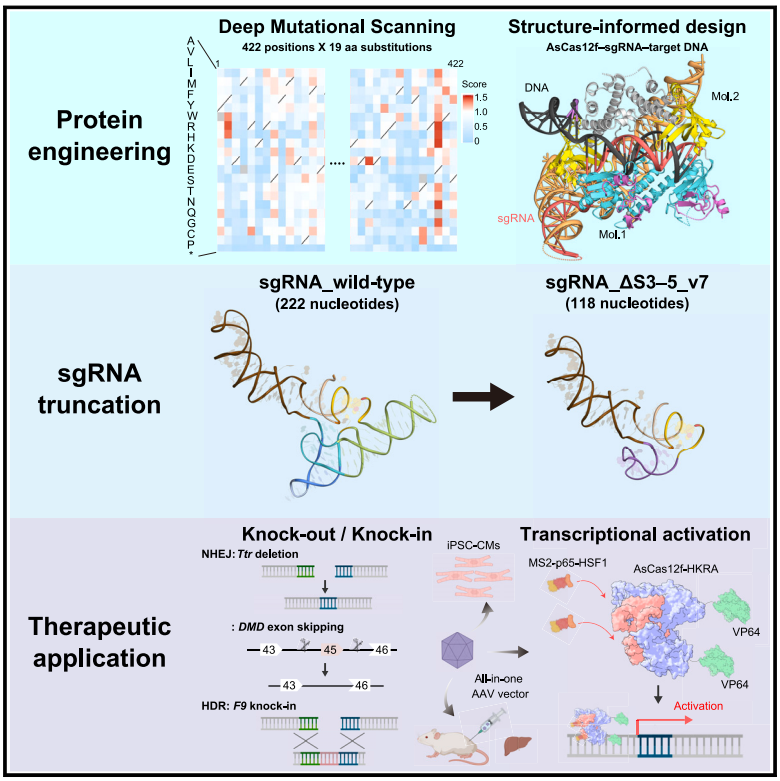
Cell

Article

An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis

Graphical abstract



Highlights

- Cryo-EM structures of compact AsCas12f in complex with sgRNA and target DNA
- AsCas12f engineering by structural analysis and deep mutational scanning methods
- Activities of enAsCas12f variants, SpCas9, and AsCas12a variants are comparable
- enAsCas12f effectors have been harnessed as a compact transcriptional activation tool

Authors

Tomohiro Hino, Satoshi N. Omura, Ryoya Nakagawa, ..., Tsukasa Ohmori, Atsushi Hoshino, Osamu Nureki

Correspondence

tohmori@jichi.ac.jp (T.O.),
a-hoshi@koto.kpu-m.ac.jp (A.H.),
nureki@bs.s.u-tokyo.ac.jp (O.N.)

In brief

A high-throughput approach, combining deep mutational scanning and structure-informed design, generated miniature AsCas12f variants with enhanced activity, enabling efficient knock-in/knock-out activities and transcriptional activation in mice using an all-in-one AAV vector.

Hino et al., 2023, Cell *186*, 4920–4935
October 26, 2023
https://doi.org/10.1016/j.cell.2023.08.031

CellPress

CellPress
OPEN ACCESS

Cell

Article

An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis

Tomohiro Hino,[1,18] Satoshi N. Omura,[2,18] Ryoya Nakagawa,[2,18] Tomoki Togashi,[3,4,18] Satoru N. Takeda,[2] Takafumi Hiramoto,[3] Satoshi Tasaka,[1] Hisato Hirano,[2] Takeshi Tokuyama,[5] Hideki Uosaki,[5] Soh Ishiguro,[6] Madina Kagieva,[6] Hiroyuki Yamano,[7] Yuki Ozaki,[7] Daisuke Motooka,[7,8] Hideto Mori,[9,10,11] Yuhei Kirita,[12] Yoshiaki Kise,[2,13] Yuzuru Itoh,[2] Satoaki Matoba,[1] Hiroyuki Aburatani,[14] Nozomu Yachie,[6,11,15] Tautvydas Karvelis,[16] Virginijus Siksnys,[16] Tsukasa Ohmori,[3,17,*] Atsushi Hoshino,[1,*] and Osamu Nureki[2,19,*]

[1]Department of Cardiovascular Medicine, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto 602-8566, Japan
[2]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan
[3]Department of Biochemistry, Jichi Medical University School of Medicine, Tochigi 329-0498, Japan
[4]Department of Clinical Laboratory Science, Division of Health Sciences, Graduate School of Medical Science, Kanazawa University, Ishikawa 920-0942, Japan
[5]Division of Regenerative Medicine, Center for Molecular Medicine, Jichi Medical University, Tochigi 329-0498, Japan
[6]School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, BC V6S 0L4, Canada
[7]Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan
[8]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives (OTRI), Osaka University, Suita, Osaka 565-0871, Japan
[9]Institute for Advanced Biosciences, Keio University, Yamagata 997-0035, Japan
[10]Graduate School of Media and Governance, Keio University, Fujisawa, Kanagawa 252-0882, Japan
[11]Premium Research Institute for Human Metaverse Medicine (WPI-PRIMe), Osaka University, Suita, Osaka, 565-0871, Japan
[12]Department of Nephrology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto 602-8566, Japan
[13]Curreio, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[14]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan
[15]Synthetic Biology Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan
[16]Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania
[17]Center for Gene Therapy Research, Jichi Medical University, Tochigi 329-0498, Japan
[18]These authors contributed equally
[19]Lead contact
*Correspondence: tohmori@jichi.ac.jp (T.O.), a-hoshi@koto.kpu-m.ac.jp (A.H.), nureki@bs.s.u-tokyo.ac.jp (O.N.)
https://doi.org/10.1016/j.cell.2023.08.031

SUMMARY

SpCas9 and AsCas12a are widely utilized as genome-editing tools in human cells. However, their relatively large size poses a limitation for delivery by cargo-size-limited adeno-associated virus (AAV) vectors. The type V-F Cas12f from *Acidibacillus sulfuroxidans* is exceptionally compact (422 amino acids) and has been harnessed as a compact genome-editing tool. Here, we developed an approach, combining deep mutational scanning and structure-informed design, to successfully generate two AsCas12f activity-enhanced (enAsCas12f) variants. Remarkably, the enAsCas12f variants exhibited genome-editing activities in human cells comparable with those of SpCas9 and AsCas12a. The cryoelectron microscopy (cryo-EM) structures revealed that the mutations stabilize the dimer formation and reinforce interactions with nucleic acids to enhance their DNA cleavage activities. Moreover, enAsCas12f packaged with partner genes in an all-in-one AAV vector exhibited efficient knock-in/knock-out activities and transcriptional activation in mice. Taken together, enAsCas12f variants could offer a minimal genome-editing platform for *in vivo* gene therapy.

INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) systems provide adaptive immunity against mobile genetic elements in bacteria and archaea and are divided into two classes (classes 1 and 2) and six types (types I–VI).[1,2] Cas9 (type II) from *Streptococcus pyogenes* (SpCas9) associates with dual RNA guides (CRISPR

RNA [crRNA] and *trans*-activating crRNA [tracrRNA] or their artificially connected single-guide RNA [sgRNA]) and cleaves double-stranded DNA (dsDNA) targets flanked by an NGG (where N is any nucleotide) protospacer adjacent motif (PAM), using its HNH and RuvC nuclease domains.[3,4] By contrast, among the diverse type V Cas12 enzymes, Cas12a from *Acidaminococcus* sp. (AsCas12a) binds a crRNA and cleaves dsDNA targets with TTTV (where V is A, G, or C) PAMs, using a single RuvC nuclease domain.[5] As SpCas9 and AsCas12a exhibit robust nuclease activities in eukaryotic cells, they are widely used as versatile genome engineering tools.[5,6] However, neither SpCas9 (1,368 amino acids) nor AsCas12a (1,307 amino acids) can be efficiently packaged into a single adeno-associated virus (AAV) vector due to their large gene sizes, which hampers their clinical applications to *in vivo* gene therapy.

Recent studies demonstrated that type V-F Cas12 effectors are exceptionally compact (400–700 amino acids) RNA-guided DNA endonucleases.[7,8] Cas12f associates with dual crRNA:tracrRNA guides and cleaves target DNAs with T-rich PAMs. Previous structural studies of Cas12f from an uncultured archaeon (UnCas12f), with 529 amino acids, revealed that UnCas12f functions as a dimer to compensate for its small size.[9,10] However, UnCas12f cleaves DNA targets only under low salt conditions *in vitro* and lacks activity in human cells, which limits its application as a genome-editing tool. The minimal Cas12f from *Acidibacillus sulfuroxidans* (AsCas12f), which consists of only 422 amino acids, can cleave DNA targets with a TTR (where R is A or G) PAM under physiological conditions *in vitro*.[8] In addition, AsCas12f exhibits low but detectable genome-editing activity in human cells.[11] Therefore, AsCas12f shows great promise as a miniature genome-editing tool that can be packaged into a single AAV vector.

In this study, we present a comprehensive effort to improve the AsCas12f system for genome editing. By combining structural analysis and deep mutational scanning (DMS) methods, we revealed the molecular basis and identified a detailed landscape of amino acid substitutions that greatly augment the nuclease activity of AsCas12f. The synergistic effects of these mutations, coupled with guide RNA engineering, remarkably enhanced the genome-editing efficiency of AsCas12f in human cells to levels comparable with those of both SpCas9 and engineered AsCas12a effectors. The compact size of AsCas12f offers an attractive feature for AAV-deliverable sgRNA and partner genes, such as base editors and epigenome modifiers. Therefore, our newly engineered AsCas12f systems could be a promising genome-editing platform.

RESULTS

Cryo-EM structure of the AsCas12f–sgRNA–target DNA ternary complex

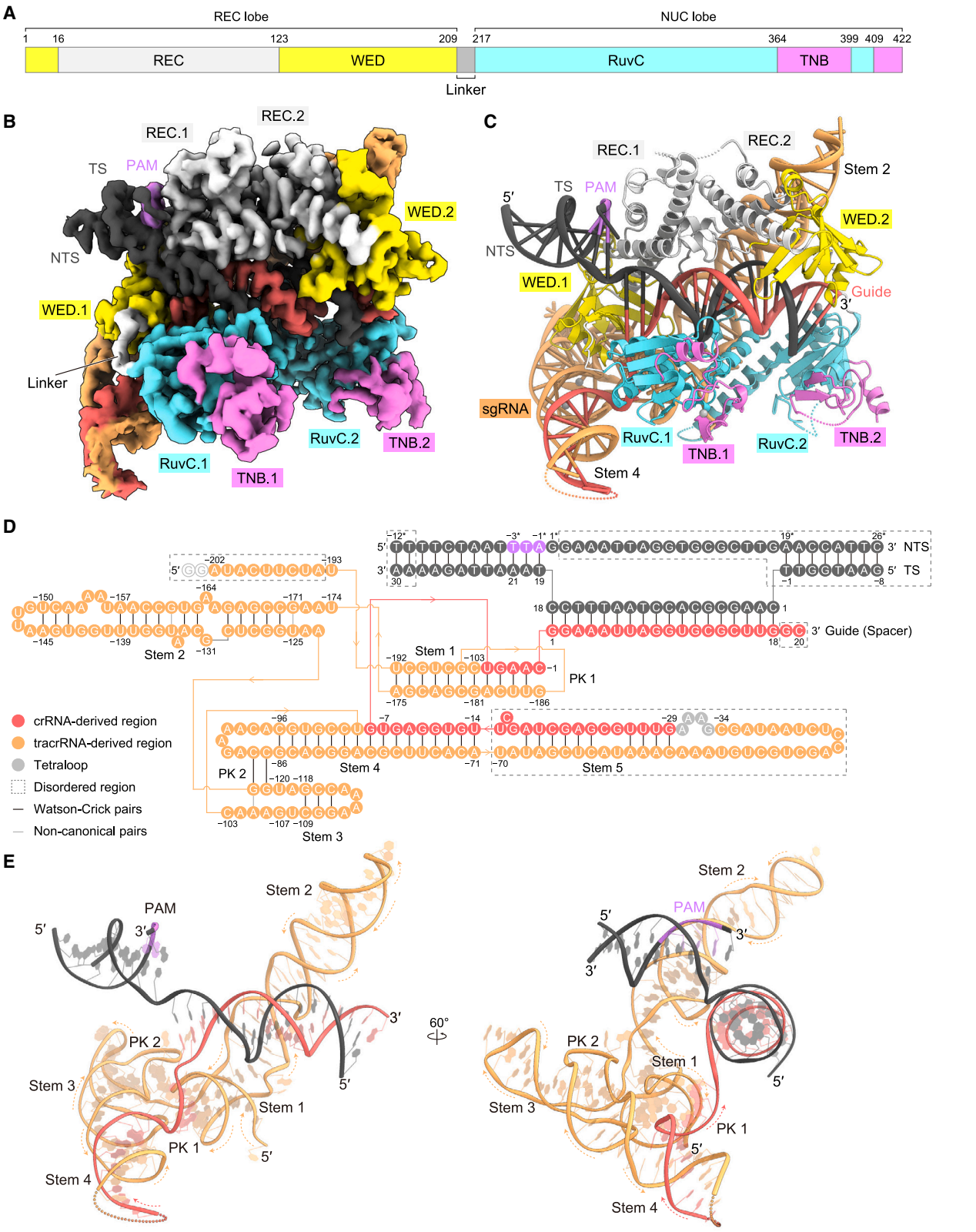
To understand the molecular mechanism of AsCas12f, we used cryoelectron microscopy (cryo-EM) to analyze the complex structure of AsCas12f with a 222-nucleotide (nt) sgRNA (natural 29-nt crRNA and 169-nt tracrRNA sequences connected by a GAAA tetraloop) and a 38-nt dsDNA with phosphorothioate modifications of the DNA backbone around the cleavage site with a TTA PAM and obtained a reconstruction at an overall resolution of 3.1 Å (Figures 1A–1C and S1A–S1D; Table 1). The structure revealed that two AsCas12f molecules (AsCas12f.1 and AsCas12f.2) assemble with one sgRNA molecule to form an asymmetric homodimer, as previously observed in UnCas12f.[9,10] The Cas12f dimer adopts a bilobed architecture comprising a recognition (REC) lobe and a nuclease (NUC) lobe, with the guide RNA–target DNA heteroduplex bound to the central channel between the two lobes. The REC lobe consists of the wedge (WED) and REC domains of both AsCas12f.1 and AsCas12f.2 (WED.1/WED.2/REC.1/REC.2), whereas the NUC lobe includes the RuvC and target nucleic-acid-binding (TNB) domains of both AsCas12f.1 and AsCas12f.2 (RuvC.1/RuvC.2/TNB.1/TNB.2).

