFz2, enNlovFz2, enCnCas12f1 and enRhCas12f1 targeting *B2M*, *CXCR4* and *DMD*-guide1 sites assessed using the PEM–seq pipeline.

https://doi.org/10.1038/s41589-025-01902-7

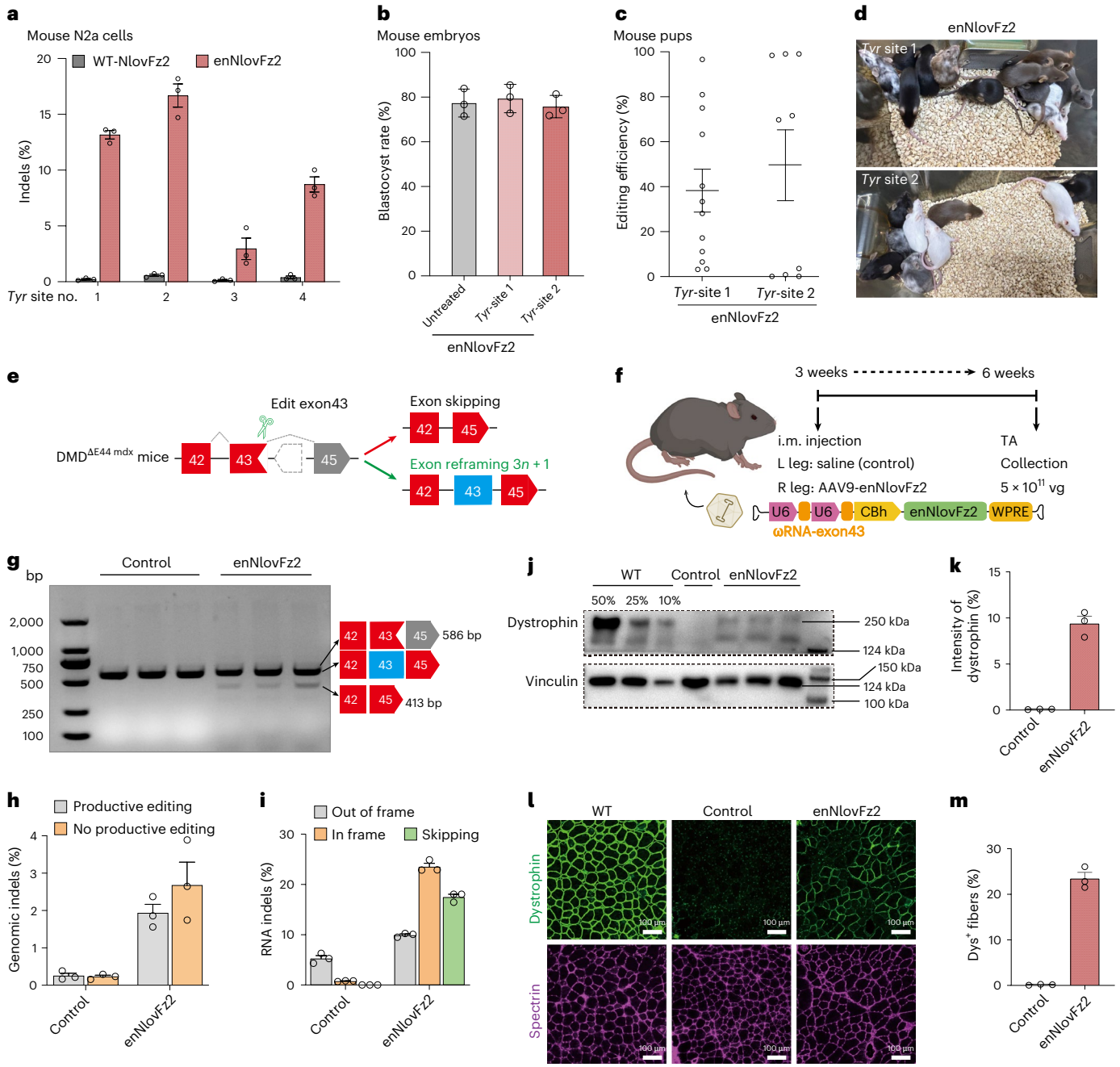


Fig. 5 | The enNlovFz2 mediates effective genome editing in mouse embryos and restores dystrophin expression in humanized DMD$^{\Delta E44\ mdx}$ mice after a single AAV injection. **a**, Comparison of the genome-editing efficiency of WT-NlovFz2 and enNlovFz2 at the *Tyr* gene in mouse N2a cells. **b**, Developmental rate from one-cell embryos to blastocyst stage for un-injected and injected embryos targeting *Tyr*-site1 and *Tyr*-site2 with enNlovFz2. **c**, Editing efficiency of enNlovFz2 at site1 and site2 of the *Tyr* gene in F0 mice. **d**, Coat color phenotype of mice edited at *Tyr*-site1 and *Tyr*-site2 by enNlovFz2. **e**, Schematic representation of the exon-skipping strategy to restore the DMD transcript's correct open reading frame using enNlovFz2. **f**, Schematics for the in vivo intramuscular (i.m.) injection of single AAV9-enNlovFz2 construct into tibialis anterior (TA) of the right (R) leg of 3-week-old DMD$^{\Delta E44\ mdx}$ mice. L, Left. **g**, Gel electrophoresis performed to analyze the RT–PCR products from the muscle of DMD$^{\Delta E44\ mdx}$ mice. **h**,**i**, Genomic (**h**) and RNA (**i**) indel editing events analyzed by targeted amplicon sequencing 3 weeks after i.m. injection. **j**, Western blotting analysis performed to assess dystrophin and vinculin expression in TA 3 weeks after injection with AAV9-enNlovFz2 or saline. The vinculin protein level was used as an internal control. **k**, Percentage of recovered dystrophin quantified using gray-scale analysis. **l**, Immunofluorescence staining for DMD showing the restoration of dystrophin expression 3 weeks after TA injection of enNlovFz2. The staining of dystrophin and spectrin proteins is depicted in green and purple, respectively. Scale bar, 100 μm. **m**, Quantification of Dys$^+$ fibers in cross-sections of TA. Values and error bars are expressed as the mean ± s.e.m. ($n = 3$ independent biological replicates). Each dot represents an individual mouse in **c**, **h**, **i**, **k** and **m**.

rescued by enNlovFz2-mediated genome editing (Fig. 5j–l) and the dystrophin protein levels in myofibers were restored to 22.53% of the WT control (Fig. 5m). Taken together, our results indicate that NlovFz2 can be engineered as a versatile genome-editing tool, providing a promising approach for both basic research and therapeutic applications.

Discussion

RNA-guided CRISPR–Cas and OMEGA systems, originating from prokaryotes, have been developed as versatile genome-editing tools. Recently, the eukaryotic Fanzor proteins (Fanzor1 and Fanzor2), which share remote homology with the OMEGA effector TnpB, have been

identified as RNA-guided programmable endonucleases[11,12]. Recent studies[11,13] have shown that Fanzor1 and ISDra2 TnpB share a bilobed structure, comprising REC and NUC domains. However, there are clear distinctions in the DNA recognition and cleavage mechanisms between these two families. Although TnpB takes advantage of the intrinsic tendency of DNA to unwind spontaneously[6], Fanzor1 proteins employ a loop insertion strategy to bridge the gap between the TAM and guide sequences, thereby initiating DNA unwinding[13]. The WED domain of the SpuFz1 structure contains three short α-helices, absent in the ISDra2 TnpB structure, which are essential for SpuFz1's recognition of the DNA duplex[11]. Unlike ISDra2 TnpB, where part of the RNA–DNA heteroduplex is exposed to the solvent, SpuFz1 features enlarged REC and RuvC domains that, together with their interactions with the ωRNA backbone, provide enhanced protection for the heteroduplex[5,6]. In addition, our structural comparison of NlovFz2 and SpuFz1, based on AlphaFold3 predictions, reveals significant differences in their ωRNA configurations: NlovFz2 contains three central stems, a PK and a triplex structure, whereas SpuFz1's ωRNA is simpler, consisting of only two primary stems and lacking both a PK and a triplex structure. In contrast, the NlovFz2 ωRNA may resemble the ωRNAs found in TnpB systems[5] and the sgRNAs present in certain miniaturized Cas12 systems[26]. Enriched in viruses and IS607 transposons, Fanzor2 is more closely related to TnpB than to other Fanzor1 proteins. This suggests that they probably originated from the phagocytosis of TnpB-containing bacteria by ameba, followed by spread through amoeba-trophic giant viruses[37].