The sgRNA consists of the 20-nt guide segment (G1 to C20) and 202-nt sgRNA scaffold (A(−202) to C(−1)) (Figures 1D and 1E). The sgRNA scaffold adopts an unexpected architecture with two pseudoknots (PK 1 and 2) and five stems (stems 1–5), which were not predicted from its primary sequence. Stem 1 contains a 6-base pair (bp) duplex (U(−192):A(−175) to G(−187):C(−180)) and PK 1 comprises a 4-bp duplex (U(−185):A(−2) to A(−182):U(−5)). Stem 2 includes a 16-bp duplex (G(−171):U(−125) to U(−150):A(−145)) with two non-canonical base pairs (G(−171):U(−125) and G(−161):U(−135)), and stem 3 consists of three base pairs (G(−118):C(−109) to C(−116):G(−111)). PK 2 contains two base pairs (G(−122):C(−88) and G(−121):C(−87)) and connects stems 2 and 3. Stem 4 comprises a 16-bp duplex with two coaxially stacked duplexes (U(−102):A(−80) to C(−96):G(−86)). Notably, C(−103), located between stems 3 and 4, is flipped out and inserted into stem 1 and PK 1 to form a base pair with G(−181), thus generating a continuous helix consisting of stem 1 and PK 1 (Figures 1D and 1E). The 5′ region (A(−202) to A(−194)) and stem 5 (U(−70) to U(−15)) are disordered in the structure, suggesting their flexibility (Figures 1D and 1E).

AsCas12f is over 100 residues shorter than UnCas12f, whereas its cognate sgRNA is about 40 nucleotides longer than that of UnCas12f. A structural comparison of AsCas12f with UnCas12f revealed that, although the domain configurations are similar, AsCas12f lacks the zinc-finger (ZF) domain inserted between the WED and REC domains, which is observed in UnCas12f (Figure S2A). Instead, in the AsCas12f structure, the ZF domain is replaced with PK 2 and stem 3, which are not present in UnCas12f (Figure S2A). These findings account for the miniaturization of AsCas12f, which is compensated by its longer sgRNA.

Miniature AsCas12f scaffold associated with the large sgRNA

AsCas12f.1 and AsCas12f.2 interact with each other mainly through hydrophobic interactions in the REC.1/REC.2 and RuvC.1/RuvC.2 interfaces, thereby facilitating the dimerization of AsCas12f, as observed in UnCas12f (Figures S2B–S2E).[9,10] REC.1 and REC.2 form a symmetrical interface by facing each other in opposite directions with W43, F48, and I116 playing central roles, whereas RuvC.1 and RuvC.2 form an asymmetrical interface with each protomer adopting a distinct conformation (Figures S2D and S2E). Notably, the α1 helix and α1-α2 loop of



(legend on next page)

RuvC.1 interact with the helices α1 and α2 of RuvC.2 (Figure S2E). In addition to the protein-protein interactions, the lower and upper regions of stem 2 extensively interact with AsCas12f.1 and AsCas12f.2, respectively, suggesting that the protein-nucleic-acid interactions also contribute to AsCas12f dimerization (Figures S2F and S2G).

The assembly of the sgRNA scaffold with AsCas12f is facilitated by both base-specific and non-specific interactions (Figures 2A–E). The continuous helix consisting of stem 1 and PK 1 is primarily recognized and accommodated within the groove formed by the WED.1 and RuvC.1 domains (Figure 2C). Stem 2 passes through the gap between AsCas12f.1 and AsCas12f.2 and extensively interacts with both protomers, likely reinforcing their dimerization, as described above (Figures S2F and S2G). PK 2 is recognized by the WED.1 and REC.1 domains through sugar-phosphate backbone interactions and further stabilized by the coordination of metal ions (Figure 2D). By contrast, stems 3 and 4 are exposed to the solvent and have minimal interactions with AsCas12f.

The guide RNA–target DNA heteroduplex is accommodated within the positively charged central channel and recognized by AsCas12f through base non-specific interactions (Figures 2B and 2E). The PAM-proximal region of the heteroduplex (G1:dC18 to G12:dC7) is mainly recognized by Cas12f.1, whereas the PAM-distal region of the heteroduplex (C13:dG6 to G18:dC1) is recognized by AsCas12f.2 (Figure 2E). Notably, P240.2 in RuvC.2 stacks with the G18:dC1 base pair in the heteroduplex, indicating that 18 nucleotides in the spacer sequence function as a guide segment (Figure 2E). A structural comparison with UnCas12f revealed that the position of RuvC.1 in AsCas12f is similar to that of RuvC.1 in UnCas12f, which cleaves both the target and non-target strands (Figure 2F). These structural observations suggest that AsCas12f.1 is responsible for cleaving the target DNA, whereas AsCas12f.2 plays a crucial role in recognizing the PAM-distal region of the heteroduplex.

Engineering of AsCas12f to enhance genome-editing efficiency in mammalian cells

Previous studies revealed that AsCas12f shows limited genome-editing activity in human cells.[8,11] To expand the utility of this compact protein, we sought to engineer an AsCas12f variant with enhanced activity. We first performed DMS to determine how all amino acid substitutions impact the genome-editing efficiency in HEK293T cells. We designed a plasmid that expresses both EGFP and AsCas12f and created an AsCas12f library that encompasses all 20-amino acid substitutions in the whole sequence (M1–K422) (Figure 3A). The sgRNA targeting GFP was transduced into HEK293T cells, and subsequently, the AsCas12f library packaged within a lentivirus was expressed at a multiplicity of infection (MOI) of less than 0.2, ensuring that no more than one mutant AsCas12f1 was expressed per cell (Figure 3B).[12] Given that the full-length sgRNA exhibited suboptimal genome-editing efficiency, a stem 5-deleted variant (sgRNA_ΔS5), which exhibited enhanced genome editing in human cells, was used for the screening process. We selected lentivirus-infected cells by sorting GFP-positive cells on day 2 post-infection and cultured them for 5 days further. On day 7 post-infection, we separately extracted mRNAs from GFP-positive and GFP-negative cells and performed the deep sequencing analysis to identify mutations (Figure 3B). The editing efficiency of each mutation was defined as the ratio of GFP-negative read-count to total read-count and normalized by the value of wild-type (WT) AsCas12f (Figure 3C). DMS experiments were performed in duplicate and yielded similar results (R^2 [coefficient of determination] = ~0.6), as in our previous study (Figure S3A).[13] We identified over 200 single amino acid substitutions that resulted in a more than 20% increase in editing efficiency compared with the WT (Figures 3C, 3D, and S3B). Among them, we first focused on the S188H and D195K mutations, which are located close to the nucleic acids in our structure and could form additional interactions to reinforce the target DNA binding by AsCas12f, and adopted them for further experiments. We next developed a split-GFP reporter system, which is connected by a frameshift linker containing the Vascular endothelial growth factor A (VEGFA) sequence, and evaluated the genome-editing activity of WT and several mutants by measuring the ratio of frameshifted GFP-positive cells (Figure S3C).[14] Consistent with the DMS analysis, the S188H and D195K mutants exhibited higher genome-editing activities than WT (Figure 3E). To create variants with higher genome-editing activity than the single S188H or D195K variants, we combined the S188H or D195K substitution with other substitutions selected based on the DMS results and our structural perspective. Although some double mutants showed decreased genome-editing activity, S188H/V232A and D195K/V232A exhibited improved activity (Figures 3E and S3D). To further increase their activities, we introduced additional mutations that enhanced the activity when combined with S188H or D195K, resulting in the generation of seven triple mutants. The S188H/V232A/E316M and D195K/D208R/V232A triple mutants showed further improved genome-editing activities (Figures 3E and S3E). Lastly, we constructed 29 quadruple mutants by combining the S188H/V232A/E316M or D195K/D208R/V232A triple mutants with other effective substitutions identified through the DMS analysis and assessed their genome-editing activities. Most of these variants showed comparable or lower activities than the triple mutants, but the F48Y/S188H/V232A/E316M

Figure 1. Cryo-EM structure of the AsCas12f–sgRNA–target DNA ternary complex

(A) Domain structure of AsCas12f.

(B and C) Cryo-EM maps (B) and structural models (C) of the AsCas12f–sgRNA–target DNA ternary complex. The zinc and magnesium ions in the TNB domain and sgRNA scaffold are shown as gray spheres.

(D) Schematic of sgRNA and the target DNA. The disordered regions are enclosed by dashed boxes. TS, target strand; NTS, non-target strand; PK, pseudoknot. To differentiate between NTS and TS within the manuscript, asterisks are used to denote NTS.

(E) Structure of the sgRNA and target DNA complex. The dotted arrows represent the backbone direction, from 5′ to 3′.

See also Figures S1 and S2.

Table 1. Data collection, processing, model refinement, and validation

Data collection and processing			
Sample	AsCas12f–sgRNA–target DNA	AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA	AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA
EMDB ID	EMD-35912	EMD-35926	EMD-35965
PDB ID	8J12	8J1J	8J3R
Microscope	Titan Krios G3i		
Detector	Gatan K3 camera		
Magnification	105,000		
Voltage (kV)	300		
Electron exposure (e^{-}/Å2)	50		
Defocus range (μm)	−0.8 to −1.6		
Pixel size (Å)	0.83		
Symmetry imposed	C_1		
Number of movies	4,027	4,154	4,398
Initial particle images (no.)	1,304,079	2,118,348	3,835,663
Final particle images (no.)	79,011	146,448	155,198
Map resolution (Å)	3.08	2.91	2.95
FSC threshold	0.143		
3DFSC analysis			
Global resolution (Å)	3.20	3.32	3.38
Sphericity	0.893	0.843	0.805
Model building and refinement			
Map sharpening *B* factor (Å2)	−73.8	−53.6	−58.1
Model composition			
Protein atoms	6,574	6,588	6,545
Nucleic acid atoms	4,121	3,202	2,977
Metal ions	9	4	4
Model refinement			
Model-map CC (CC_{mask}/CC_{box}/CC_{peaks}/CC_{volume})	0.84/0.84/0.82/0.84	0.85/0.84/0.84/0.85	0.84/0.84/0.82/0.84
Resolution (Å) by model to map FSC, threshold 0.50 (masked/unmasked)	3.11/3.05	2.89/2.89	2.94/2.94
Average *B* factor (Å2) (protein/nucleotides/metal ion)	74.20/91.91/75.44	69.20/79.37/96.22	72.01/93.51/73.69
RMS deviations			
Bond lengths (Å)	0.003	0.004	0.003
Bond angles (°)	0.445	0.528	0.498
Validation			
MolProbity score	1.81	1.67	1.60
CaBLAM outliers (%)	1.40	2.29	0.76
Clashscore	6.90	6.59	5.57
Rotamer outliers (%)	2.14	1.43	2.46
C_β outliers (%)	0.00	0.00	0.00
EMRinger score	2.67	3.09	3.14
Ramachandran plot			
Favored (%)	96.99	96.87	97.99
Allowed (%)	3.01	2.88	2.01
Outliers (%)	0.00	0.25	0.00

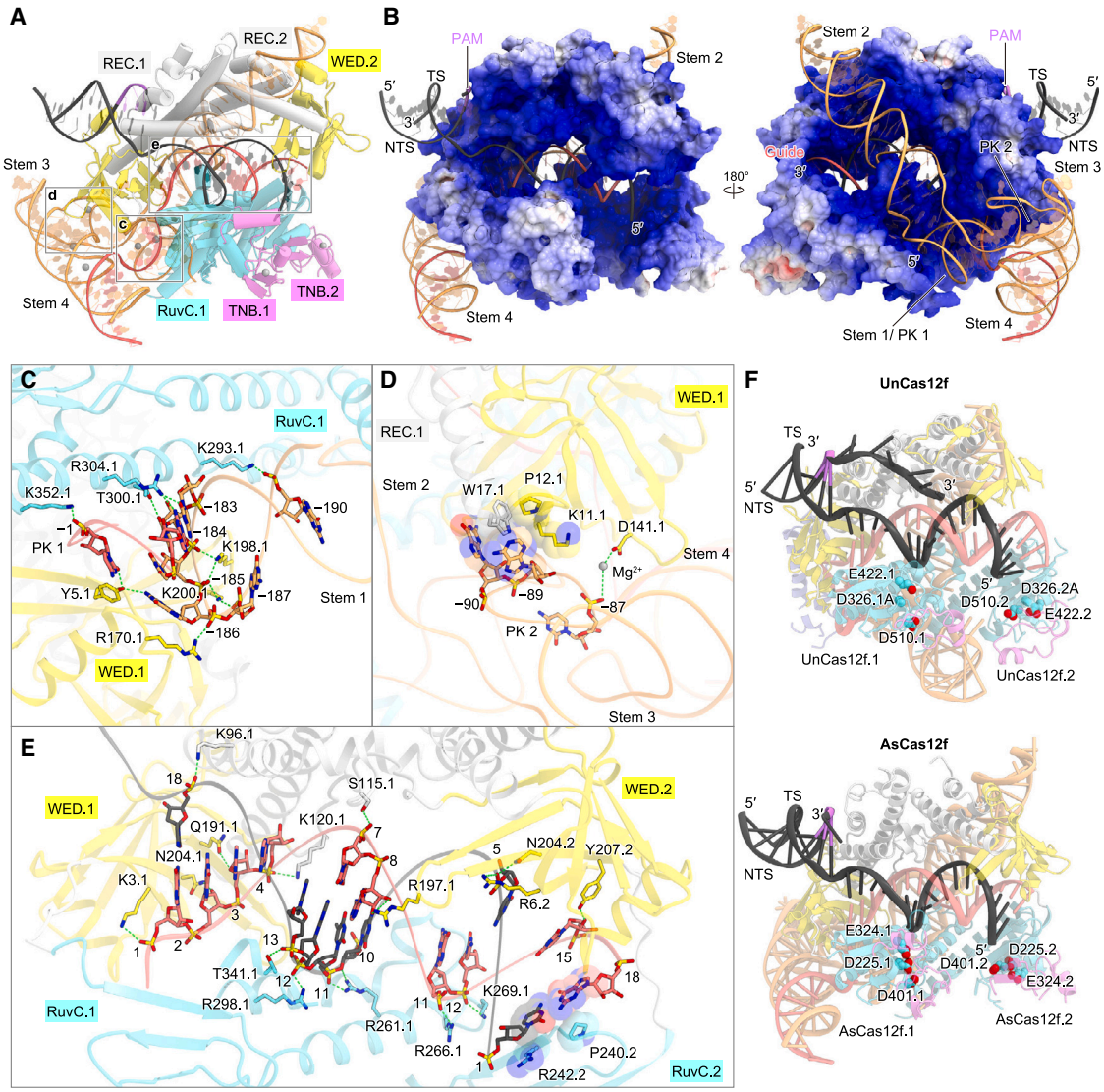


Figure 2. Recognition of the guide RNA and target DNA

(A) Recognition sites of the guide RNA scaffold and target DNA.

(B) Electrostatic surface potential of AsCas12f. The sgRNA–target DNA heteroduplex and the stem 1, stem 2, and PK 1 regions of the sgRNA scaffold are accommodated within the positively charged grooves of the AsCas12f dimer.

(C–E) Recognition of stem 1 and PK 1 (C), stem 3 (D), and the guide RNA–target DNA heteroduplex (E).

(F) Structural comparison of the TS, NTS, and RuvC active site of AsCas12f with those of UnCas12f (Cas12f from an uncultured archaeon) (PDB: 7C7L). The positions of the RuvC.1 active site relative to the target DNA are similar in both structures, suggesting that AsCas12f also cleaves target DNA by using its RuvC.1 domain.

See also Figure S2.

and I123H/D195K/D208R/V232A quadruple mutants exhibited further enhanced genome-editing activities (Figures 3E and S3F). We designated the F48Y/S188H/V232A/E316M and I123H/D195K/D208R/V232A variants as AsCas12f-YHAM and AsCas12f-HKRA, respectively.