Despite testing hypercompact Fanzor orthologs for genome editing in human cells, their extremely low editing efficiency and stringent TAM requirements have hindered their application. Through ωRNA and protein engineering, we developed the enNlovFz2 system, which demonstrated significantly higher genome-editing efficiency compared with WT-NlovFz2 and some reported Cas12f1 systems (enCnCas12f1 and enRhCas12f1) with the same 5′-NCCG TAM, and also showed the comparable gene-editing efficiency with the previously reported IS200 or IS605 transposon-encoded TnpB proteins (enISDra2 TnpB, TnpBmax and ISAba30 TnpB). Furthermore, the 5′-NMYG TAM preference of enNlovFz2 provides a promising solution to overcome the 5′-C-rich TAM constraint observed in WT-NlovFz2 (Fig. 3a,b and Extended Data Fig. 3b–e). Previous studies[22–24,28,38–40] have reported that the editing efficiency, targeting range and fidelity of CRISPR–Cas or OMEGA systems can be enhanced through structure-guided sgRNA or ωRNA design and protein engineering. These approaches enhance sgRNA or ωRNA stability and increase the interaction of nuclease protein with nucleic acids. Through systematic truncations and base-pair substitutions of the cognate ωRNA, we obtained three optimized ωRNA variants (UA97CG + S3Loop-Δ4U, AU103GC + S3Loop-Δ5U and S3-Δ1bp + S3Loop-Δ4U) that exhibited efficient genome-editing activity, achieving a fivefold increase compared with WT-ωRNA in human cells (Fig. 1d). Moreover, through protein engineering, we identified enhanced protein variants (Pro6Arg, Gln44Arg, Glu64Arg, Thr178Arg and Gln285Arg) that exhibited a 1.2-fold increase in editing activity compared with WT-NlovFz2 (Fig. 1e). Consistent with previous engineering efforts on the Cas12f and TnpB systems[18,21,23,41], our findings suggest that structural flaws in the ωRNA of NlovFz2 are key factors limiting its editing efficiency. In contrast, NlovFz2 protein engineering contributed less to genome-editing efficiency, probably because Fanzor family nucleases exhibit greater diversity in both protein and RNA structures compared with other nucleases, such as Cas9, Cas12a and IscB. We also observed structural defects in the ωRNA of other Fanzor nucleases, such as MmeFz2 and MmFNuc, including internal pentauridinylate (UUUUU) sequences and mismatched G–U base-pairs, which suggests that ωRNA modifications could be a more effective strategy than protein engineering for enhancing the genome-editing activity of Fanzor systems. In summary, by engineering ωRNA and NlovFz2 nuclease, we generated the enNlovFz2 system, which exhibited significantly improved editing activity and minimal off-target effect in mammalian cells.

Efficient in vivo delivery of large Cas9 or Cas12a nucleases (>1,000 amino acids) and their derived base or prime editors using AAVs is highly challenging as a result of its limited cargo capacity (approximately 4.7 kb)[42,43]. NlovFz2, with only 489 amino acids, is a highly sought-after genome-editing enzyme for gene therapy using single AAV delivery. In the present study, we employed the engineered enNlovFz2 for in vivo gene-editing therapy in mice with DMD, demonstrating its efficiency as a miniature nuclease in modulating exon skipping and restoring dystrophin expression. Since the discovery of eukaryotic Fanzor proteins, our work has shown extensively the rational optimization of Fanzor2 (NlovFz2) for achieving excellent gene-editing performance both in vitro and in vivo. Our study provides a solid foundation for engineering eukaryotic OMEGA systems, paving the way for their broader and more accessible use in research and therapeutic applications. Considering its hypercompact size, enNlovFz2 is particularly attractive for developing various derivative genome editing tools, including base editing[44,45], prime editing[46], retron editing[47] and epigenome editing[48]. We envision that these genome-editing tools could be developed by coupling the dead NlovFz2 protein with context-specific deaminases or other architectural components. In summary, our study demonstrated the enhanced gene-editing activity of NlovFz2 through ωRNA and protein engineering in mammalian cells and mouse embryos, showcasing its disease correction ability in animal models and indicating the potential of the hypercompact NlovFz2-ωRNA system as an effective miniature gene-editing platform for diverse gene therapeutic and cell engineering 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/s41589-025-01902-7.

References

1. Makarova, K. S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
2. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
3. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
4. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
5. Sasnauskas, G. et al. TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature* **616**, 384–389 (2023).
6. Nakagawa, R. et al. Cryo-EM structure of the transposon-associated TnpB enzyme. *Nature* **616**, 390–397 (2023).
7. Meers, C. et al. Transposon-encoded nucleases use guide RNAs to promote their selfish spread. *Nature* **622**, 863–871 (2023).
8. Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
9. Schuler, G., Hu, C. & Ke, A. Structural basis for RNA-guided DNA cleavage by IscB-omegaRNA and mechanistic comparison with Cas9. *Science* **376**, 1476–1481 (2022).
10. Hirano, S. et al. Structure of the OMEGA nickase IsrB in complex with omegaRNA and target DNA. *Nature* **610**, 575–581 (2022).
11. Saito, M. et al. Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature* **620**, 660–668 (2023).
12. Jiang, K. et al. Programmable RNA-guided DNA endonucleases are widespread in eukaryotes and their viruses. *Sci. Adv.* **9**, eadk0171 (2023).

13. Xu, P. et al. Structural insights into the diversity and DNA cleavage mechanism of Fanzor. *Cell* **187**, 5238–5252.e5220 (2024).
14. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
15. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
16. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).
17. Bao, W. & Jurka, J. Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mob. DNA* **4**, 12 (2013).
18. Li, Z. et al. Engineering a transposon-associated TnpB-ωRNA system for efficient gene editing and phenotypic correction of a tyrosinaemia mouse model. *Nat. Commun.* **15**, 831 (2024).
19. Marquart, K. F. et al. Effective genome editing with an enhanced ISDra2 TnpB system and deep learning-predicted ωRNAs. *Nat. Methods* **21**, 2084–2093 (2024).
20. Xiang, G. et al. Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. *Nat. Biotechnol.* **42**, 745–757 (2024).
21. Su, M. et al. Molecular basis and engineering of miniature Cas12f with C-rich PAM specificity. *Nat. Chem. Biol.* **20**, 180–189 (2024).
22. Kong, X. et al. Engineered CRISPR-OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. *Nat. Commun.* **14**, 2046 (2023).
23. 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 (2022).
24. Han, D. et al. Development of miniature base editors using engineered IscB nickase. *Nat. Methods* **20**, 1029–1036 (2023).
25. Chen, W. et al. Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Mol. Cell* **83**, 2768–2780.e6 (2023).
26. Wu, Z. et al. Structure and engineering of miniature *Acidibacillus sulfuroxidans* Cas12f1. *Nat. Catal.* **6**, 695–709 (2023).
27. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
28. 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).
29. Zhang, H. et al. An engineered xCas12i with high activity, high specificity, and broad PAM range. *Protein Cell* **14**, 538–543 (2023).
30. Yang, Y. et al. Highly efficient and rapid detection of the cleavage activity of Cas9/gRNA via a fluorescent reporter. *Appl. Biochem. Biotechnol.* **180**, 655–667 (2016).
31. Leenay, R. T. et al. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* **62**, 137–147 (2016).
32. 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).
33. Liu, Y. et al. PEM-seq comprehensively quantifies DNA repair outcomes during gene-editing and DSB repair. *STAR Protoc.* **3**, 101088 (2022).
34. Olson, E. N. Toward the correction of muscular dystrophy by gene editing. *Proc. Natl Acad. Sci. USA* **118**, e2004840117 (2021).
35. Min, Y. L. et al. CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Sci. Adv.* **5**, eaav4324 (2019).
36. Long, C. et al. Correction of diverse muscular dystrophy mutations in human engineered heart muscle by single-site genome editing. *Sci. Adv.* **4**, eaap9004 (2018).
37. Boyer, M. et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl Acad. Sci. USA* **106**, 21848–21853 (2009).
38. Xu, X. et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* **81**, 4333–4345.e4 (2021).
39. Han, L. et al. Engineering miniature IscB nickase for robust base editing with broad targeting range. *Nat. Chem. Biol.* **20**, 1629–1639 (2024).
40. Yan, H. et al. Assessing and engineering the IscB-ωRNA system for programmed genome editing. *Nat. Chem. Biol.* **20**, 1617–1628 (2024).
41. Wang, Y. et al. Guide RNA engineering enables efficient CRISPR editing with a miniature *Syntrophomonas palmitatica* Cas12f1 nuclease. *Cell Rep.* **40**, 111418 (2022).
42. Wang, D., Zhang, F. & Gao, G. CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. *Cell* **181**, 136–150 (2020).
43. Raguram, A., Banskota, S. & Liu, D. R. Therapeutic in vivo delivery of gene editing agents. *Cell* **185**, 2806–2827 (2022).
44. 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).
45. 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).
46. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
47. Kong, X. et al. Precise genome editing without exogenous donor DNA via retron editing system in human cells. *Protein Cell* **12**, 899–902 (2021).
48. Nuñez, J. K. et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503–2519.e2517 (2021).