Optimization of sgRNA for AsCas12f

A recent study reported that deleting stem 5 from the sgRNA enhances the genome-editing activity of AsCas12f.[15] Consistent with this finding, stem 5 is completely disordered in our structure. Our present structure also revealed that, although PK1, PK2, stem 1, and stem 2 form extensive interactions with the AsCas12f protein, stems 3 and 4 are exposed to the solvent and form few interactions with the protein (Figure 2A). Thus, we sought to further engineer the sgRNA by truncating these domains. We designed seven sgRNA variants (sgRNA_ΔS3–5_v1–v7) to eliminate stems 3, 4, and 5 while connecting the original architecture of the remaining regions by different linker sequences (Figure S4A). As expected, all the sgRNA variants exhibited increased genome-editing activities compared with

sgRNA_ΔS5, with sgRNA_ΔS3–5_v7 being the most effective, albeit by a slight margin (Figure S4B). In our initial DMS experiment, the sgRNA transduced by a multi-copy lentivirus led to higher genome-editing activities than that by a single-copy lentivirus. Therefore, we hypothesized that the expression level of sgRNA in cells may influence the activity of AsCas12f. Indeed, the expression level of sgRNA_ΔS3–5_v7 was 4.5-fold higher than that of sgRNA_ΔS5 in HEK293 cells (Figure S4C), suggesting that shortening the sgRNA sequence led to an increase in its expression, and consequently enhanced the genome-editing activity. To further elucidate the effects of our sgRNA modification, we transduced the GFP-targeting sgRNA_ΔS5 and sgRNA_ΔS3–5_v7 by single-copy lentivirus infection and evaluated the deletion rate of GFP. sgRNA_ΔS3–5_v7 significantly improved the GFP deletion efficiencies of both AsCas12f-WT and AsCas12f-HKRA, compared with sgRNA_ΔS5 (Figures 3F and S4D), suggesting that sgRNA_ΔS3–5_v7 is effective in situations with a limited expression level, such as *in vivo* gene delivery.

Cryo-EM structures of AsCas12f variants with optimized sgRNA

To gain mechanistic insights into the enhanced DNA cleavage activity exhibited by our AsCas12f variants, we determined the cryo-EM structures of the AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA complexes, both at overall resolutions of 2.9 Å (Figures 4A and S1E–S1L; Table 1). The overall structures of the AsCas12f-YHAM and AsCas12f-HKRA variants are similar to that of WT AsCas12f, revealing that the mutations introduced by our DMS experiment do not substantially affect the overall structures of the complexes.

The sgRNA_ΔS3–5_v7 consists of the 20-nt guide segment (G1 to C20) and 96-nt sgRNA scaffold (A(−96) to C(−1)) (Figures S4E–S4H). As we expected, sgRNA_ΔS3–5_v7 lacks the stem 3, 4, and 5 regions while maintaining the stem 1, 2, and PK 1 regions and extensively interacts with AsCas12f. Notably, C(−6), which corresponds to the flipped C(−103) in the WT sgRNA, base pairs with G(−83) and connects PK 1/stem 1 to form a continuous helix (Figures S4I and S4J). The designed linker region (G(−24) to A(−7)) forms the stem-loop structure connected by a typical GAAA tetraloop, containing the G(−24):C(−8) to G(−23):C(−9) base pairs and C(−21):G(−14) to C(−20): G(−15) base pairs (Figures S4F and S4H). Unexpectedly, G(−12) in the stem loop is flipped out and stacks with Trp17, akin to the interaction between G(−90) and Trp17 in the WT, contributing to the stable interaction between sgRNA_ΔS3–5_v7 and AsCas12f (Figures S4K and S4L). Collectively, the sgRNA_ΔS3–5_v7 can stably bind to AsCas12f through both structurally maintained scaffolds and mimicked regions (e.g., C(−6) and G(−12)), explaining why sgRNA_ΔS3–5_v7 exhibits efficient genome-editing activity.

In the AsCas12f-YHAM structure, in addition to the hydrophobic interactions observed in WT AsCas12f, Tyr48.1 (F48Y.1), located at the dimer interface, may form a hydrogen bond with the main chain of G57.2, likely promoting more stable dimerization (Figure 4B). The side chain of His188.1 (S188H.1) is stabilized by Ile2 and interacts with the backbone phosphate group between dT19 and dC18, indicating that the S188H mutation promotes the unwinding of the target DNA (Figure 4C). Met316.1 and.2 (E316M.1 and.2) form additional hydrophobic interactions with Thr239 and Ala241, thereby potentially enhancing the overall stability (Figure 4D).

In the AsCas12f-HKRA structure, His123.1 (I123H.1) and Lys195.1 (D195K.1) provide hydrophobic and electrostatic interactions with the ribose moiety of A5 and the backbone phosphate groups of A3 and A5, respectively, thus stabilizing the guide RNA–target DNA heteroduplex (Figure 4E). Arg208.1 (D208R.1) forms hydrogen-bonding and hydrophobic interactions with dT19 and Ile2, respectively, analogous to His188.1 of AsCas12f-YHAM, thereby facilitating the initial heteroduplex formation (Figure 4F). Furthermore, Arg208.2 of another subunit (D208R.2) also forms an electrostatic interaction with the phosphate backbone of G(−43) in the sgRNA scaffold to stabilize the complex structure (Figure 4G).

Despite the presence of the V232A mutation in both the AsCas12f-YHAM and AsCas12f-HKRA variants, our structures do not provide clear insight into its contribution to the enhanced DNA cleavage activities of both variants. These findings suggest that DMS-based engineering approaches offer significant potential for generating highly active mutants that cannot be predicted from structural information alone.

Characterization of AsCas12f variants in human cells

To assess the PAM specificity of enAsCas12f variants toward a broad range of targets, we developed a self-targeting library composed of both the sgRNA and its target sequence and measured insertion–deletion (indel) formations induced by WT AsCas12f (referred to as AsCas12f for simplicity), AsCas12f-YHAM, and AsCas12f-HKRA toward 750 different spacer sequences with 16 NTTN PAMs (Figures 5A and S5A). The deep sequence analysis demonstrated that the three enzymes induced indels at the NTTR PAM, but not the NTTY (where Y is T or C) PAM, indicating that they commonly recognize the NTTR sequences as the PAM (Figure 5B). Our cryo-EM structures revealed that the nucleobases of dT(−3*) and dT(−2*) in the PAM duplex form hydrophobic interactions with Tyr76.1, whereas the N6 of dA(−1*) forms a hydrogen bond with His72.1 (Figure S5B). When dA(−1*) is replaced by dG, the O6 of dG(−1*) can form a hydrogen bond with His72.1, providing a clear explanation for the NTTR PAM preference. At the NTTR PAM, AsCas12f induced indels at 3.0% on average, whereas AsCas12f-YHAM and AsCas12f-HKRA induced indels at 40.8% and 44.7% on average, respectively (Figure 5B). These results established that the enAsCas12f variants exhibit genome-editing activities much higher than that of AsCas12f at various target sequences with the NTTR PAM.

Next, we compared the genome-editing efficiencies of AsCas12f and enAsCas12f variants with those of AsCas12a and AsCas12a variants (UltraCas12a and enAsCas12a),[16,17] at five target sites with the NTTG PAM in HEK293T cells. AsCas12f, AsCas12f-YHAM, AsCas12f-HKRA, AsCas12a, UltraCas12a, and enAsCas12a generated indels at these five sites with 8.5%, 22.3%, 14.3%, 9.0%, 11.0%, and 11.1% frequencies on average, respectively (Figure 5C). These results indicated that the enAsCas12f variants, but not AsCas12f, can induce indels

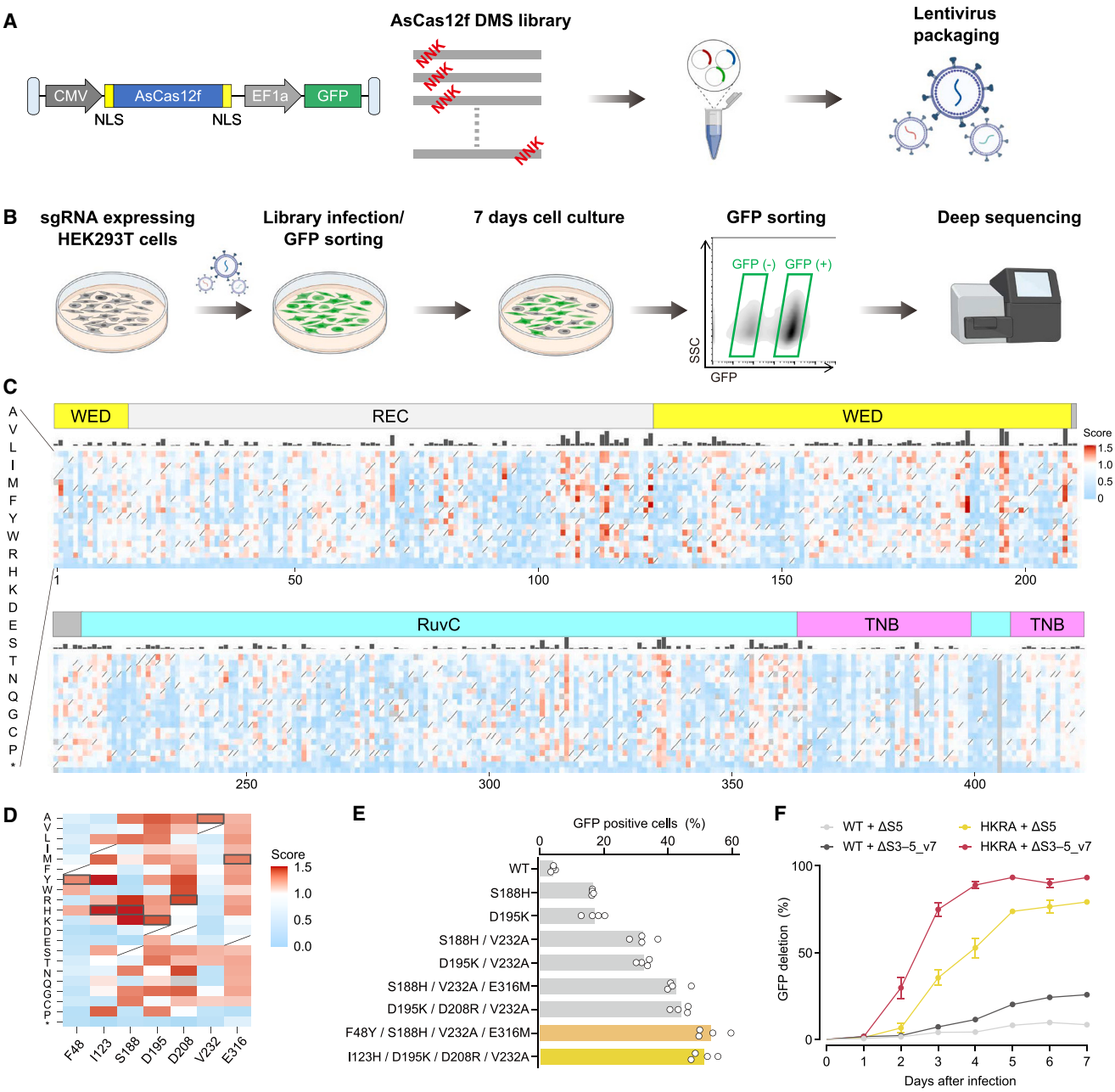


Figure 3. Engineering AsCas12f variants with deep mutational scanning

(A) The design of the deep mutational scanning (DMS) library for AsCas12f.

(B) A schematic of the DMS approach to evaluate the genome-editing efficiency in the context of the GFP gene deletion.

(C) Heatmap illustrating how all single mutations affect the genome-editing activity. Squares are colored by mutational effect according to the scale bars on the right, with blue indicating deleterious mutations. Squares with a diagonal line through them indicate wild-type amino acid.

(D) Close-up view of the selected amino acid substitutions. Squares are colored by mutational effect according to the scale bars on the right, with blue indicating deleterious mutations. Mutations selected for two quadruple mutants (AsCas12f-YHAM and AsCas12f-HKRA) are highlighted with gray-colored borders.

(E) The genome-editing activities of AsCas12f mutants carrying editing efficiency-enhancing mutations, determined by flow cytometry (n = 4).

(F) Time course of GFP deletion induced by AsCas12f and AsCas12f-HKRA in HEK293T cells expressing d2EGFP (n = 3, mean ± SD). sgRNA_ΔS5 and sgRNA_ΔS3–5_v7 were transduced by a single-copy lentivirus infection. See also Figures S3 and S4.

at efficiencies comparable with those of the AsCas12a and AsCas12a variants. We also compared the genome-editing efficiencies with those of SpCas9 at eight target sites, which had the same spacer sequences with a 5′-TTTG PAM for AsCas12f and an NGG-3′ PAM for SpCas9, in HEK293T cells (Figure 5D). AsCas12f, AsCas12f-YHAM, AsCas12f-HKRA, and SpCas9

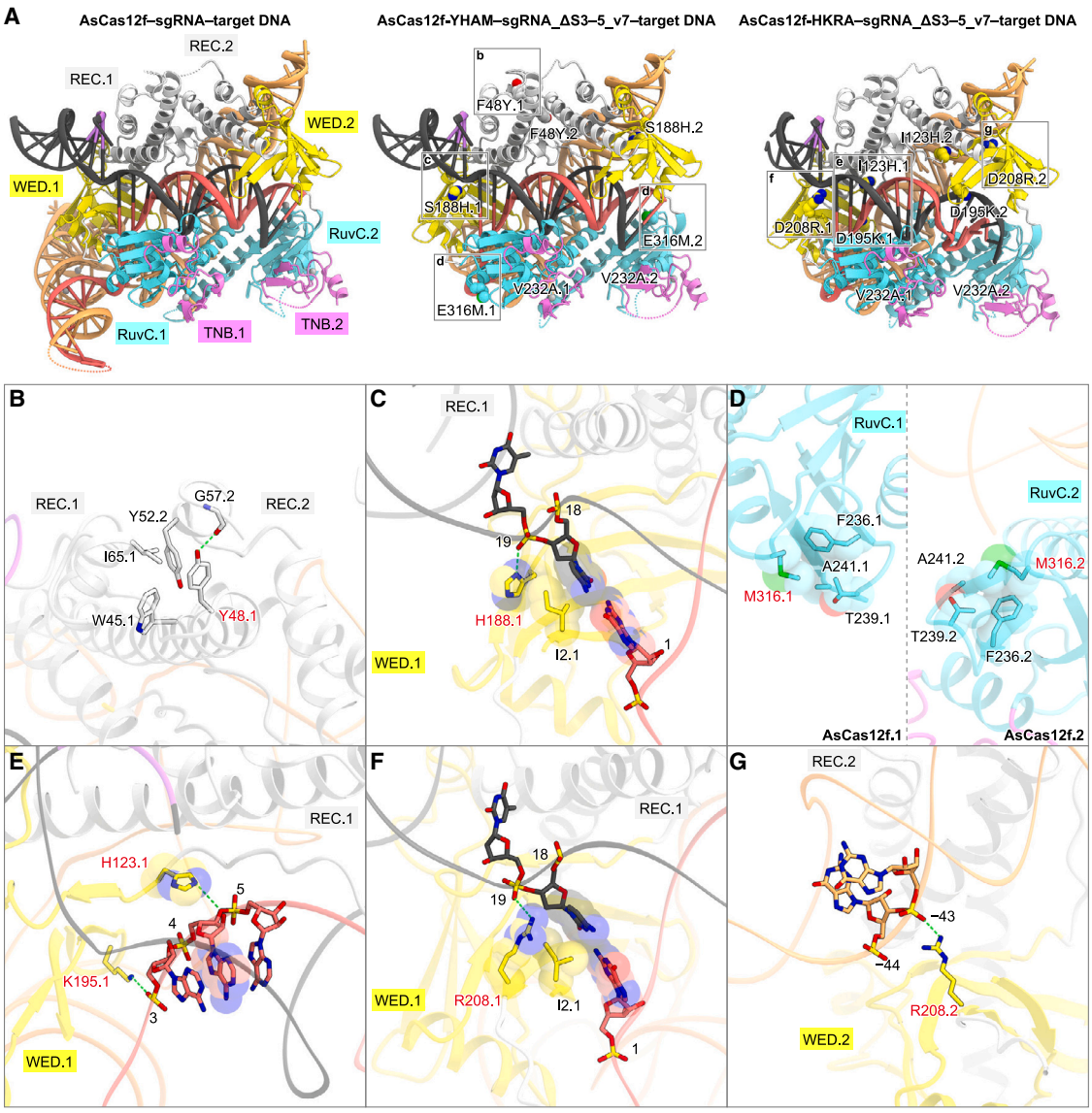


Figure 4. Cryo-EM structure of AsCas12f variants with sgRNA_ΔS3–5_v7

(A) Structural models of the AsCas12f–sgRNA–target DNA (left), AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA (center), and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA (right) complexes. Mutated residues are shown as Corey-Pauling-Koltun (CPK) models.