Methods

Structure prediction by AlphaFold3

The sequences of the WT-NlovFz2 protein and its cognate full-length ωRNA with a 20-nt *B2M* guide, along with a 40-bp endogenous *B2M* target DNA, were submitted to the AlphaFold3 online web server (https://golgi.sandbox.google.com), to generate a predicted ternary complex structure. The resulted structure was fine-tuned by using COOT. Molecular visualization figures were generated using CueMol (http://www.cuemol.org).

Plasmid construction

NlovFz2 and ωRNA sequences were synthesized by HuaGene Co., Ltd. and cloned to generate the plasmid construct named pCAG_NLS-NlovFz2-NLS_pA_pU6_ωRNA scaffold-2x Eco31I_pCMV_mCherry_pA, using the *pEASY*-Basic Seamless Cloning and Assembly Kit from TransGen Biotech. The ωRNA oligos were annealed and ligated into the Eco31I cleavage site of the construct. The ωRNA sequences are listed in Supplementary Table 1.

To generate NlovFz2-ωRNA mutants, we divided the three stem loops of NlovFz2-ωRNA into three fragments. These fragments were subsequently replaced with Eco31I recognition sequences using PCR and Gibson assembly techniques, utilizing the *pEASY*-Basic Seamless Cloning and Assembly Kit. This process facilitated the creation of three backbone mutants for the ωRNA scaffold. Subsequently, specific mutations were introduced by incorporation of annealed oligos containing the desired mutations through Eco31I digestion and ligation with T4 DNA ligase. For the generation of NlovFz2 mutants, the protein sequence was divided into 29 distinct segments, each comprising 17 amino acid residues. The methodology employed to obtain NlovFz2 mutants followed the same approach for generating ωRNA mutations. The protein sequences are listed in Supplementary Table 2.

NlovFz2-ωRNA RNP purification

E. coli BL21-AI cells containing the pDuet-NlovFz2/ωRNA plasmid were incubated at 37 °C with vigorous shaking until the optical density at 600 nm reached 0.8. The expression of NlovFz2 and ωRNA was induced by adding isopropyl β-D-1-thiogalactopyranoside and arabinose to final concentrations of 0.25 mM and 2% (w:v), respectively, and the incubation was continued at 16 °C overnight. Cells were collected by centrifugation and resuspended in buffer A (20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 10 mM $MgCl_2$ and 15 mM imidazole). The cells were then disrupted by high-pressure homogenization and clarified by centrifugation. The clear lysate was loaded on to a 5-ml HisTrap HP Ni-NTA column pre-equilibrated with buffer A. Unbound proteins were washed away with buffer A containing 50 mM imidazole and the His-tagged NlovFz2-ωRNA RNP was eluted with buffer A containing 250 mM imidazole. The eluted RNP was concentrated to 1 ml by ultrafiltration and then loaded on to a HiLoad 16/600 Superdex 200-pg column pre-equilibrated with buffer A without imidazole. Peak fractions were merged and concentrated to ~0.2 mg ml^{-1} for the cleavage assay. The entire RNP purification procedure was carried out at 4 °C.

In vitro cleavage assays

The dsDNA substrates were produced by PCR amplification of pUC19 plasmids or synthesized DNA fragments containing the target loci and TAM sequences. ssDNA substrates were ordered as FAM-labeled oligonucleotides. Target cleavage assays were performed in 10 µl of reaction mixture containing 5 nM substrate and 1 µl of RNP in 1× reaction buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl and 10 mM $MgCl_2$). The assay was incubated at 37 °C for 30 min. After the incubation, reactions were treated with RNase A and proteinase K. The cleavage products were analyzed using 2% agarose gels or 15% TBE (tris base, boric acid and EDTA)–polyacrylamide gels.

TAM depletion assay and analysis

A 6-N TAM library was constructed by PCR amplifying a pUC19-based plasmid with a 20-bp *B2M* sequence as the target for NlovFz2 and ligated by Golden Gate Assembly. The 6-N TAM library was in vitro cleaved by NlovFz2/ωRNA RNP at 37 °C for 30 min. The digest product underwent two-step PCR amplifications to add barcodes and indices, for next-generation sequencing (NGS) on an Illumina Novaseq 6000 sequencing platform in HaploX Genomics Center. TAM randomized regions were extracted and the abundance of each TAM was counted and normalized to total reads. Depletion values were computed by calculating the ratio compared with the abundance in RNP-free control sample.

Cell culture, transfection and flow cytometry analysis

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% penicillin–streptomycin in a humidified incubator at 37 °C with 5% CO_2. Then, 1×10^5 cells were seeded on to 24-well plates (Corning) before transfection. For screening experiments involving NlovFz2 mutants, cells were co-transfected with 1.5 µg of the reporter plasmid and the NlovFz2 plasmid at a 1:1 molar ratio using poly(etherimide) (PEI). After 48 h, the activation efficiency of EGFP fluorescence was evaluated using a BD FACS Aria III. Data analysis was performed using FlowJo X (v.10.0.7). For determining genome-editing efficiency at endogenous loci, cells were transfected with 1.5 µl of PEI and 0.75 µg of NlovFz2 plasmid. The top 30% or all $mCherry^+$ cells were sorted by FACS 60–72 h post-transfection (Supplementary Fig. 7).

Mouse N2a cells were cultured under the same conditions as human HEK293T cells. For screening efficient guide sequences targeting the mouse *Tyr* gene, N2a cells were seeded in 12-well plates at 30–50% confluence. After 18 h of incubation, 1.5 µg of plasmid was transfected using Lipofectamine 3000 (Invitrogen) following the manufacturer's instructions. After the transfected cells were cultured for ~60–72 h, we carefully resuspended the cell pellet and sorted all $mCherry^+$ cells using FACS for subsequent genome-editing evaluation by targeted amplicon sequencing.

Targeted amplicon sequencing and analysis

Approximately 6,000 sorted cells were lysed in 20 µl of lysis buffer containing proteinase K (Vazyme Biotech), following the manufacturer's manual. To analyze the in vivo gene-editing efficiency of enNlovFz2, DNA was extracted from successfully born enNlovFz2-edited mouse tails or muscle tissues using TIANamp Genomic DNA Kit (TIANGEN). For targeted amplicon sequencing, genomic regions were amplified by nested PCR with Phanta Max Super-Fidelity DNA Polymerase (Vazyme) using primers containing barcodes. The PCR products were pooled and purified using a Gel Extraction Kit (Omega). The amplicon-seq libraries were prepared using the VAHTS Universal DNA Library Prep Kit (Vazyme), purified and then sequenced on an Illumina NovaSeq 6000 platform using the PE150 strategy. The sequencing data were initially demultiplexed using Cutadapt (v.2.8) and then processed by CRISPResso2 (v.2.0.20b)[49] to quantify genome-editing efficiency (refer to Supplementary Table 3 for the target site sequence and primer information).

Animals

All mouse experiments were approved by the Biomedical Research Ethics Committee of HuidaGene Therapeutics Co. Ltd. The mice were housed in a controlled barrier facility with a 12-h light:dark cycle, maintained at temperatures ranging from 18 °C to 23 °C and humidity levels between 40% and 60%. Food and water were readily available at all times. $DMD^{\Delta E44\,mdx}$ mice were generated on the C57BL/6J background using the CRISPR–Cas9 system. Considering that DMD is the most common sex-linked lethal disease in humans, male mice were selected for this investigation.

In vitro transcription of enNlovFz2 mRNA and *Tyr*-ωRNAs

The enNlovFz2 plasmids were linearized using the FastDigest EcoRI restriction enzyme (Thermo Fisher Scientific), purified with Gel Extraction Kit and employed as the template for in vitro transcription using the mMESSAGE mMACHINE T7 Ultra kit (Life Technologies). The T7 promoter was integrated into the ωRNA template by PCR amplification with a specific primer pair. The PCR products purified with the Gel Extraction Kit as templates were transcribed using the MEGAshortscript T7 Kit (Life Technologies). The enNlovFz2 mRNA and *Tyr*-ωRNAs were purified using the MEGAclear Kit (Life Technologies) and eluted in RNase-free water. In vitro transcribed RNAs were aliquoted and stored at −80 °C until use. Before microinjection, the mixture of enNlovFz2 mRNA and *Tyr*-ωRNA was prepared by centrifugation for 10 min at 18,000*g* and 4 °C and the supernatant was transferred to 0.2-ml RNA-free PCR tubes for injection.