(B–D) Dimer interface (B), target DNA recognition (C), and hydrophobic interactions (D) of AsCas12f-YHAM. Mutated residues are highlighted in red.

(E–G) Recognition of the guide RNA–target DNA heteroduplex (E), target DNA (F), and sgRNA scaffold (G) of AsCas12f-HKRA. Mutated residues are highlighted in red.

See also Figures S1 and S4.

generated indels at the eight target sites with 1.7%, 14.8%, 16.9%, and 22.9% frequencies on average (Figure 5D). Furthermore, we measured indel formations at three therapeutic targets, *PCSK9* and *ANGPTL3* for atherosclerosis[18] and *TTR* for transthyretin amyloidosis,[19] in HEK293T cells. We observed higher indel frequencies for AsCas12f-YHAM and AsCas12f-HKRA than those for SpCas9 at two of the three target sites (Figure 5E). Similar results were obtained at other target sites and in other human cell lines, Huh-7 hepatoma cells, and HT-1080 sarcoma cells (Figures S5C–S5F), indicating that both AsCas12f-YHAM and AsCas12f-HKRA exhibit genome-editing activities comparable with those of SpCas9 and AsCas12a variants regardless of the target sites or cell lines.

To examine the specificities of enAsCas12f variants, we investigated their mismatch tolerance. Both enAsCas12f variants showed broad tolerance of a single mismatch but negligible tolerance of double mismatches, except for those at the PAM-distal region (positions 16–20), as observed with Cas12a and other Cas12fs (Figure S6A).[15,20,21] We also examined the genome-wide specificities of AsCas12f-HKRA and SpCas9

using GUIDE-seq (genome-wide, unbiased identification of double-stranded breaks (DSBs) enabled by sequencing)[22–24] at nine target sites. AsCas12f-HKRA and SpCas9 had comparable numbers of off-target sites, and AsCas12f-HKRA tolerated mismatches at the PAM-distal 5-nt region, consistent with our mismatch experiments (Figures S6B and S6C). These results demonstrated that enAsCas12f variants with sgRNA_ΔS3–5_v7 exhibit genome-editing activities and specificities comparable with those of SpCas9 and AsCas12a, despite their extremely compact size (Figure S5F).

Therapeutic potential of AsCas12f-mediated genome editing

To evaluate the potentials of the variants for therapeutic applications, we generated induced pluripotent stem cells (iPSCs) with the deletion of Duchenne muscular dystrophy (DMD) exon 44 (hDMDΔEx44), which is the major mutation of DMD, and differentiated them into cardiomyocytes (iPSC-CMs) (Figures S7A–S7D). Dystrophin deficiency causes myopathy and cardiomyopathy, and the skipping of exon 45 to restore the DMD protein is now in clinical trials.[25,26] We first expressed AsCas12f-HKRA and paired sgRNAs designed to skip DMD exon 45, using an all-in-one AAV serotype 6 vector in WT iPSC-CMs, and found that AsCas12f-HKRA efficiently reduced the amount of dystrophin protein (Figure 6A). Furthermore, this all-in-one treatment partially restored the dystrophin protein in *hDMD*ΔEx44-iPSC-CMs (Figures 6B and 6C), confirming the potential of AsCas12f variants to treat DMD using AAV gene delivery.

Next, to investigate the therapeutic potential of AsCas12f *in vivo*, we applied AsCas12f-HKRA to mouse liver genome editing. We constructed an AAV vector encoding AsCas12f or AsCas12f-HKRA under the HCRhAAT promoter and a U6 promoter-driven sgRNA targeting the *TTR* gene for transthyretin amyloidosis (Figure 6D).[27] We injected 7-week-old mice with 3×10^{11} or 1×10^{12} vg of the hepatotropic AAV serotype 8 and evaluated the expression level of plasma transthyretin. AsCas12f failed to modulate the plasma transthyretin level, whereas AsCas12f-HKRA reduced it in a dose-dependent manner (Figure 6E). 8 weeks after AAV-based gene delivery, the target locus was analyzed by targeted amplicon deep sequencing, which showed that a high editing rate (66.3%) was achieved by 1×10^{12} vg of AsCas12f-HKRA (Figure 6F).

We then evaluated the *in vivo* knock-in efficiency of the AsCas12f system to insert the *EGFP* gene at the *mAlb* 3′ UTR. The knock-in of genes at the *Alb* locus is an intensively explored platform to ectopically produce therapeutic proteins, including coagulation factors and lysosomal enzymes, from the liver.[28,29] We prepared two AAV vectors, one expressing AsCas12f or AsCas12f-HKRA and the sgRNA targeting the *mAlb* 3′ UTR, and the other providing a donor template to knock in *EGFP* at the DSB only by homology-directed repair (HDR) (Figure 6G). We intraperitoneally injected these two AAV vectors into C57BL/6 WT neonatal mice and assessed EGFP expression in the liver at 4 weeks after the vector injection by immunofluorescence microscopy. We confirmed a significant increase in EGFP-positive liver cells with the injection of the AAV vector harboring AsCas12f-HKRA (Figures 6H and 6I). We next replaced the *EGFP* gene of the donor vector with the coagulation factor IX (*F9*) cDNA with the Padua mutation and then injected the vectors into neonatal hemophilia B mice (*F9*-deficient mice). The knock-in of the *F9* gene at the *Alb* locus using AsCas12f-HKRA, but not AsCas12f, significantly increased the plasma coagulation factor IX (FIX) activity beyond the therapeutic range (Figure 6J). These data indicate that our engineered AsCas12f variants with the optimal sgRNA can be harnessed for *in vivo* gene therapies (e.g., for hemophilia).

Applications of compact enAsCas12f

Given its small size, the AsCas12f gene can be packaged into an all-in-one AAV vector along with multiple sgRNAs and/or large partner genes, enabling its application to genome-editing treatments that were not possible with previous genome-editing tools. We designed a single AAV vector, encoding AsCas12f (or AsCas12f-HKRA) under a liver-tropic *Ttr* promoter, a donor sequence, and a U6 promoter-driven sgRNA, to insert the *F9* cDNA with the Padua mutation into the *Alb* 3′ UTR locus (Figure 7A).[30] We injected the single AAV serotype 8 vector into neonatal hemophilia B mice and found that both the plasma FIX activity (FIX:C) and antigen (FIX:Ag) were significantly increased into the therapeutic range by the AAV vector harboring AsCas12f-HKRA, but not by that carrying AsCas12f (Figures 7B and 7C). Consistent with these results, the coagulation time assessed by activated partial thromboplastin time (APTT) and the expression level of *F9* mRNA assessed by quantitative real-time PCR (RT-PCR) were significantly improved (Figures 7D and 7E). The amylose gel analysis of the mRNA PCR fragments revealed that cDNA insertion via HDR had occurred, although some insertions mediated by non-homologous end joining (NHEJ) were also observed (Figure 7F).

Finally, to investigate the utility of enAsCas12f for epigenome editing, we performed a transcriptional activation assay in Huh-7 cells, which stably express luciferase driven by the minimal cytomegalovirus (CMV) promoter with two *HEXA* sgRNA recognition sites (Figure 7G). To enhance the transcriptional activity of enAsCas12f, we engineered an sgRNA with the MS2 aptamer inserted into its stem loops.[31] We transfected the Huh-7 cells with the plasmids expressing catalytically inactive AsCas12f-HKRA (dAsCas12f-HKRA) conjugated with VP64 and an MS2-fused activator (MS2-p65-HSF1), as well as engineered sgRNAs targeting *HEXA* (Figure 7G). As shown in Figures 7H and S7E, dAsCas12f-HKRA in combination with the MS2 aptamer-inserted sgRNA significantly enhanced luciferase expression (Figure 7H). The conjugation of VP64 at the end of dAsCas12f-HKRA was most effective for transcriptional activation, although the direct conjugation of VPR (VP64, p65, and Rta) failed to enhance the transcription by MS2-p65-HSF1 (Figures S7F and S7G). In addition, we transduced Huh-7 cells with the all-in-one AAV serotype 6 vector, encoding dAsCas12f or dAsCas12f-HKRA conjugated with VP64, MS2-p65-HSF1, and sgRNA, and confirmed the significant increase of luciferase expression in a dose-dependent manner (Figures 7I and S7H). Lastly, to investigate the utility of enAsCas12f-mediated transcriptional activation *in vivo*, we administered the AAV vector to mice (Figure S7I). We observed the significant increase of luciferase activity in mice treated with dAsCas12f-HKRA, but not with dAsCas12f (Figures 7J and 7K). These results indicated

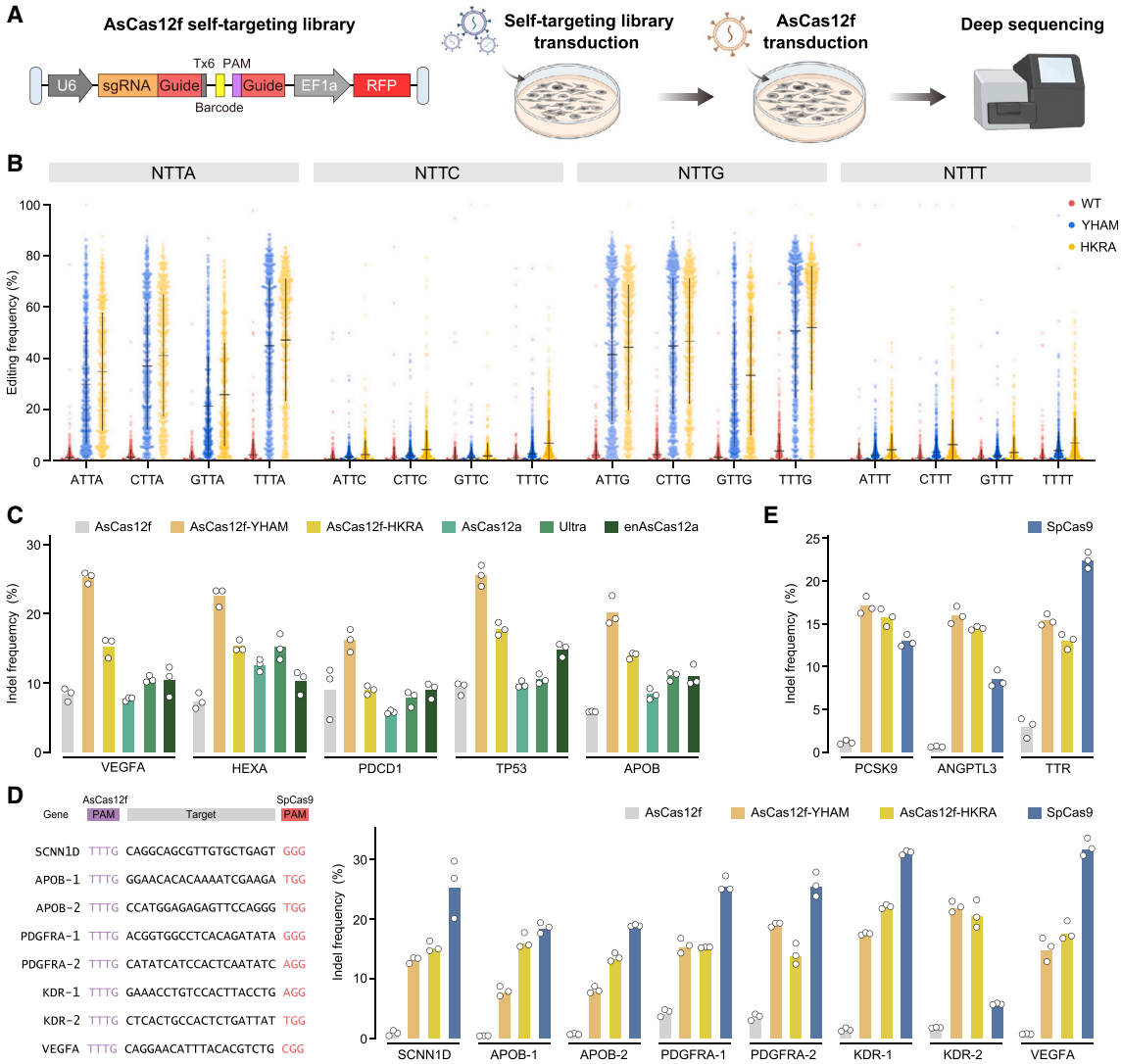


Figure 5. Genome-editing activities of AsCas12f variants in human cell lines

(A) A schematic of the self-targeting library screening for large-scale genome-editing efficiency in HEK293T cells.

(B) Indel efficiencies on 750 target sequences in each NTTN PAM. Red, blue, and yellow represent AsCas12f, -YHAM, and -HKRA, respectively. Bars indicate median indel rates.

(C) Indel efficiencies by AsCas12f, AsCas12f-YHAM, AsCas12f-HKRA, AsCas12a, Ultra, and enAsCas12a at five endogenous target loci in HEK293T cells (n = 3).

(D) Indel efficiencies by AsCas12f, AsCas12f-YHAM, AsCas12f-HKRA, and SpCas9 at eight endogenous target loci in HEK293T cells. The same spacer sequences with a 5′-TTTG PAM for AsCas12f and an NGG-3′ PAM for SpCas9 were used (n = 3).

(E) Indel efficiencies by AsCas12f, AsCas12f-YHAM, AsCas12f-HKRA, and SpCas9 at three therapeutic target loci in HEK293T cells (n = 3).

See also Figures S5 and S6.

that enAsCas12f conjugated with a transcriptional activator can be harnessed for a transcriptional activation tool deliverable using a single AAV vector.

DISCUSSION

In this study, we determined the cryo-EM structures of the AsCas12f–sgRNA–target DNA ternary complex and revealed that two AsCas12f molecules (AsCas12f.1 and AsCas12f.2) assemble with a single sgRNA to form an asymmetric homodimer, similar to UnCas12f. This observation suggests that the compact type V-F CRISPR-Cas nucleases commonly function as dimers.