Zygote microinjection and embryo transplantation

The 8-week-old superovulated B6D2F1 females were mated with B6D2F1 male mice and fertilized embryos were collected from the oviducts 21 h after human chorionic gonadotropin injection. The mixture of enNlovFz2 mRNA (50 ng μl^{-1}) and *Tyr*-ωRNAs (100 ng μl^{-1}) was injected into the cytoplasm of fertilized eggs in a droplet of M2 medium using a FemtoJet microinjector (Eppendorf) with constant flow settings. The injected embryos were cultured overnight in M16 medium supplemented with amino acids at 37 °C, under 5% CO_2 in a humidified incubator, and then transferred into the oviducts of pseudo-pregnant, 8-week-old ICR foster mothers at 0.5 d post-coitus.

The gRNA-dependent off-target analysis

The specificity of the NlovFz2 system was evaluated by predicting potential off-target sites using CRISPR RGEN Tools (Cas-OFFinder, http://www.rgenome.net/cas-offinder). The search queries covered both 'CCG' and a 20-nt target sequence. For the research, the PAM sequence was set to 'NNN' and the DNA or RNA bulge to 1–2, allowing up to 5 mismatches. All other parameters were left at their default settings. The off-target sites for each guide sequence were selected in ascending order of mismatch numbers. Potential off-target sites with one or more mismatches were chosen for primer design using the online Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast). The top 20 predicted potential off-target sites were amplified and sequenced to assess the specificity of enNlovFz2, enCnCas12f1 and enRhCas12f1. All predicted off-target site sequences and corresponding primers are provided in Supplementary Table 4.

PEM–seq analysis

Genome-wide off-target analysis was conducted following a previously established PEM–seq protocol[33,50]. In brief, expression plasmids encoding NlovFz2-WT, enNlovFz2, enCnCas12f1 and enRhCas12f1 targeting the *B2M*, *CXCR4* and *DMD*-guide1 genes were transfected into HEK293T cells using PEI. After 72 h, all positively transfected cells were harvested for DNA extraction. Genomic DNA (20 µg) was fragmented using a Covaris sonicator to produce fragments ranging from 300 bp to 700 bp in length. Biotin labeling of DNA fragments was achieved by one-round PCR extension using biotinylated primers at the 5′-end, followed by primer removal using AMPure XP beads and purification with streptavidin beads. The ssDNA bound to streptavidin beads was ligated with a bridge adapter containing a 14-bp random molecular barcode. The PCR product was generated via nested PCR to enrich DNA fragments containing the bait dsDNA break (DSB) and tagged with illumine adapter sequences. The prepared sequencing library underwent high-throughput sequencing on the Hi-seq 2500 platform with PE150 strategy. PEM–seq data analysis was conducted using the PEM-Q pipeline (https://github.com/liumz93/PEM-Q) with default parameters. Primers for amplifying the target genomic DNAs are detailed in Supplementary Table 5.

AAV9 production and delivery to $DMD^{\Delta E44\ mdx}$ mice

The AAVs used in the present study were produced by HuidaGene Therapeutics Inc. and purified using iodixanol density gradient centrifugation. The $DMD^{\Delta E44\ mdx}$ mice were generated by crossing humanized $DMD^{\Delta E44}$ mice with mdx mice carrying a stop mutation (p.Gln993*) in murine exon 23 on chromosome X. For intramuscular administration, 3-week-old $DMD^{\Delta E44\ mdx}$ mice were anesthetized and tibialis anterior (TA) was injected with either 50 µl of AAV9 (5×10^{11} vector genome (vg)) preparations or an equivalent volume of saline solution. Then, 3 weeks after injection, mice were anesthetized and euthanized and tissues were dissected into distinct segments for targeted assessment. Specifically, the distal region was designated for evaluating DNA editing and exon skipping efficiency, the middle portion was allocated for western blotting analysis of dystrophin expression and the proximal segment was reserved for immunofluorescent analysis of dystrophin levels.

Western blotting analysis

Muscle samples were homogenized in radioimmunoprecipitation buffer supplemented with a protease inhibitor cocktail. Lysate supernatants were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) and then adjusted to a uniform concentration with water. Subsequently, equal amounts of the samples were mixed with NuPAGE LDS sample buffer (Invitrogen) and 10% β-mercaptoethanol and then boiled at 70 °C for 10 min. Then, 10 µg of total protein per lane was loaded on to a 3–8% tris-acetate gel (Invitrogen) and subjected to electrophoresis at 200 V for 1 h. The proteins were transferred to poly(vinylidene fluoride) membranes for 3.5 h at 350 mA under wet conditions. Afterward, the membranes were blocked in 5% nonfat milk in Tris-buffered saline with Tween 20 (TBST) and probed with primary antibodies specific to the target proteins. After three washes with TBST, the membranes were incubated with horseradish peroxidase-conjugated secondary antibodies (1:1,000 dilution, Beyotime, cat. no. A0216) specific to the immunoglobulin G (IgG) of the species from which the primary antibodies against dystrophin (1:1,000 dilution, Sigma-Aldrich, cat. no. D8168) or vinculin (1:1,000 dilution, CST, cat. no. 13901s) were obtained. Finally, the target proteins were visualized using chemiluminescent substrates (Invitrogen, cat. no. WP20005).

Histology and immunofluorescence

For histological analysis, tissue samples embedded in paraffin were first deparaffinized in xylene, followed by rehydration with ethanol gradient ranging from 100% to 50%. Subsequently, the sections were washed in distilled water and stained with hematoxylin and eosin and Picro-Sirius Red solution (0.1%) for histological examination.

For immunofluorescence, tissues were embedded in an optimal cutting temperature compound and snap-frozen in liquid nitrogen. Sequential frozen sections (10-µm thick) were fixed at 37 °C for 2 h and then permeabilized with phosphate-buffered saline (PBS) containing 0.4% Triton X-100 for 30 min. Afterward, the specimens were blocked with 10% goat serum for 1 h at room temperature after rinsing with PBS. The slides were incubated overnight with primary antibodies against dystrophin (1:100 dilution, Abcam, cat. no. ab15277) and spectrin (1:500 dilution, Millipore, cat. no. MAB1622) at 4 °C. The next day, the samples were thoroughly washed with PBS and incubated with compatible secondary antibodies (Alexa Fluor-488 AffiniPure donkey anti-rabbit IgG (1:1,000 dilution, Jackson ImmunoResearch Labs, cat. no. 711-545-152) or Alexa Fluor-647 AffiniPure donkey anti-mouse IgG (1:1,000 dilution, Jackson ImmunoResearch Labs, cat. no. 715-605-151)) along with DAPI for 2 h at room temperature. After a 15-min PBS wash, the slides were sealed using a Fluoromount-G mounting medium. The proportion of dystrophin-positive (Dys^+) muscle fibers is calculated as a percentage of the total spectrin-positive muscle fibers.

Statistical analysis
Data from two or three biological replicate experiments are presented as the mean ± s.e.m. Statistical analyses were conducted using GraphPad Prism 10 (v.9.5.1) with an unpaired, two-tailed Student's *t*-test. A $P < 0.05$ was considered to be statistically significant.

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

Data availability
NGS data are available at the National Center for Biotechnology Information's Sequence Read Archive database under the BioProject accession no. PRJNA1147587. Source data are provided with this paper.

References
49. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
50. Yin, J. et al. Optimizing genome editing strategy by primer-extension-mediated sequencing. *Cell Discov.* **5**, 18 (2019).

Acknowledgements
We appreciate the support from the Gene Editing Scientific Teaching (NWAFU-GEST), High-Performance Computing and Life Science Research Core Service platforms (K. R. Huang, X. R. Liu, L. Chen, M. Zhou and L. Q. Li) at Northwest A&F University (NWAFU). We also thank the members of HuidaGene Therapeutics Co., Ltd. for their contributions in supplying experimental materials and insightful discussions. This work is supported by the Biological Breeding-Major Projects (grant nos. 2023ZD04074 to K.X., 2023ZD04051 to Y.W. and 2022ZD04014 to X.W.), the National Natural Science Foundation of China (grant nos. 32441080 and 32301251 to Y.W., 22207074 to Z.W. and 32272848 to X.W.), the National Science and Technology Major Project of China (grant no. 2023ZD0500500 to Z.W.), the China Agricultural Research System (grant no. CARS-39-03 to X.W.) and local grants (grant nos. QCYRCXM-2023-104 and Z1013023006 to Y.W., 22YF1428100 to Z.W. and 2023A02011-2-4 and 2022GDTSLD-46 to X.W.).

Author contributions
Y.W. and X.W. conceived the project. Y.W., Z.W. and K.X. designed the experiments. Y.W., P.G., S.L. and G.L. performed data analysis. Z.W. and D.P. conducted the structural prediction and biochemical assays. P.G., Y.F.C., H.J., Y.Y., S.L., Z.W., Z.L. and M.Z. performed cell transfection and FACS. Y.W. and G.L. performed animal experiments. Y.W. and Z.W. wrote the manuscripts. Y.W., X.W., Z.W., Y.L.C. and K.X. supervised the project.