We applied the DMS technique to CRISPR-Cas effectors and successfully engineered compact AsCas12f variants with enhanced activities. The DMS approach combines exhaustive protein mutagenesis and functional screening with deep sequencing, enabling the assessment of the effects of thousands of mutations in a single experiment. One of the typical applications of DMS is the introduction of the library into the yeast

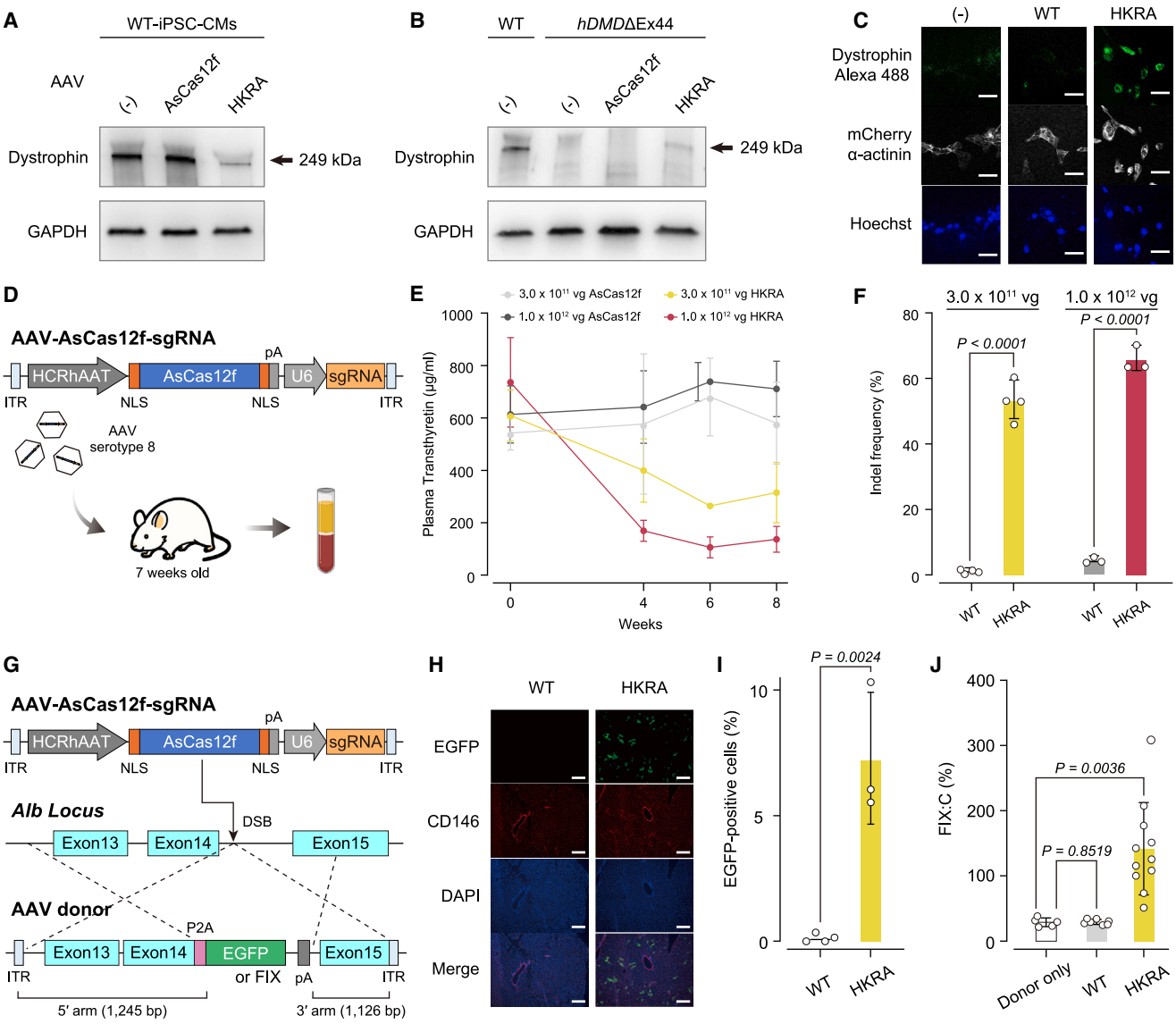


Figure 6. Applications of AsCas12f-HKRA for therapeutic mouse and human models

(A) Western blot analysis of dystrophin protein in wild-type iPSC-CMs (AR21-5) after AAV transduction at a titer of 1 × 10^5 vg/cell.

(B and C) In *hDMD*ΔEx44-iPSC-CMs, AsCas12-HKRA (5 × 10^5 vg/cell) partially restored dystrophin, which was confirmed by western blotting (B) and immunostaining (C). Scale bar represents 40 μm.

(D) A schematic of the AAV vector used for mouse liver genome editing. NLS, SV40; pA, poly A sequence.

(E) Time course of plasma transthyretin levels (n = 3–4, mean ± SD).

(F) Indel frequency at the *Ttr* locus in mouse livers (n = 3–4, mean ± SD). Statistical significance between AsCas12f and AsCas12f-HKRA was analyzed by two-tailed Student's t test.

(G) A schematic of the AAV vectors for EGFP or coagulation factor IX (*F9*) knockin at the *Alb* 3′ UTR. NLS, SV40; pA, poly A sequence.

(H) Immunohistochemistry (IHC) staining of livers from AAV-injected mice at 4 weeks after the vector injection. Endothelial cells were stained with antibodies specific to CD146 (red). Nuclei were counterstained with DAPI (blue). Scale bars, 200 μm.

(I) The ratio of EGFP-positive cells. Statistical significance between AsCas12f and AsCas12f-HKRA was analyzed by two-tailed Student's t test.

(J) The increase in plasma factor IX activity (FIX:C). Statistical significance between donor only and donor plus AsCas12f-HKRA was analyzed by two-tailed Student's t test.

See also Figure S7.

surface display system and the evaluation of the binding affinity of ligands, including antibodies and viral glycoproteins.[32] In this study, to evaluate genome-editing activities, we applied the yeast screening system to mammalian cell-based screening. We constructed a library that encompasses all 20 single amino acid substitutions at each position in the whole sequence of

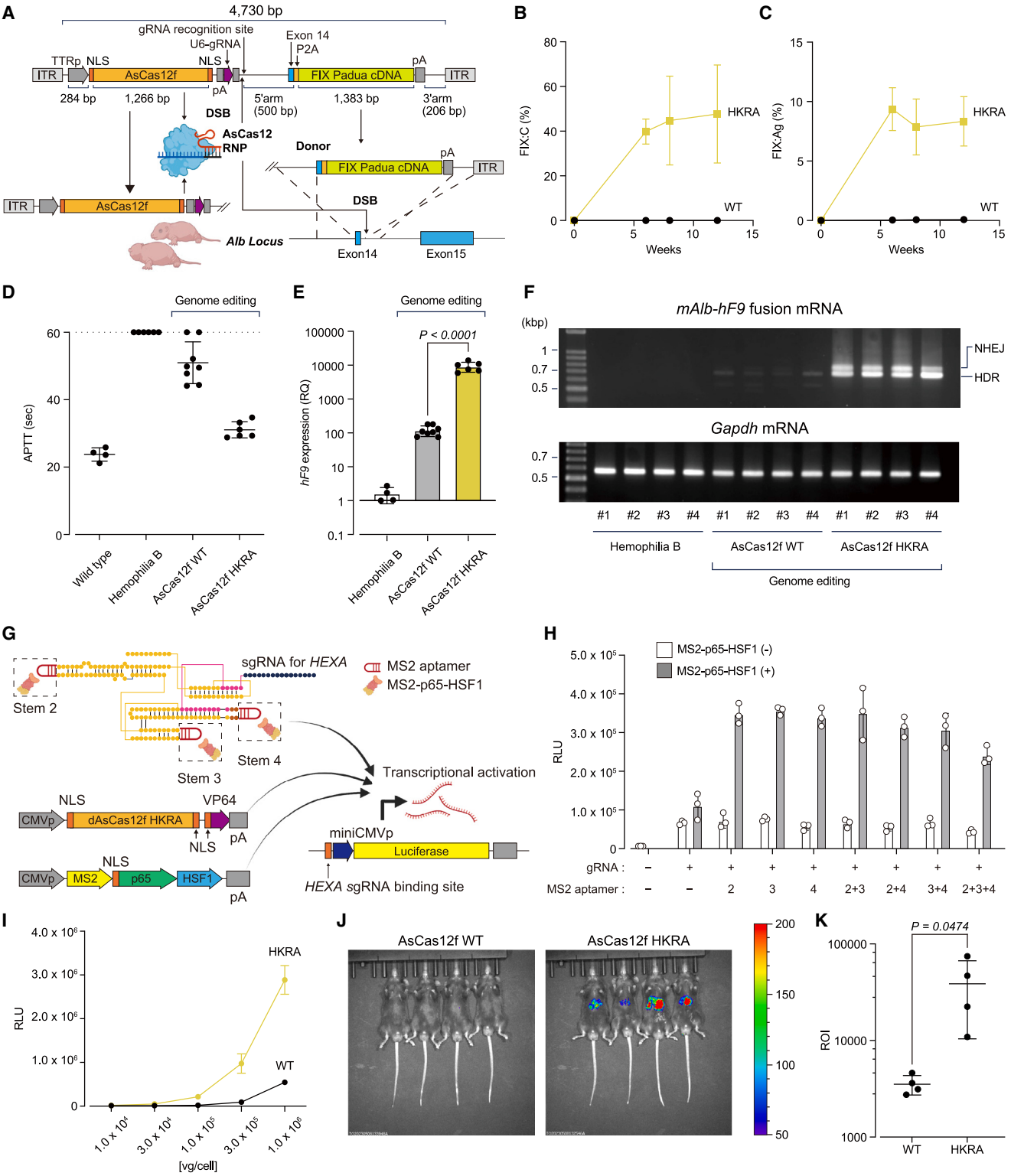


Figure 7. Application of enAsCas12f for knock-in and transcriptional activation by a single AAV vector

(A) A schematic representation of the knock-in strategy targeting the *Alb* locus by a single AAV vector *in vivo*. Neonatal hemophilia B mice were treated with the AAV vector (3×10^{11} vg/body). NLS, SV40; pA, poly A sequence.

(B and C) Increase in plasma factor IX activity (FIX:C) (B) and antigen (FIX:Ag) (C) (n = 6–8, mean ± SD).

(legend continued on next page)

AsCas12f (422 amino acids) and identified over 200 effective mutations at various positions. To efficiently explore the effective combinations of these mutants, we employed a structural perspective. In the absence of an experimentally determined structure, structure prediction models like AlphaFold would be a useful approach for efficiently selecting mutations identified through DMS.

Various CRISPR-Cas variants have been engineered with the assistance of structural information[33,34] and directed evolution with random mutagenesis.[35] However, the former approach has limited capacities to test the number of mutations, whereas the latter method carries the risk of introducing unnecessary mutations that do not contribute to activity improvement. By contrast, by combining comprehensive DMS with rationally structure-informed design, our approach enables the development of variants with enhanced activity in a more reliable and efficient manner. By applying this powerful approach to other Cas enzymes with different PAM sequences, we can potentially generate efficient genome-editing enzymes capable of targeting a wide range of genes. Moreover, with suitable adaptations to the evaluation system, this approach can be applied to enzymes beyond the scope of genome editing.

Limitations of the study

Although our DMS technique discovered numerous effective substitutions, it is not clear whether the best set of mutations was selected in the final optimized variants, due to the lack of a reliable method for selecting the optimal combination of effective mutations. One future solution is to use computational modeling to predict effective combinations without compromising protein stability.[36] Another is a machine learning-based approach. By training a machine learning algorithm using a dataset from the DMS containing multiple mutations,[37] the resultant model could predict genome-editing activity values for all possible mutation combinations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.08.031.

ACKNOWLEDGMENTS

Experimental schemes were created in part through BioRender.com. V.S. acknowledges the Research Council of Lithuania for the support of the European Joint Programme on Rare Diseases project GET-READY (grant number S-EJPRD-21-1). T.O. was supported by AMED grant numbers JP22fk0410037 and JP22ae0201007. A.H. was supported by AMED grant number JP19am0401005 and the Japan Foundation for Applied Enzymology. O.N. was

(D) Plasma coagulation time assessed by activated partial thromboplastin time (APTT) in C57BL/6 mice, hemophilia B mice, and hemophilia B mice treated with the AAV vector (n = 4–8, mean ± SD).
(E) The expression of *hF9* in liver, measured by quantitative RT-PCR (n = 4–8, mean ± SD). Statistical significance was analyzed by one-way ANOVA with post hoc Tukey's multiple comparison test.
(F) Amylose gel analysis of mRNA PCR fragments. When a donor AAV fragment is inserted into the DSB by NHEJ, the duplication of exon 14 results in the 68-bp longer transcript, compared with those modified by HDR.
(G) A schematic representation of transcriptional activation with the insertion of the MS2 aptamer in the stem loop of the sgRNA of AsCas12f. The plasmid and sgRNA targeting *HEXA* were transduced into Huh-7 cells stably expressing luciferase under the control of the minimal CMV promoter with *HEXA* sgRNA-binding sites. NLS, SV40; pA, poly A sequence.
(H) Increased luciferase activity in the cells transduced with the plasmid vector (n = 3, mean ± SD).
(I) Increased luciferase activity in the cells transduced with the AAV vector (n = 4, mean ± SD).
(J) Representative IVIS imaging system data at 3 weeks after the vector administration.
(K) Quantitative data expressed as photon units. Values are presented as mean ± SD (n = 4). Statistical significance was analyzed by two-tailed Student's t test. See also Figure S7.

supported by AMED grant numbers JP233fa627001 and JP19am0401005, the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research [BINDS]) from AMED under grant number JP23am121002 (support number 3272) and JP23ama121012 (support number 4848, A.H.), and the Cabinet Office, the Government of Japan, and the Public/Private R&D Investment Strategic Expansion Program (PRISM) grant number JPJ008000.

AUTHOR CONTRIBUTIONS

T.H., S.T., Y.K., and A.H. conducted deep mutational scanning and subsequent data analysis with assistance from S.M. T.H. and S.T. performed *in vitro* assays to identify enhanced mutants. A.H. performed self-targeting library experiments. H.Y., Y.O., D.M., and H.M. contributed to the data analysis. T.H. performed the GUIDE-seq analysis. S.I., M.K., H.A., and N.Y. contributed to data collection. T.T. and H.U. conducted the iPSC-derived cardiomyocyte experiments. T.T., T.H., and T.O. conducted AAV preparation and mouse experiments. S.N.O. and R.N. performed biochemical and structural analyses with assistance from S.N.T., H.H., Y.K., and Y.I. T.K. and V.S. conceived the project. S.N.O., R.N., H.U., T.O., A.H., and O.N. wrote the manuscript with help from all authors. T.O., A.H., and O.N. supervised the research.

DECLARATION OF INTERESTS

T.H., S.N.O., R.N., Y.K., S.M., T.O., A.H., and O.N. have filed a patent application related to this work.