Competing interests
Y.W., X.W. and K.X. have filed a patent application related to this work through NWAFU (patent no. 202411627205.2). The other authors declare no competing interests.

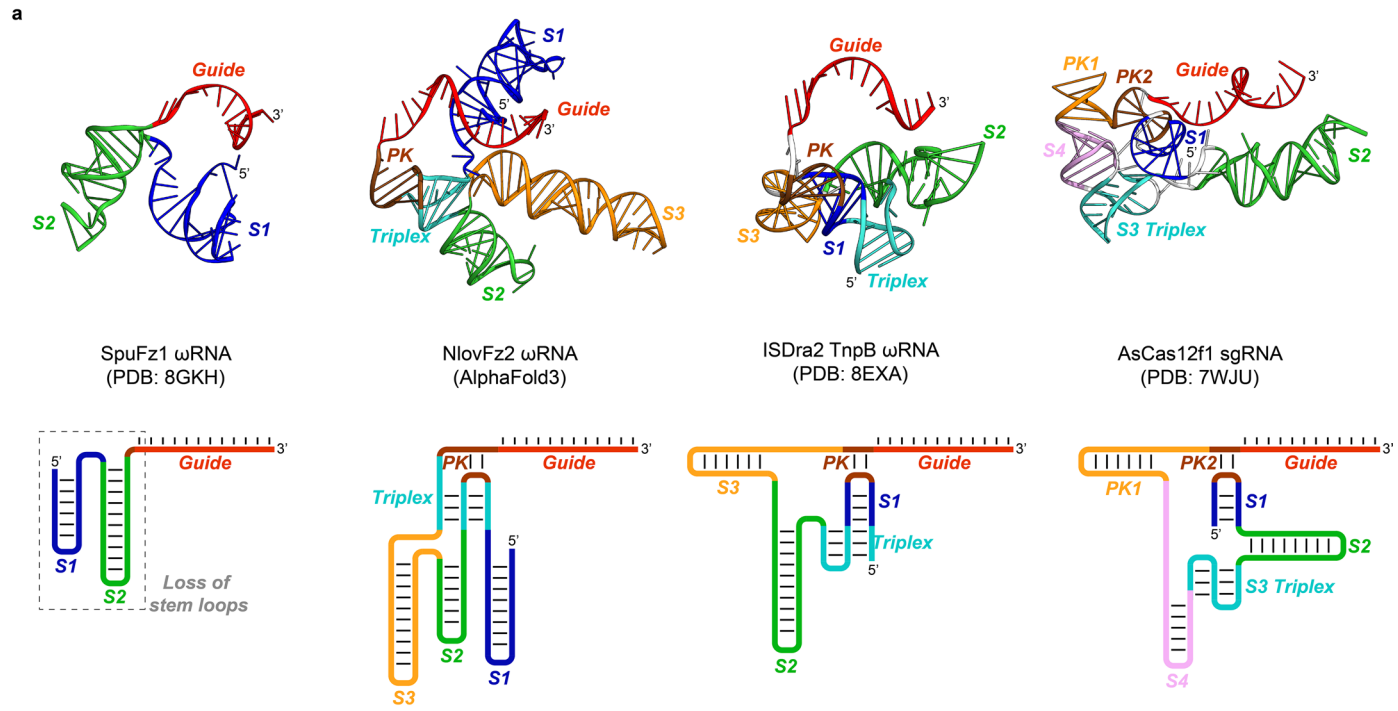
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41589-025-01902-7.

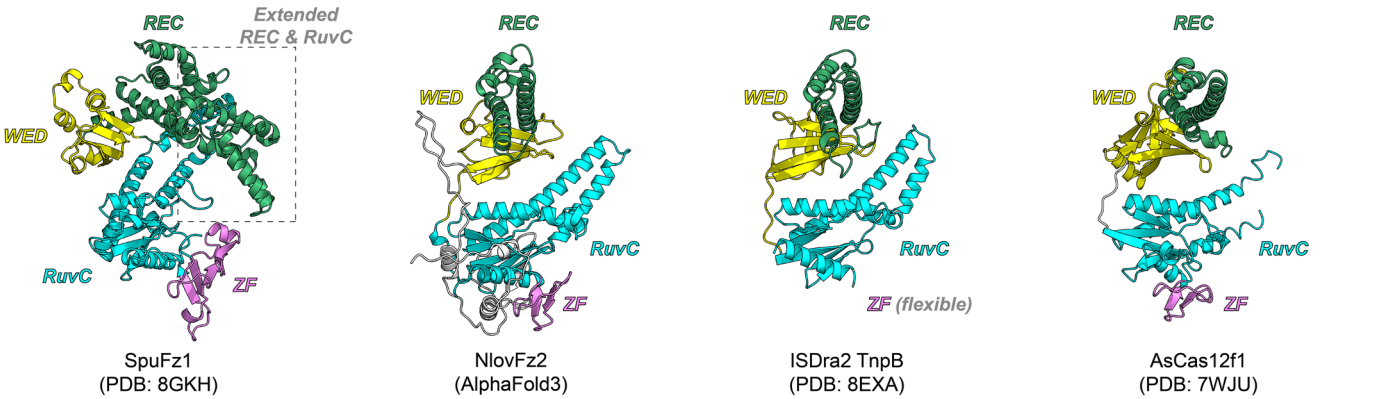
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-025-01902-7.

Correspondence and requests for materials should be addressed to Yinghui Wei, Kun Xu, Zhaowei Wu or Xiaolong Wang.

Peer review information *Nature Chemical Biology* thanks Chunyi Hu, Shahid Mansoor and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

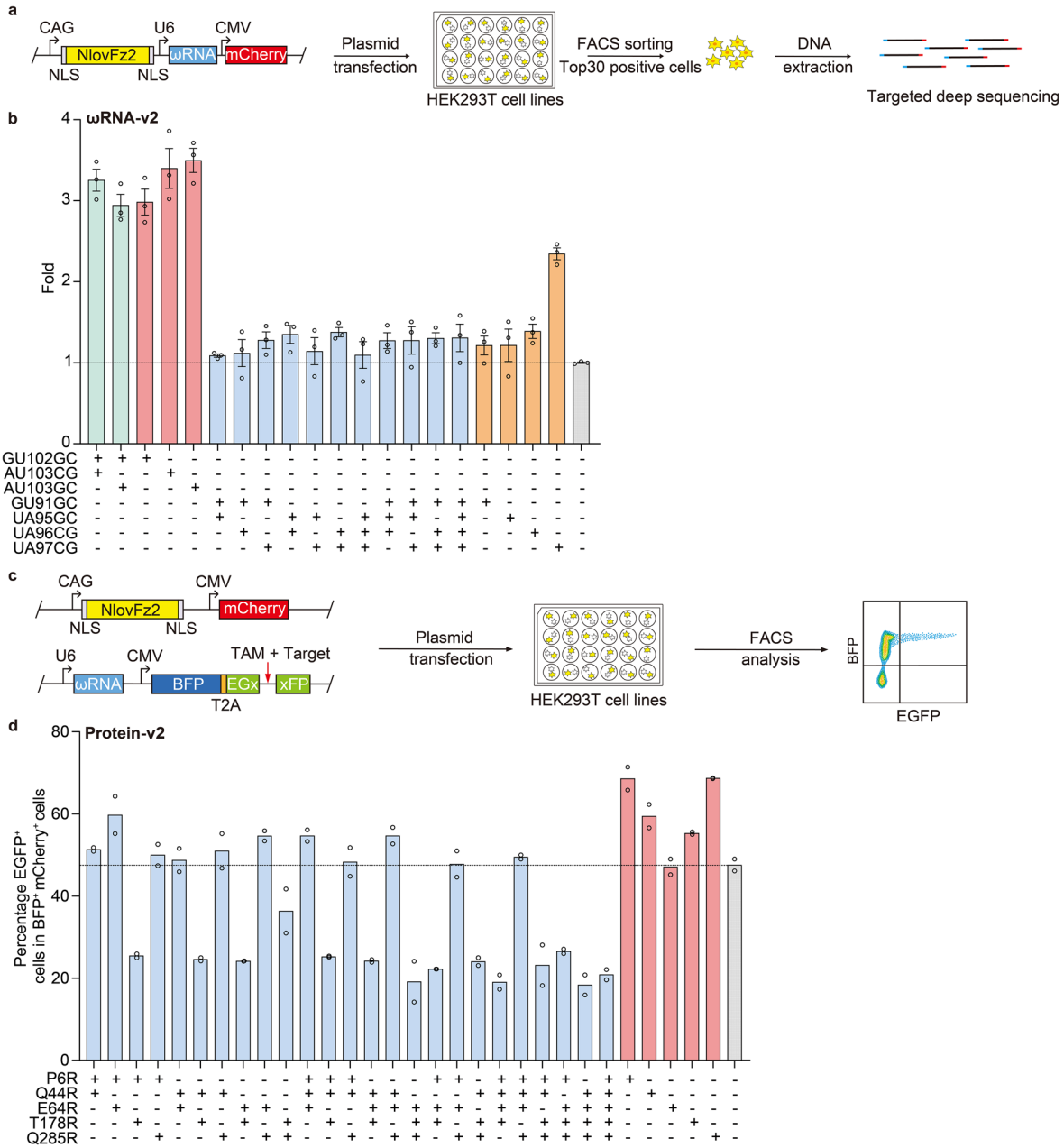
https://doi.org/10.1038/s41589-025-01902-7





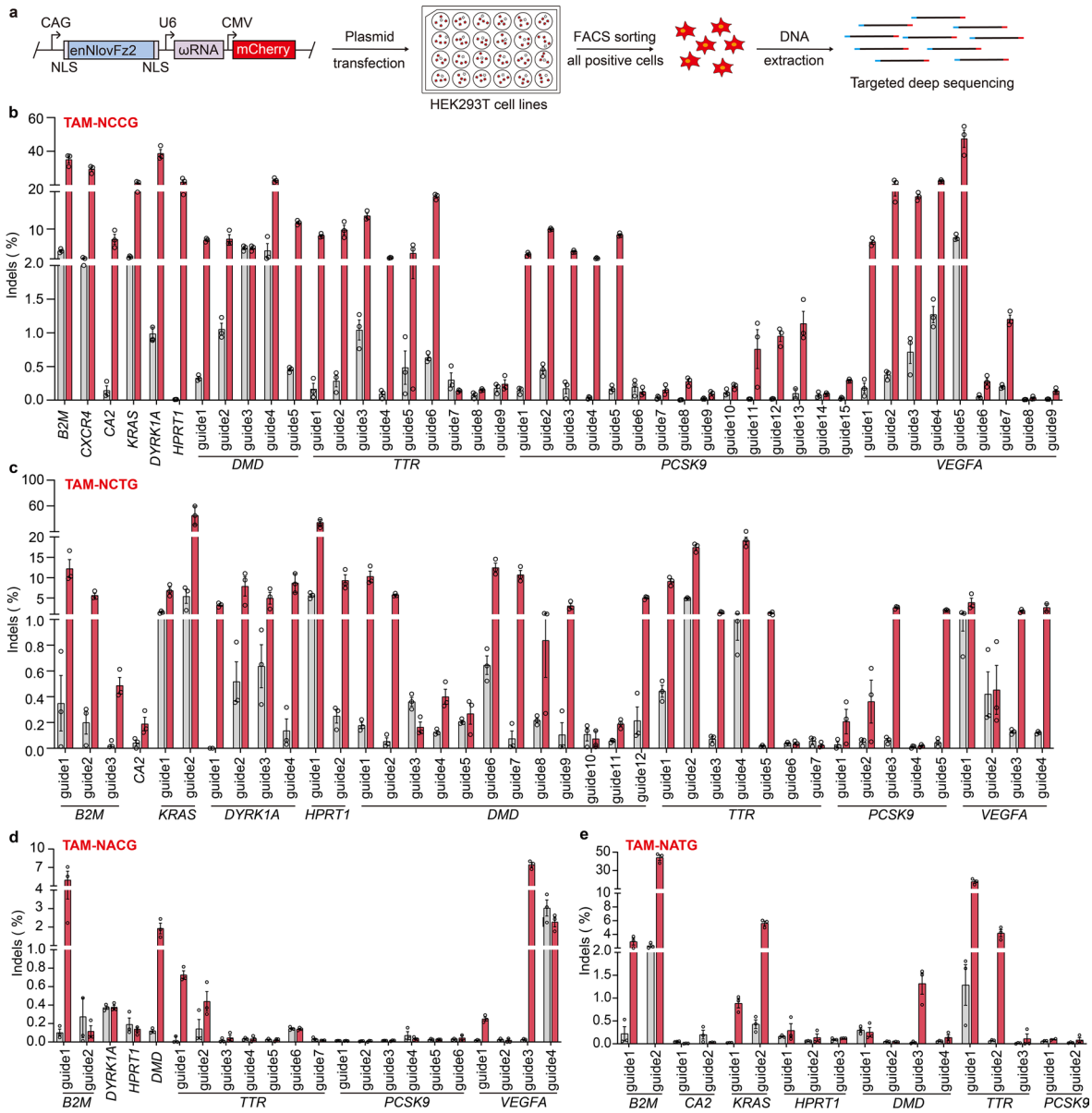
Extended Data Fig. 1 | Structural comparisons of the Fanzor1-ωRNA, Fanzor2-ωRNA, TnpB-ωRNA, and CRISPR-Cas12f1 systems. a, Experimental or predicted structures of SpuFz1-ωRNA (PDB: 8GKH), NlovFz2-ωRNA (AlphaFold3), ISDra2-ωRNA (PDB: 8EXA), and AsCas12f1-sgRNA (PDB: 7WJU), along with their topological schematics. **b**, Experimental or predicted structures of SpuFz1 (PDB: 8GKH), NlovFz2 (AlphaFold3), ISDra2 (PDB: 8EXA), and AsCas12f1 (PDB: 7WJU).

https://doi.org/10.1038/s41589-025-01902-7



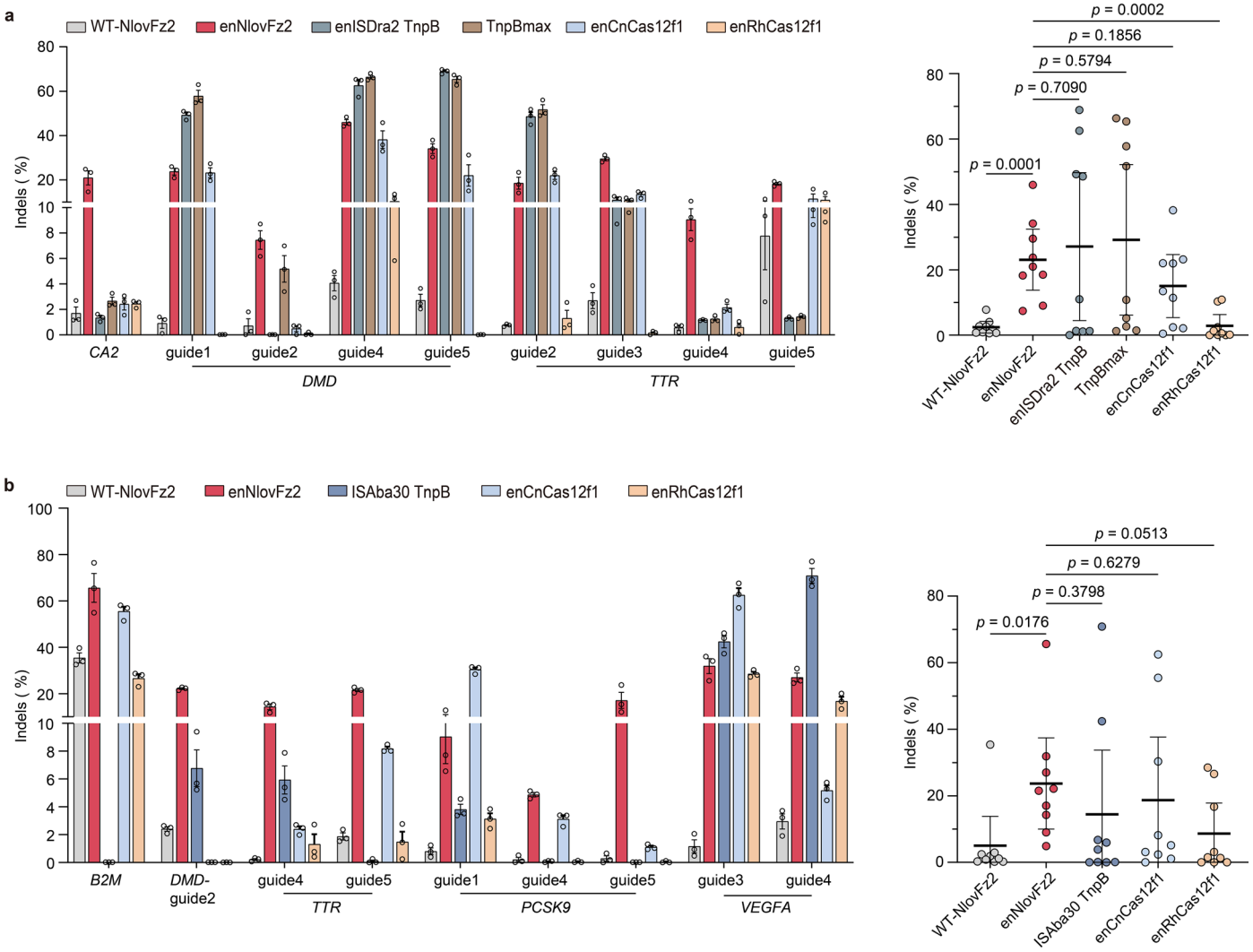
Extended Data Fig. 2 | Engineering NlovFz2 ωRNA and protein to enhance editing efficiency in human cells. a, Experimental workflow for detecting NlovFz2 genome-editing activity at the endogenous *B2M* locus by optimizing the structure of the ωRNA scaffold in HEK293T cells. **b**, Second round of ωRNA optimization involves combining high-efficiency variants of A-U and G-U base pair substitutions, building upon the results from the first round. Fold-change represents the ratio of the editing efficiency of ωRNA variants to that of WT-ωRNA. Values and error bars are expressed as mean ± SEM (n = 3 independent biological replicates). **c**, Schematics illustrating the detection of indel activity based on the fluorescence signal of the EGFP reporter activation in HEK293T cells. **d**, EGFP activation efficiency was evaluated by combining single NlovFz2 mutations (P6R, Q44R, E64R, T178R, Q285R) with an ωRNA variant (S3Loop-Δ4U + UA97CG). The dashed line indicates the cleavage activity of WT-NlovFz2.