Received: April 21, 2023
Revised: June 27, 2023
Accepted: August 23, 2023
Published: September 29, 2023

REFERENCES

1. Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The biology of CRISPR-Cas: backward and forward. Cell *172*, 1239–1259.
2. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83.
3. Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA *109*, E2579–E2586.
4. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.
5. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.
6. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.
7. Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., and Doudna, J.A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science *362*, 839–842.
8. Bigelyte, G., Young, J.K., Karvelis, T., Budre, K., Zedaveinyte, R., Djukanovic, V., Van Ginkel, E., Paulraj, S., Gasior, S., Jones, S., et al. (2021). Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. Nat. Commun. *12*, 6191.
9. Takeda, S.N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Nishimasu, H., and Nureki, O. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol. Cell *81*, 558–570.e3.
10. Xiao, R., Li, Z., Wang, S., Han, R., and Chang, L. (2021). Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. Nucleic Acids Res. *49*, 4120–4128.
11. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., Li, F., Liu, C., Nan, H., Chen, W., et al. (2021). Programmed genome editing by a miniature CRISPR-Cas12f nuclease. Nat. Chem. Biol. *17*, 1132–1138.
12. Higuchi, Y., Suzuki, T., Arimori, T., Ikemura, N., Mihara, E., Kirita, Y., Ohgitani, E., Mazda, O., Motooka, D., Nakamura, S., et al. (2021). Engineered ACE2 receptor therapy overcomes mutational escape of SARS-CoV-2. Nat. Commun. *12*, 3802.
13. Ikemura, N., Taminishi, S., Inaba, T., Arimori, T., Motooka, D., Katoh, K., Kirita, Y., Higuchi, Y., Li, S., Suzuki, T., et al. (2022). An engineered ACE2 decoy neutralizes the SARS-CoV-2 Omicron variant and confers protection against infection in vivo. Sci. Transl. Med. *14*, eabn7737.
14. Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. *23*, 102–107.
15. Kim, D.Y., Lee, J.M., Moon, S.B., Chin, H.J., Park, S., Lim, Y., Kim, D., Koo, T., Ko, J.H., and Kim, Y.S. (2022). Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. Nat. Biotechnol. *40*, 94–102.
16. Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., Scarfò, I., et al. (2019). Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol. *37*, 276–282.
17. Zhang, L., Zuris, J.A., Viswanathan, R., Edelstein, J.N., Turk, R., Thommandru, B., Rube, H.T., Glenn, S.E., Collingwood, M.A., Bode, N.M., et al. (2021). AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. Nat. Commun. *12*, 3908.
18. Musunuru, K. (2022). Moving toward genome-editing therapies for cardiovascular diseases. J. Clin. Invest. *132*, e148555.
19. Finn, J.D., Smith, A.R., Patel, M.C., Shaw, L., Youniss, M.R., van Heteren, J., Dirstine, T., Ciullo, C., Lescarbeau, R., Seitzer, J., et al. (2018). A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. Cell Rep. *22*, 2227–2235.
20. Kim, D., Lim, K., Kim, D., and Kim, J.S. (2020). Genome-wide specificity of dCpf1 cytidine base editors. Nat. Commun. *11*, 4702.
21. Wang, Y., Wang, Y., Pan, D., Yu, H., Zhang, Y., Chen, W., Li, F., Wu, Z., and Ji, Q. (2022). Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease. Cell Rep. *40*, 111418.
22. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197.
23. Malinin, N.L., Lee, G.H., Lazzarotto, C.R., Li, Y., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Iafrate, A.J., Le, L.P., et al. (2021). Defining genome-wide CRISPR–Cas genome-editing nuclease activity with GUIDE-seq. Nat. Protoc. *16*, 5592–5615.
24. Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., and Joung, J.K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat. Biotechnol. *34*, 869–874.
25. Kenjo, E., Hozumi, H., Makita, Y., Iwabuchi, K.A., Fujimoto, N., Matsumoto, S., Kimura, M., Amano, Y., Ifuku, M., Naoe, Y., et al. (2021). Low immunogenicity of LNP allows repeated administrations of CRISPR-Cas9 mRNA into skeletal muscle in mice. Nat. Commun. *12*, 7101.

26. Wagner, K.R., Kuntz, N.L., Koenig, E., East, L., Upadhyay, S., Han, B., and Shieh, P.B. (2021). Safety, tolerability, and pharmacokinetics of casimersen in patients with Duchenne muscular dystrophy amenable to exon 45 skipping: a randomized, double-blind, placebo-controlled, dose-titration trial. Muscle Nerve *64*, 285–292.

27. Gillmore, J.D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M.L., Seitzer, J., O'Connell, D., Walsh, K.R., Wood, K., et al. (2021). CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis. N. Engl. J. Med. *385*, 493–502.

28. Pagant, S., Huston, M.W., Moreira, L., Gan, L., St Martin, S., Sproul, S., Holmes, M.C., Meyer, K., Wechsler, T., Desnick, R.J., et al. (2021). ZFN-mediated in vivo gene editing in hepatocytes leads to supraphysiologic α-Gal A activity and effective substrate reduction in Fabry mice. Mol. Ther. *29*, 3230–3242.

29. He, X., Zhang, Z., Xue, J., Wang, Y., Zhang, S., Wei, J., Zhang, C., Wang, J., Urip, B.A., Ngan, C.C., et al. (2022). Low-dose AAV-CRISPR-mediated liver-specific knock-in restored hemostasis in neonatal hemophilia B mice with subtle antibody response. Nat. Commun. *13*, 7275.

30. VandenDriessche, T., and Chuah, M.K. (2018). Hyperactive factor IX Padua: a game-changer for hemophilia gene therapy. Mol. Ther. *26*, 14–16.

31. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

32. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.

33. Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., et al. (2016). Structure and engineering of Francisella novicida Cas9. Cell *164*, 950–961.

34. Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science *361*, 1259–1262.

35. Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464–471.

36. Arimori, T., Ikemura, N., Okamoto, T., Takagi, J., Standley, D.M., and Hoshino, A. (2022). Engineering ACE2 decoy receptors to combat viral escapability. Trends Pharmacol. Sci. *43*, 838–851.

37. Taft, J.M., Weber, C.R., Gao, B., Ehling, R.A., Han, J., Frei, L., Metcalfe, S.W., Overath, M.D., Yermanos, A., Kelton, W., et al. (2022). Deep mutational learning predicts ACE2 binding and antibody escape to combinatorial mutations in the SARS-CoV-2 receptor-binding domain. Cell *185*, 4008–4022.e14.

38. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

39. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

40. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

41. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. D Struct. Biol. *74*, 531–544.

42. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. *66*, 12–21.

43. Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat. Methods *17*, 1214–1221.

44. Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

45. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. *37*, 224–226.

46. Marquart, K.F., Allam, A., Janjuha, S., Sintsova, A., Villiger, L., Frey, N., Krauthammer, M., and Schwank, G. (2021). Predicting base editing outcomes with an attention-based deep learning algorithm trained on high-throughput target library screens. Nat. Commun. *12*, 5114.

47. Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell *160*, 1246–1260.

48. Mimuro, J., Mizukami, H., Hishikawa, S., Ikemoto, T., Ishiwata, A., Sakata, A., Ohmori, T., Madoiwa, S., Ono, F., Ozawa, K., et al. (2013). Minimizing the inhibitory effect of neutralizing antibody for efficient gene expression in the liver with adeno-associated virus 8 vectors. Mol. Ther. *21*, 318–323.

49. Ohmori, T., Nagao, Y., Mizukami, H., Sakata, A., Muramatsu, S.I., Ozawa, K., Tominaga, S.I., Hanazono, Y., Nishimura, S., Nureki, O., et al. (2017). CRISPR/Cas9-mediated genome editing via postnatal administration of AAV vector cures haemophilia B mice. Sci. Rep. *7*, 4159.

50. Ahmed, R.E., Chanthra, N., Anzai, T., Koiwai, K., Murakami, T., Suzuki, H., Hanazono, Y., and Uosaki, H. (2021). Sarcomere shortening of pluripotent stem cell-derived cardiomyocytes using fluorescent-tagged sarcomere proteins. J. Vis. Exp., 1–18.

51. Atmanli, A., Chai, A.C., Cui, M., Wang, Z., Nishiyama, T., Bassel-Duby, R., and Olson, E.N. (2021). Cardiac myoediting attenuates cardiac abnormalities in human and mouse models of Duchenne muscular dystrophy. Circ. Res. *129*, 602–616.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mouse anti-dystrophin	Sigma-Aldrich	# D8168, RRID:AB_259245
mouse anti-dystrophin antibody	Santa Cruz Biotechnology	# sc-73592, RRID:AB_1122390
mouse anti-GAPDH	Santa Cruz Biotechnology	# sc-32233, RRID:AB_627679
anti-mouse CD146	BioLegend	# 134702, RRID:AB_1731991
Anti-GFP rabbit polyclonal antibody	MBL	# 598, RRID:AB_591816
Goat anti-Rabbit IgG (H+L) Cross-Absorbed Secondary AlexaFluor 488	Thermo Fisher Scientific	# A32731
Goat anti-Rat IgG (H+L) Cross-Absorbed Secondary AlexaFluor 594	Thermo Fisher Scientific	# A11007
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 488	Thermo Fisher Scientific	# A32766
Peroxidase AffiniPure donkey anti-mouse IgG (H+L)	Jackson ImmunoResearch	# 715-035-150
Bacterial and virus strains		
Competent Quick DH5α	TOYOBO	DNA-913F
Endura ElectroCompetent Cells	Lucigen	60242-2
AAV6 AsCas12f DMD	This paper	N/A
AAV6 AsCas12f-HKRA DMD	This paper	N/A
AAV8 AsCas12f Ttr	This paper	N/A
AAV8 AsCas12f-HKRA Ttr	This paper	N/A
AAV8 AsCas12f Alb	This paper	N/A
AAV8 AsCas12f-HKRA Alb	This paper	N/A
AAV8 Alb-EGFP	This paper	N/A
AAV8 Alb-F9 Padua	This paper	N/A
AAV8 mTTR AsCas12f1 sgRNA_Alb-F9 Padua donor	This paper	N/A
AAV8 mTTR AsCas12f1-HKRA sgRNA_Alb-F9 Padua donor	This paper	N/A
AAV6 CMV AsCas12f (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
AAV6 CMV AsCas12f-HKRA (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
AAV8 mTTR AsCas12f (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
AAV8 mTTR AsCas12f-HKRA (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
AAV8 CMVmini (AsCas12f1-HEXA)-Luciferase	This paper	N/A
Chemicals, peptides, and recombinant proteins		
AsCas12f	Bigelyte et al.[8]	N/A
AsCas12f variants	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DMEM	Sigma-Aldrich	# D6429
DMEM	WAKO	044-29765
FBS	Thermo Fisher Scientific	#10270-106
RPMI1640	Sigma-Aldrich Thermo Fisher Scientific	# R8758 # 11875-093
StemFit AK02N	ReproCell	RCAK02N
B-27 supplement (50X), serum free	Thermo Fisher Scientific	17504044
B-27 Supplement, minus insulin	Thermo Fisher Scientific	A1895601
iMatrix-511 silk	Matrixome	892021
Y-27632	Fujifilm/Wako Pure Chemical	257-00614
CHIR99021	Cayman	13122
C59, Wnt Antagonist (WntC59)	Abcam	ab142216
human EGF	Thermo Fisher Scientific	#13247-051
ITS-X Supplement	Thermo Fisher Scientific	#51500-056
GlutaMAX	Thermo Fisher Scientific	#35050-061
penicillin/streptomycin	Thermo Fisher Scientific	# 15140122
Fugene HD	Promega	E2311
Lipofectamine 3000	Thermo Fisher Scientific	#3000015
Cellmatrix type I-C	Nitta Gelatin	# 637-00773
TRIzol	Thermo Fisher Scientific	# 15596018
Direct-zol RNA MiniPrep	Zymo Research Corporation	# R2052
PrimeScript II Reverse Transcriptase	Takara Bio	# 2690A
SimplePrep reagent for DNA	Takara Bio	# 9180
ExTaq DNA polymerase	Takara Bio	# RR006A
PrimeSTAR HS (Premix)	Takara Bio	# R040A
KOD One PCR Master Mix	TOYOBO	KMM-101
T7 endonuclease	Nippon Gene	# 313-08801
NEBuilder HiFi DNA Assembly Master Mix	NEB	E2621X
DNA Ligation Kit	Takara Bio	# 6023
PrimeScript RT Master Mix	Takara Bio	# RR036A
KAPA SYBR FAST	Kapa Biosystems	# 7959630001
SE Cell Line 4D-Nucleofecto X Kit L	Lonza	# V4XC-1024
Tissue-Tek optimal cutting temperature compound	Sakura Fintek Japan	# 45833
VECTASHIELD Mounting Medium with DAPI	Vector Laboratories	#H-1500
Alt-R® S.p. Cas9 Nuclease V3 Alt-R® S.p. Cas9 Nuclease V3	Integrated DNA Technologies	1081058
Alt-R® CRISPR-Cas9 tracrRNA	Integrated DNA Technologies	1072533
4–15% Mini-PROTEAN® TGX™ Precast Protein Gels	Bio-Rad	#4561083
10x Tris/Glycine/SDS	Bio-Rad	#1610732
Trans-Blot Turbo Midi 0.2 μm PVDF Transfer Packs	Bio-Rad	#1704157
PVDF Blocking Reagent for *Can Get Signal*	TOYOBO	NYPBR01
Can Get Signal® Immunoreaction Enhancer Solution	TOYOBO	NKB-101
Tris Buffered Saline with Tween®20 (TBS-T) Tablets, pH 7.6	Takara Bio	T9142

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Immobilon Western Chemiluminescent HRP Substrate	Merck	WBKLS0500
cOmplete™ EDTA-free protease inhibitor cocktail	Roche	11836170001
phosphatase inhibitor cocktail 2	Sigma-Aldrich	P5726
phosphatase inhibitor cocktail 3	Sigma-Aldrich	P0044
Droplet Generation Oil for Probes	Bio-Rad	#1863005
ddPCR Droplet Reader Oil	Bio-Rad	#1863004
ddPCR™ Buffer Control for Probes	Bio-Rad	#1863052
ddPCR™ Supermix for Probes (No dUTP)	Bio-Rad	#1863023
Thrombocheck APTT	Sysmex	#15350
Thrombocheck Factor IX	Sysmex	#15450
Luciferase Assay System	Promega	E1501
Passive lysis buffer 5x	Promega	E194A
TaqMan Gene Expression Assays_ *F9*	Thermo Fisher Scientific	Hs01592597_m1
TaqMan Gene Expression Assays_*Gapdh*	Thermo Fisher Scientific	Mm99999915_g1
anti-human FIX antibody	CEDARLANE	CL20039AP
anti-human FIX antibody conjugated with horseradish peroxidase	Affinity Biologicals	GAFIX-APHRP
D-Luciferin sodium salt	OZ Biosciences	#LN10000
ABTS Microwell Peroxidase Substrate	Seracare	5120-0032
Critical commercial assays		
NucleoSpin Plasmid EasyPure	Takara Bio	# U0727
NucleoBond Xtra Midi Plus EF	Takara Bio	# U0420
NucleoSpin Gel and PCR Clean-Up	Takara Bio	# U0609
NucleoSpin Tissue	Takara Bio	# U0952
Prealbumin ELISA Kit	Aviva Systems Biology	#OKIA00111
Deposited data		
AsCas12f–sgRNA–DNA complex coordinates	This paper	PDB: 8J12
AsCas12f-YHAM–sgRNA_ΔS3–5_v7–DNA complex coordinates	This paper	PDB: 8J1J
AsCas12f-HKRA–sgRNA_3–5_v7–DNA complex coordinates	This paper	PDB: 8J3R
AsCas12f–sgRNA–DNA complex EM map	This paper	EMDB: EMD-35912
AsCas12f-YHAM–sgRNA_ΔS3–5_v7–DNA EM map	This paper	EMDB: EMD-35926
AsCas12f-HKRA–sgRNA_3–5_v7–DNA EM map	This paper	EMDB:EMD-35965
Data sets from DMS, self-targeting library, and GUIDE-seq	This paper	BioProject ID: PRJNA937472
Experimental models: Cell lines		
E. coli Mach1	Thermo Fisher Scientific	C862003
E. coli Rosetta 2 (DE3)	Novagen	71397
HEK293	JCRB Cell Bank	#JCRB9068
HT1080	JCRB Cell Bank	#JCRB9113
AAVpro 293T	Takara Bio	#632273
Huh-7	RIKEN BRC	#RCB1942
TLR3	JCRB Cell Bank	IFO50380