https://doi.org/10.1038/s41589-025-01902-7



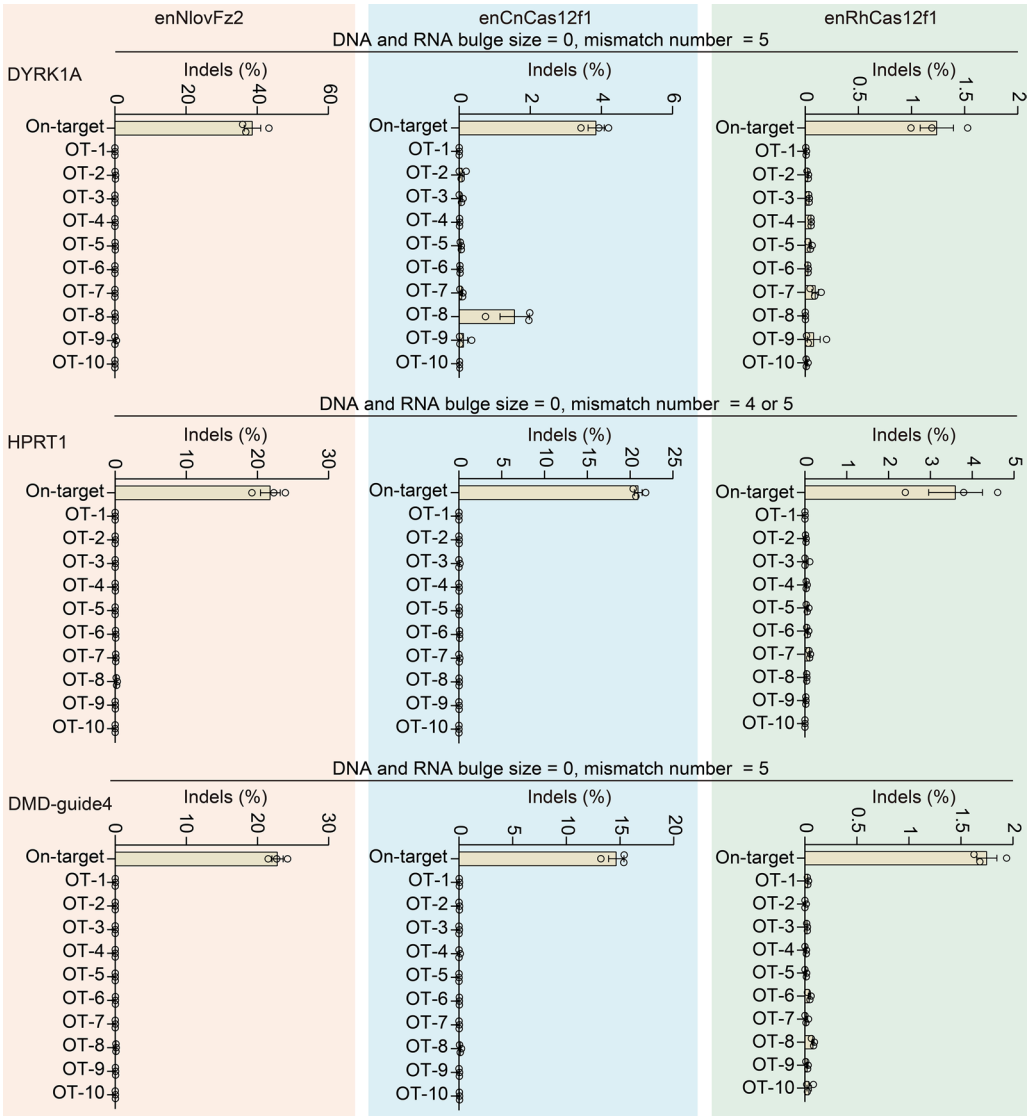
Extended Data Fig. 3 | Comprehensive validation of TAM preferences of WT-NlovFz2 and enNlovFz2 at endogenous loci in human cells. a, Schematics illustrating the detection of nuclease genome-editing activity at endogenous loci in HEK293T cells. **b**-**e**, Comparing the genome-editing efficiency of WT-NlovFz2 and enNlovFz2 with different TAM preferences (5′-NCCG, **b**; 5′-NCTG, **c**; 5′-NACG, **d**; 5′-NATG, **e**) at endogenous loci in human cells. Values and error bars are expressed as mean ± SEM (n = 3 independent biological replicates).

https://doi.org/10.1038/s41589-025-01902-7



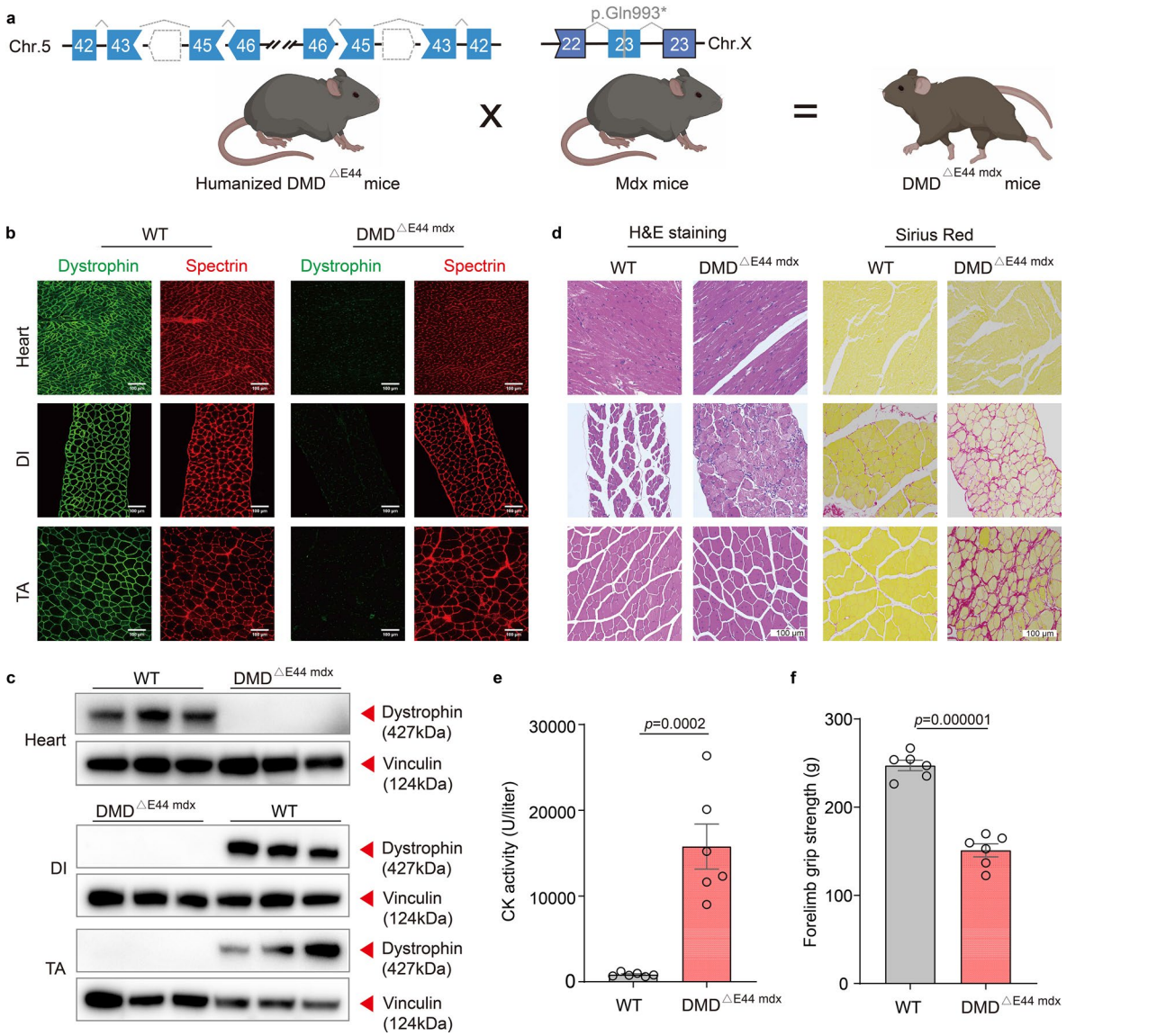
Extended Data Fig. 4 | Comparison of genome-editing efficiency among WT-NlovFz2, enNlovFz2, IS200/IS605 transposon-encoded TnpBs, enCnCas12f1, and enRhCas12f1 in human cells. a, Comparison of genome-editing activity at 9 human endogenous gene loci targeted by WT-NlovFz2, enNlovFz2, enISDra2 TnpB, TnpBmax, enCnCas12f1, and enRhCas12f1 in HEK293T cells. Adjusted *P* (P_{adj}) values are 0.0001, 0.7090, 0.5794, 0.1856, 0.0002, respectively. **b**, Comparison of gene editing activity at 9 human endogenous gene loci targeted by WT-NlovFz2, enNlovFz2, ISAba30 TnpB, enCnCas12f1, and enRhCas12f1 in HEK293T cells. Adjusted *P* (P_{adj}) values are 0.0176, 0.3798, 0.6279, 0.0513, respectively. Values and error bars are expressed as mean ± SEM (n = 3 independent biological replicates). *P* values were derived using a two-sided Student's t-test. The top 30% of mCherry$^+$ cells were sorted by FACS to evaluate the genome-editing efficiency of NlovFz2, TnpBs, and Cas12f1 nucleases at endogenous genomic loci.

https://doi.org/10.1038/s41589-025-01902-7



Extended Data Fig. 5 | The sgRNA/ωRNA-dependent off-target effects of enNlovFz2, enCnCas12f1, and enRhCas12f1 at in-silico predicted off-target sites were determined by targeted-amplicon sequencing. The left, middle, and right panels represent the on-target and top10 off-target sites of enNlovFz2, enCnCas12f1, and enRhCas12f1 targeting *DYRK1A*, *HPRT1*, and *DMD*-guide 4 genes, respectively. Values and error bars are expressed as mean ± SEM (n = 3 independent biological replicates).

https://doi.org/10.1038/s41589-025-01902-7



Extended Data Fig. 6 | Establishment and characterization of the DMD$^{\Delta E44\ mdx}$ mouse model. a, Strategy for generating the DMD$^{\Delta E44\ mdx}$ mouse model. The human *DMD* exon44 was deleted using the traditional CRISPR-Cas9 system. The DMD$^{\Delta E44\ mdx}$ mice were generated by mating humanized DMD$^{\Delta E44}$ mice with mdx mice. **b**, Dystrophin immunofluorescence in indicated muscles of wild-type (WT) and DMD$^{\Delta E44\ mdx}$ mice. WT mice were generated by crossing STOCK Tg (DMD) 72Thoen/J mice (#018900) with mdx mice, which carry a c.2977 C > T, p.Gln993* mutation in exon 23 on Chr.X. The staining of dystrophin (Abcam) and spectrin (Millipore) proteins is shown in green and red, respectively. **c**, Western blotting was employed to confirm the absence of dystrophin (Sigma) in heart, diaphragm (DI), and tibialis anterior (TA) muscles of DMD$^{\Delta E44\ mdx}$ mice. **d**, HE staining and Sirius red staining were performed on the TA, DI, and heart muscles of WT and DMD$^{\Delta E44\ mdx}$ mice. **e**, Serum CK, a muscle damage and membrane leakage marker, was measured in WT and DMD$^{\Delta E44\ mdx}$ mice (n = 6 independent biological replicates). The P_{adj} value is 0.0002. **f**, WT and DMD$^{\Delta E44\ mdx}$ mice underwent forelimb grip strength testing to assess muscle performance (n = 6 independent biological replicates). The P_{adj} value is 0.000001. All mice were 4 weeks old at the time of the experiment. Data are presented as mean ± SEM. Each dot represents an individual mouse. *P* values were derived using a two-sided Student's t-test. Scale bar: 100 μm.

nature portfolio

Corresponding author(s): Xiaolong Wang

Last updated by author(s): Mar 10, 2025

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	BD FACS Aria III for flow cytometry data collection. Next-generation sequencing data were generated using the Illumina NovaSeq 6000 platform.
Data analysis	Flow cytometry data were analyzed using FlowJoX (version 10.0.7). Frequency, mean, and standard error of the mean were calculated with GraphPad Prism 10 (version 9.5.1). Deep-seq data were analyzed using Cutadapt (version 2.8) and CRISPResso2 (version 2.0.20b). PEM-seq analysis was conducted following the pipeline described on GitHub (https://github.com/liumz93/PEM-Q).

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

Next-generation sequencing data are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive database under the BioProject accession code PRJNA1147587. Source data are provided with this paper.

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 included two biological replicates, all other experiments used at least three independent biological replicates. Sample sizes were listed in the corresponding figure legends. The sample sizes for three experiments were chosen based on 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 from the analysis.
Replication	All experiments were independently reproduced at least twice, with each replication being successful.
Randomization	Human cells were grown under identical condition, and after seeding them into 24-wells plates, we randomly selected cells for the test and control groups. DMD mice used for intramuscular injection were randomly assigned to either the control group or the AAV9-treated group.
Blinding	Blinding was not performed because the data analyzed (DNA cleavage, genome editing efficiency, etc.) are not subject to biased interpretation.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

Antibodies for western blot: primary antibodies against dystrophin (1:1000 dilution, Sigma, D8168) and vinculin (1:1000 dilution, CST, 13901S); secondary antibody (1:1000 dilution, Beyotime, A0216)
Antibodies for immunofluorescence: primary antibodies against dystrophin (1:100 dilution, Abcam, ab15277) and spectrin (1:500 dilution, Millipore, MAB1622); secondary antibodies (Alexa Fluor 488 AffiniPure donkey anti-rabbit IgG (1:1000 dilution, Jackson ImmunoResearch labs, 711-545-152) or Alexa Fluor 647 AffiniPure donkey anti-mouse IgG (1:1000 dilution, Jackson ImmunoResearch labs, 715-605-151)).

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/primary-antibodies/dystrophin-antibody-ab15277.html; https://www.merckmillipore.com/HK/en/product/msds/MM_NF-MAB1622?ReferrerURL=https%3A%2F%2Fwww.bing.com%2F

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

HEK293T and N2a cells were obtained from the Stem Cell Bank of the Chinese Academy of Sciences.

Authentication

HEK293T and N2a cells were authenticated using STR profiling by the supplier.

Mycoplasma contamination

All cell lines tested negative for mycoplasma contamination via PCR.

Commonly misidentified lines
(See ICLAC register)

No misidentified cell lines listed in the ICLAC database 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

The human DMD exon44 was deleted in STOCK Tg (DMD) 72Thoen/J mice (#018900, 8-week-old) using the traditional CRISPR/Cas9 system. DMDΔE44 mdx mice were then generated by mating humanized DMDΔE44 mice (8-week-old) with mdx mice (8-week-old), which carry a c.2977C>T, p.Gln993* mutation in exon 23 on Chr.X.
The D2B6F1 mice (8-week-old) result from a cross between DBA2 males (8-week-old) and C57BL/6J females (8-week-old).
ICR mice: females, 8-week-old

Wild animals

No wild animals were used in this study.

Reporting on sex

Duchenne muscular dystrophy (DMD) is the most common sex-linked lethal disease in humans, so male mice were selected for this study.

Field-collected samples

This study did not involve any field-collected samples.

Ethics oversight

Experiments involving mice were approved by the Biomedical Research Ethics Committee at HuidaGene Therapeutics Co. Ltd., Shanghai, China. All animal studies adhered to relevant ethical regulations for animal testing and research.

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

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	After transfection, cells were trypsinized, resuspended in cell culture medium, and analyzed by flow cytometry.
Instrument	BD FACSAria III
Software	FlowJo X version 10.0.7
Cell population abundance	A minimum of 20,000 cells per sample were acquired for flow cytometry.
Gating strategy	HEK293T cells were identified using forward and side scatter, followed by doublet exclusion. Cells were then examined for mCherry positivity. Cells transfected with a control plasmid (without targeting spacer) were used as a negative control to establish boundaries between mCherry-positive and mCherry-negative cells. For indel analysis, all mCherry-positive cells were gated.

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