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Lenti-X 293T	Clontech	Cat# 632180
Human iPS cells	PMID: 33749676	AR21-5
DMD Knock-out human iPS cells	This paper	DMD #2-16, 33, 61, 84
Experimental models: Organisms/strains		
Mouse: C57BL/6J wild type	SLC Japan, Inc.	N/A
Mouse: B6.129P2-*F9*^{tm1Dws}	The Jackson Laboratory	RRID:IMSR_JAX:004303
Oligonucleotides		
Sequences used for structural analysis, see Table S1	This paper	Table S1
Spacer and designed full sgRNA sequences, see Table S2	This paper	Table S2
Primers for qPCR and NGS, see Table S3	This paper	Table S3
Recombinant DNA		
pE-mH6-AsCas12f	This paper	https://benchling.com/s/seq-NjTfYidGYdXGiBCkyxBQ?m=slm-vlGrnO4tH1mikHG33rGB
lentiCRISPR v2	Addgene	Cat# 52961
lentiGuide-RFP670	This paper	N/A
pBS sgRNA (AsCas12f)	This paper	N/A
pBS sgRNA (SpCas9)	This paper	N/A
pBS gRNA (AsCas12a)	This paper	N/A
pMD2.G	Addgene	Cat# 12259
psPAX2	Addgene	Cat# 12260
Cbhv5 AsCas12f DMD	This paper	N/A
Cbhv5 AsCas12f-HKRA DMD	This paper	N/A
pAAV HCRhAAT AsCas12f Ttr	This paper	N/A
pAAV HCRhAAT AsCas12f-HKRA Ttr	This paper	N/A
pAAV HCRhAAT AsCas12f Alb	This paper	N/A
pAAV HCRhAAT AsCas12f-HKRA Alb	This paper	N/A
pAAV Alb-EGFP	This paper	N/A
pAAV Alb-F9 Padua	This paper	N/A
pRC8	This paper	N/A
pRC6	Takara Bio	Cat# 6651
pHelper	Takara Bio	Cat# 6651
pLenti CMV AsCas12f EF1a EGFP	This paper	N/A
pLenti CMV SplitGFP-VEGFA	This paper	N/A
pLenti CMV SplitGFP-TTR	This paper	N/A
pLenti CMV d2EGFP	This paper	N/A
pLenti U6 sgRNA EF1a RFP670	This paper	N/A
pcDNA4TO mCherry	This paper	N/A
pcDNA4TO AsCas12f	This paper	N/A
pcDNA4TO SpCas9	This paper	N/A
pcDNA4TO AsCas12a	This paper	N/A
pLenti CMV AsCas12f PGK mCherry	This paper	N/A
pLenti CMV SpCas9 PGK mCherry	This paper	N/A
pLenti CMV AsCas12a PGK mCherry	This paper	N/A
pAAV mTTR AsCas12f1 sgRNA_Alb-F9 Padua donor	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pAAV mTTR AsCas12f1-HKRA sgRNA_Alb-F9 Padua donor	This paper	N/A
pSI545 CMV AsCas12f1-HKRA (D225A) VP64	This paper	N/A
pcDNA3 MS2-p65-HSF1	This paper	N/A
pcDNA3 miniCMV (AsCas12f1-HEXA)-Luciferase	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem2	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem3	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem4	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem2+3	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem2+4	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem3+4	This paper	N/A
pBS sgRNA for AsCas12f1_HEXA MS2 stem2+3+4	This paper	N/A
pAAV CMV AsCas12f (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
pAAV CMV AsCas12f-HKRA (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
pSI545 CMV VP64-AsCas12f1-HKRA (D225A)	This paper	N/A
pSI545 CMV VP64-AsCas12f1-HKRA (D225A)-VP64	This paper	N/A
pSI545 CMV AsCas12f1-HKRA (D225A)-VPR	This paper	N/A
pSI545 CMV AsCas12f1-HKRA (D225A)	This paper	N/A
pSI545 CMV AsCas12f1 (D225A)-VP64	This paper	N/A
pAAV mTTR AsCas12f (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
pAAV mTTR AsCas12f-HKRA (D225A)-VP64_P2A_MS2-p65-HSF1_sgRNA (HEXA) Stem4	This paper	N/A
pAAV CMVmini (AsCas12f1-HEXA)-Luciferase	This paper	N/A
Software and algorithms		
EPU software	Thermo Fisher Scientific	https://www.thermofisher.com/jp/ja/home/about-us/events/industrial/smart-epu.html
cryoSPARC v3.3.2 software	Punjani et al.[38]	https://cryosparc.com/updates
AlphaFold2	Jumper et al.[39]	https://alphafold.ebi.ac.uk/download
COOT	Emsley and Cowtan[40]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PHENIX	Afonine et al.[41]	https://www.phenix-online.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MolProbity	Chen et al.[42]	https://www.phenix-online.org/documentation/reference/molprobity_tool.html
CueMol	N/A	http://www.cuemol.org
ImageJ	NIH	https://imagej.nih.gov/ij/
FlowJo	Three Star	https://www.flowjo.com/
Prism9	GraphPad	https://www.graphpad.com/scientific-software/prism/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Osamu Nureki (nureki@bs.s.u-tokyo.ac.jp).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- The atomic models have been deposited in the Protein Data Bank under the accession codes 8J12 (AsCas12f–sgRNA–DNA complex), 8J1J (AsCas12f-YHAM–sgRNA_ΔS3–5_v7–DNA) and 8J3R (AsCas12f-HKRA–sgRNA_ΔS3–5_v7–DNA). The cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under the accession codes EMD-35912 (AsCas12f–sgRNA–DNA complex), EMD-35926 (AsCas12f-YHAM–sgRNA_ΔS3–5_v7–DNA) and EMD-35965 (AsCas12f-HKRA–sgRNA_ΔS3–5_v7–DNA).
- Data sets from DMS, self-targeting library, and GUIDE-seq have been deposited with the National Center for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA937472.
- Microscopy data reported in this paper will be shared by the lead contact upon request.
- The data of unprocessed image files have been deposited in the Mendeley Data repository (https://doi.org/10.17632/dn9h5k4tpf.1).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Protein and RNA preparation for structural analysis

The N-terminally His_6-tagged wild-type AsCas12f, AsCas12f-YHAM, and AsCas12f-HKRA proteins were expressed in *Escherichia coli* Rosetta2 (DE3). Transformed *E. coli* cells were cultured at 37°C until the OD_{600} reached 0.8, and protein expression was then induced by the addition of 0.1 mM isopropyl β-D-thiogalactopyranoside (Nacalai Tesque). *E. coli* cells were further cultured at 20°C overnight and harvested by centrifugation. The cells were then resuspended in buffer A (20 mM HEPES-NaOH, pH 7.6, 20 mM imidazole, and 1 M NaCl), lysed by sonication, and centrifuged. The supernatant was mixed with 3 ml Ni-NTA Superflow resin (QIAGEN), and the mixture was loaded into an Econo-Column (Bio-Rad). Proteins were eluted with buffer B (20 mM HEPES-NaOH, pH 7.6, 0.3 M imidazole, 0.5 M NaCl) and then loaded onto a 5-ml HiTrap Heparin HP column (GE Healthcare) equilibrated with buffer C (20 mM HEPES-NaOH, pH 7.6, and 0.5 M NaCl). The proteins were eluted with a linear gradient of 0.5–2 M NaCl and further purified by chromatography on a HiLoad 16/600 Superdex 200 column (GE Healthcare) equilibrated in buffer D (20 mM HEPES-NaOH, pH 7.6, 0.5 M NaCl). The purified proteins were stored at −80°C until use. The wild-type sgRNA and sgRNA_ΔS3–5_v7 were transcribed *in vitro* with T7 RNA polymerase and purified by 10% denaturing (7 M urea) polyacrylamide gel electrophoresis.

Electron microscopy sample preparation and data collection

The AsCas12f–sgRNA–target DNA ternary complex was reconstituted by mixing purified AsCas12f, the 222-nt sgRNA, the 38-nt target DNA, and the 38-nucleotide non-target DNA at a molar ratio of 1:1:1:1. Each DNA strand has phosphorothioate modifications within the phosphate backbone around the cleavage site to inhibit DNA hydrolysis (Table S1). The AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA ternary complexes were reconstituted in the same way. The ternary complexes were purified by size-exclusion chromatography on a Superdex 200 Increase 10/300 column (GE Healthcare)

equilibrated with buffer E (20 mM HEPES-NaOH, pH 7.6, 50 mM NaCl, 10 mM $MgCl_2$, and 10 μM $ZnCl_2$). The purified complex solution ($A_{260\ nm}$ = 10 for wild type and $A_{260\ nm}$ = 4 for mutants) was then applied to Au 300-mesh R1.2/1.3 grids (Quantifoil) that were glow-discharged after adding 3 μl of amylamine in a Vitrobot Mark IV (FEI) at 4°C, with a waiting time of 10 sec and a blotting time of 4 sec under 100% humidity conditions. The grids were plunge-frozen in liquid ethane and cooled to the temperature of liquid nitrogen.

Micrographs for all datasets were collected with a Titan Krios G3i microscope (Thermo Fisher Scientific) running at 300 kV and equipped with a Gatan Quantum-LS Energy Filter (GIF) and a Gatan K3 Summit direct electron detector in the electron counting mode (The University of Tokyo, Japan). Datasets of the AsCas12f–sgRNA–target DNA ternary complex were collected with a total dose of approximately 50 electrons per $Å^2$ per 48 frames by the standard mode, and datasets of the AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA ternary complexes were collected with a total dose of approximately 50 electrons per $Å^2$ per 64 frames by the CDS mode, using the EPU software (Thermo Fisher Scientific). The dose-fractionated movies were subjected to beam-induced motion correction and dose weighting using Patch Motion Correction, and the contrast transfer function (CTF) parameters were estimated using Patch-based CTF estimation in cryoSPARC v3.3.2.

Single-particle cryo-EM data processing

Data were processed using cryoSPARC v3.3.2.[38] For the AsCas12f–sgRNA–target DNA ternary complex, 1,304,079 particles were initially selected from the 4,027 motion-corrected and dose-weighted micrographs using Blob picker and Topaz, and extracted at a pixel size of 3.32 Å. These particles were subjected to several rounds of 2D classification to curate particle sets. The particles were further curated by heterogeneous refinement, using a map derived from *ab initio* reconstruction as the template. The selected 150,071 particles were then re-extracted at a pixel size of 1.33 Å and subjected to 3D variability analysis. The resulting maps with different conformations were used for subsequent heterogeneous refinement. The selected particles after heterogeneous refinement were refined using non-uniform refinement with optimization of the CTF value,[43] yielding a map at 3.08 Å, according to the Fourier shell correlation (FSC) criterion of 0.143.[44]

The datasets for the AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA ternary complexes were processed using cryoSPARC, in a similar manner to that used for the wild-type AsCas12f–sgRNA–target DNA ternary complex. For data processing details, see Figure S1.

Model building and validation

The model was built using the predicted model of the AsCas12f protein created by AlphaFold2[39] (the AlphaFold model) as the reference, followed by manual model building with COOT[40] against the density map sharpened using DeepEMhancer. The model was refined using phenix.real_space_refine ver. 1.20.1[41] with secondary structure, base pair, and metal coordination restraints. The metal coordination restraints were generated using ReadySet, as implemented in PHENIX. Structure validation was performed using MolProbity.[42] In the model of the wild-type AsCas12f–sgRNA–target DNA ternary complex, residues 56–65, 268–284, 326–330, 380–384, and 422 of AsCas12f.2, nucleotides (−202) to (−194), (−70) to (−14), and 19 to 20 of the sgRNA, nucleotides (−8) to (−1) and 29 to 30 of the TS, and nucleotides (−12*) to (−11*) and 2* to 26* of the NTS were not included in the final model due to their poor resolutions in the density map. In the model of the AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA ternary complex, residues 59–64, 266–284, 327–330, 380–384, and 422 of AsCas12f.2, nucleotides (−96) to (−95) and 19 to 20 of the sgRNA, nucleotides (−8) to (−1) and 29 to 30 of the TS, and nucleotides (−12*) to (−11*) and 2* to 26* of the NTS were not included in the final model due to their poor resolutions in the density map. In the model of the AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA ternary complex, residues 59–64, 266–284, 327–330, 380–384, and 422 of AsCas12f.2, nucleotides (−96) to (−95) and 19 to 20 of the sgRNA, nucleotides (−8) to 2 and 26 to 30 of the TS, and nucleotides (−12*) to (−8*) and 2* to 26* of the NTS were not included in the final model due to their poor resolutions in the density map.

Plasmid generation for cell biological experiments

Using a plasmid encoding both AsCas12f and sgRNA (addgene #171614), AsCas12f and its mutants were cloned into pcDNA4TO (Thermo Fisher Scientific) using a NEBuilder HiFi DNA Assembly Kit (New England Biolabs). The sgRNAs were separately cloned into pBS or lentiGuide (addgene #52963). Other Cas enzymes, including AsCas12a, UnCas12f, SpCas9 and their mutants, were mutagenized and cloned into pcDNA4TO (Thermo Fisher Scientific) with the following plasmids: addgene #114091, #176269, and #98290, respectively. The engineered sgRNA was designed by extracting stem sequences. Both the 5′ and 3′ side fragments were cloned into pBS. The reporter for genome editing was constructed with GFP1-10, GFP11 and the linker containing the target sequence. This sequence was cloned into pLenti CMV (addgene #17448).

Cell culture

HEK293T cells (Clontech) were cultured at 37 °C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM, WAKO), containing 10% fetal bovine serum (FBS, Thermo Fisher Scientific) and penicillin/streptomycin (100 U/ml, Thermo Fisher Scientific). HEK293 cells (JCRB Cell Bank), HT1080 cells (JCRB Cell Bank), and AAVpro 293T cells (Takara Bio) were cultured in DMEM (Sigma-Aldrich), supplemented with 10% FBS (Thermo Fisher Scientific) and GlutaMAX (Thermo Fisher Scientific). Huh-7 cells (RIKEN BRC) were cultured in RPMI 1640 (Sigma-Aldrich) supplemented with 10% FBS and 2 mM L-glutamine. Murine immortalized liver cells (TLR3

cells, JCRB Cell Bank) were maintained in DMEM containing 2% FBS, 5 ng/ml of human EGF, and ITS-X Supplement (Thermo Fisher Scientific). All cell lines tested negative for mycoplasma contamination.

Lentivirus production

Ten-centimeter plates of 70% confluent HEK293T (Clontech) cells were transfected with 9 μg of the plasmid library, 6 μg of the psPAX2 vector and 3 μg pMD2.G, using Fugene HD (Promega) according to the manufacturer's instructions. Supernatants were collected after 48 h, centrifuged at 3,000 rpm for 10 min at 4 °C, filtered through a 0.45 μm low protein-binding filter (SFCA), and then frozen at –80 °C.

T7 endonuclease assay and amplicon sequencing

One day before transduction, the cells (5×10^4 cells/well) were seeded into 48 well plates coated with collagen type I (Cellmatrix type I-C, Nitta Gelatin). The plasmids (200 ng) were incubated together with Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's recommendations, and then directly added to the cell culture. At 72 h after the transduction, the cells were lysed with the SimplePrep reagent for DNA (Takara Bio). The supernatants were directly used for polymerase chain reaction (PCR). DNA fragments were amplified with ExTaq DNA polymerase (Takara Bio). Purified PCR products were denatured and re-annealed using a thermal cycler, and then treated with T7 endonuclease (Nippon Gene). DNA fragments were analyzed by a microchip electrophoresis system (MCE-202 MultiNA; Shimadzu). The target gene sequences and primer sequences used in the analysis are described in Tables S2 and S3, respectively. When indicated, PCR amplicons were subjected to 150 bp pair-end read sequencing using the Illumina MiSeq at The University of Tokyo, or to 300 bp pair-end read sequencing using the Illumina MiSeq at Bioengineering Lab. Co., Ltd. (Kanagawa, Japan). The frequencies of the mutations were assessed by CRISPResso2.[45]

DMS Library construction, FACS, and Illumina sequencing analysis

The AsCas12f sequence was divided into 6 units, and saturation mutagenesis was introduced separately in these 6 units. Pooled oligonucleotides with degenerate NNK codons were synthesized by Integrated DNA Technologies, Inc. Synthesized oligos were extended by overlap PCR and cloned into pLenti CMV EF1-GFP, modified from the plasmid (addgene #17448) using an NEBuilder HiFi DNA Assembly Kit (New England Biolabs). An 80 μl ligation reaction was performed using 10 μg of the gel-purified inserts and 5 μg of the vector, and was then purified with a Gel and PCR Clean-up kit (Takara Bio) and eluted by 20 μl of distilled water. Next, 200 μg (~4 μl) of the purified reaction was transformed into 50 μl of electrocompetent cells (Lucigen) and expanded according to the manufacturer's protocol with 1500 V electroporation by an ECM 399 system (BTX). A 1,000-fold dilution of the full transformation was plated to estimate the scale of the mutant library. The six libraries were packaged into the lentivirus separately to avoid swapping contamination, and then transduced into HEK293T cells. At 48 h after transduction, the GFP-expressing library cells were sorted. The six libraries were pooled and further cultured for 5 days, and then cells were sorted on a MA900 cell sorter (Sony). Dead cells, doublets, and debris were excluded by first gating on the main population by forward and side scatter. GFP-positive and -negative cells were collected. The total numbers of collected cells were about 10 million cells for each group. Total RNA was extracted from collected cells with TRIzol (Life Technologies) and Direct-zol RNA MiniPrep (Zymo Research Corporation), according to the manufacturers' protocols. First-strand complementary DNA (cDNA) was synthesized with PrimeScript II Reverse Transcriptase (Takara Bio) primed with a gene-specific oligonucleotide. After cDNA synthesis, each library was amplified with specific primers. Following a second round of PCR, primers with added adapters for annealing to the Illumina flow cell and sequencing primers, together with barcodes for experiment identification, were used. The PCR products were sequenced on an Illumina NovaSeq 6000 using a 2 × 150 nucleotide paired-end protocol, in the Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University. Data were analyzed by comparing the read counts within each group normalized relative to the WT sequence read-count. Editing rates for all individual mutations were calculated as the ratio of GFP-negative reads to the total (GFP-positive and -negative) reads.

Self-targeting library construction, FACS, and Illumina sequencing analysis

The custom pooled oligonucleotides containing pairs of sgRNAs, corresponding target sequences and barcodes were synthesized by Twist Bioscience. The library includes NTTN 16 PAM sequences and 750 random target sequences, adapted from a previous paper,[46] per PAM sequence. Synthesized pooled oligonucleotides were extended by overlap PCR and cloned into lentiGuide EF1-RFP (adapted from addgene #52963), packaged into lentivirus, and transduced into HEK293T cells in a multiple copy infection manner, using the same protocol as for the DMS library. These library cells were infected with lentivirus carrying AsCas12f or enAsCas12f and PGK-mCherry, and sorted with mCherry 2 days after infection. The cells were maintained for 7 days and then harvested to extract DNA. This library contains 12,000 different sequences and for more than 500-fold coverage, at least 6 million cells were maintained in each step. Genomic DNA was extracted by the salt precipitation method, as previously reported.[47] From one confluent 100 mm dish, ~15 million cells were harvested in a 15 ml conical tube, and 6 ml of NK Lysis Buffer (50 mM Tris, pH 8.0, 50 mM EDTA, 1% SDS) and 30 μl of 20 mg/ml Proteinase K (QIAGEN 19131) were added and the mixture was incubated at 55°C overnight. The next day, 30 μl of 10 mg/ml RNase A (QIAGEN 19101) was added to the lysed sample, which was then inverted 25 times and incubated at 37°C for 30 min. Samples were then cooled on ice. To precipitate the proteins, 2 ml of pre-chilled 7.5 M ammonium acetate (Sigma-Aldrich A1542) was added and the samples were vortexed at high speed for 20 s or until a white, turbid solution was obtained. After

centrifugation at 8,000 × g for 10 min, a tight pellet was visible in each tube and the supernatant was carefully decanted into a new 15 ml conical tube. After 6 ml 100% isopropanol was added, the tube was briefly vortexed and centrifuged at 8,000 × g for 10 min. Genomic DNA was visible as a small white pellet in each tube. The supernatant was discarded, and 6 ml of freshly prepared 70% ethanol was added. The tube was vortexed briefly, and then centrifuged at 8,000 × g for 1 min. The supernatant ethanol was removed completely, and the pellet was air-dried for 10–30 min. After the pellet turned slightly translucent, 500 μl of TE buffer was added and the tube was incubated at 65°C for 1 h and at room temperature overnight to fully resuspend the DNA. After gDNA extraction, 100 μg portions were subjected to PCR to attach the adapters and barcodes. The PCR products were purified with a Gel and PCR Clean-up kit (Takara Bio) and sequenced on an Illumina NovaSeq 6000, using a 2 × 150 nucleotide paired-end protocol, in the Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University.

GUIDE-Seq

GUIDE-seq experiments were performed essentially in accordance with the protocols described previously.[23] Briefly, 2 x 10^5 HEK293T cells were transfected with 360 ng enAsCas12f- or SpCas9-encoding plasmid, 360 ng sgRNA expression plasmid, and 50 pmol end-protected double-strand oligodeoxynucleotide (dsODN) GUIDE-seq tag that encodes an internal NdeI restriction site, on a 4D-Nucleofector with the SE Cell Line Nucleofector Kit and the CM-137 program, according to the manufacturer's instructions (Lonza). Genomic DNA was isolated with a NucleoSpin Tissue Kit (Takara Bio) and sheared to an average size of 500-bp with 6 cycles of 30 s sonication in mode H of a Bioruptor UCD-250 (Cosmo Bio Co., Ltd.). Restriction enzyme assays with NdeI and T7E1 were performed at the target loci to determine tag integration efficiencies and modification percentages, respectively. High-throughput sequencing libraries were generated in accordance with the protocols and subsequently sequenced with about 10 million reads per sample, using an Illumina Novaseq 6000 system in a 150-bp paired-end mode. The obtained sequence data were analyzed using the published pipeline (https://github.com/tsailabSJ/guideseq) with the following modifications: demultiplexing with bcl2fastq (v2.20.0.422), human reference genome (hg19), window size of 80-bp, and configurable PAM sequence.

Plasmid constructs for AAV vector production

A DNA fragment encoding a promoter, the AsCas12f cDNA, the SV40 polyadenylation signal, and an sgRNA sequence driven by the U6 promoter was introduced between inverted terminal repeats into the pAAV plasmid. We employed the cytomegalovirus (CMV) promoter and the liver-tropic promoter [HCRhAAT promoter (an enhancer element of the hepatic control region of the Apo E/C1 gene and the human anti-trypsin promoter) or murine transthyretin (*Ttr*) promoter] used for the cell transduction and the mouse liver transduction *in vivo*, respectively. The guide RNA sequences for AsCas12f used in this study are shown in Table S2. We also created a pAAV donor plasmid containing a P2A self-cleaving peptide sequence-conjugated EGFP sequence or human FIX with the Padua mutation, possessing homologous arms at the target site for inserting the target gene only by HDR. For DMD experiments, AsCas12f and sgRNA sequences were inserted in the AgeI-EcoRI sites and XhoI-NotI sites of the Cbhv5 plasmid, respectively. The AAV genes were packaged by triple plasmid transfection of AAVpro293T cells to produce the AAV vector (helper-free system), as described previously.[48] The titers of recombinant AAV vectors were determined by quantitative PCR.[49]

Human iPS cell-derived cardiomyocyte model of DMD

To model the cardiomyocytes of Duchenne muscular dystrophy, the parental hiPS cell line (AR21-5) and four *DMD* exon 44 deleted cell lines (2-16, 33, 61, 84) were differentiated toward cardiomyocytes using the previously reported method.[50] Briefly, 125,000 undifferentiated hiPS cells were plated in the wells of a 6-well plate, cultured in AK02N for 4 days, and then induced to cardiomyocytes by serial treatments with CHIR99021 (6 μM, days 0–2) and Wnt-C59 (2 μM, days 2–4), in RPMI 1640 supplemented with B27 minus insulin and GlutaMAX. Cells were further cultured in RPMI 1640+B27-Insulin for another 3 days, and purified by puromycin selection in RPMI 1640 with B27 for 3 to 8 days. At day 11 or 15, cardiomyocytes were replated in iMatrix and 0.1% gelatin coated dishes for genome editing experiments. One day after replating, the AAVs encoding AsCas12f or enAsCas12f-HKRA targeting *DMD* exon 45 were transduced to iPSC-derived cardiomyocytes at 1 x 10^4 vg/cell to 5 x 10^5 vg/cell. Two separate AAVs encoding split SpCas9-N and -C were used as controls.

Confirmation of DMD exon skipping by genomic and droplet digital PCR

DNA and proteins were isolated at 6 to 13 days after transduction, and subsequently subjected to PCR and western blotting to confirm exon skipping. The deletion of DMD exon 45 was detected by genomic PCR (Fwd, CTTGTGGGGACAAGAAATCG; Rev, GCTCAAGTTCCCCTTCAAGA) and droplet digital PCR (HEX-conjugated dHsaCP2506429 for *DMD* exon 26 as a control; FAM-conjugated dHsaCNS992188911 for *DMD* exon 45; Bio-Rad). Quantitative digital droplet PCR was performed using the QX100 Droplet Digital PCR system (Bio-Rad), according to the manufacturer's instructions. The plate was heat-sealed with a foil seal and amplified on the C1000 Touch thermal cycler (Bio-Rad) using the following protocol: 95°C for 10 min, 40 cycles of 94°C for 30 s and 53°C for 1 min, and 98°C for 10 min. Quantification was determined based on the number of positive droplets in each reaction, using the QuantaSoft Analysis software (Bio-Rad).

Western blotting of dystrophin

For western blotting of the dystrophin protein, iPSC-CMs were lysed in Laemmli SDS sample buffer supplemented with DTT (100 mM), cOmplete™ EDTA-free protease inhibitor cocktail (11836170001, Roche), phosphatase inhibitor cocktail 2 (P5726, Sigma-Aldrich) and phosphatase inhibitor cocktail 3 (P0044, Sigma-Aldrich). After heating at 95°C for 5 min, samples were resolved on 4–15% polyacrylamide gels (Mini Protean TGX gels, Bio-Rad) and transferred to 0.2 μm PVDF membranes with a trans-blot turbo transfer system (Bio-Rad). The blot was blocked using PVDF blocking reagent (NYPBR01, TOYOBO). Primary and secondary antibodies were each diluted in Can Get Signal Solution (NKB-101, TOYOBO). The blot was incubated with mouse anti-dystrophin antibody (1:1,000, D8168, Sigma-Aldrich), mouse anti-dystrophin antibody (1:1,000, sc-73592, Santa Cruz) or mouse anti-GAPDH antibody (1:1,000, sc-32233, Santa Cruz) at room temperature for 1 h, and then with peroxidase AffiniPure donkey anti-mouse IgG (H+L) (1:2,000, Jackson ImmunoResearch) at room temperature for 30 min. The protein band was visualized using Immobilon Western Chemiluminescent HRP Substrate (WBKLS0500, Merck), according to the manufacturer's instructions.

Immunofluorescent staining of dystrophin

We also replated cardiomyocytes for an immunofluorescent analysis to confirm the restoration of dystrophin in cardiomyocytes. Immunofluorescence staining of DMD was performed as described previously.[51] Cells were fixed with ice-cold acetone for 10 minutes at −20°C, and then blocked with 10% donkey serum (S30-100ML, Sigma-Aldrich) / TBS-T (T9141, Takara Bio) for 1 h at room temperature. Primary and secondary antibodies were added to cells in blocking buffer for 2 h and 1 h, respectively. Nuclei were counterstained using Hoechst 33342 (62249, Thermo Fisher Scientific). Antibodies used in this article were dystrophin (MANDYS8, D8168, Sigma-Aldrich, 1:800 dilution) and Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 488 (A32766, Thermo Fisher Scientific, 1:500 dilution). All imaging was performed using an Olympus FV1000 laser scanning confocal microscope.

Animal experimentation

All animal experimental procedures were approved by The Institutional Animal Care and Concern Committee of Jichi Medical University (permission number: 20051-8), and animal care was conducted in accordance with the committee's guidelines and ARRIVE guidelines. Coagulation factor IX (FIX)-deficient mice (B6.129P2-F9tm1Dws) were obtained from The Jackson Laboratory (Sacramento, CA). C57BL/6 mice were purchased from SLC Japan (Shizuoka, Japan). To obtain plasma samples, mice were anesthetized with isoflurane (1–3%), and the blood sample was drawn from the jugular vein using a 29G micro-syringe (TERUMO) containing 1/10 (volume/volume) sodium citrate. Platelet-poor plasma was obtained by centrifugation and then frozen and stored at −80°C until analysis. The AAV vector was administered intravenously through the jugular vein (100–150 μL) and intraperitoneally (10 μL) in adult and neonatal mice, respectively.

Quantitative RT-PCR

Total RNA was isolated from cells with an RNeasy Mini Kit (QIAGEN). The RNA samples were reverse-transcribed using a PrimeScript RT Reagent Kit (Takara Bio). Quantitative real-time PCR was performed using TaqMan Gene Expression Assays [(*F9*: Hs01592597_m1; *Gapdh*: Mm99999915_g1 (Thermo Fisher Scientific)] and THUNDERBIRD Probe qPCR Mix (TOYOBO) in a QuantStudio 12K Flex Real-Time PCR system (Thermo Fisher Scientific). Reactions were analyzed in duplicate, and *F9* expression levels were normalized to *Gapdh* mRNA levels.

FIX activity and ELISA

Human FIX activity (FIX:C) was measured with a one-stage clotting-time assay and an automated coagulation analyzer (Sysmex CS-1600). Plasma coagulation time was measured as activated partial thromboplastin time (APTT), using an automated coagulation analyzer (Sysmex CA-500). The FIX antigen (FIX:Ag) was measured as follows: microtiter plates were coated with an anti-human FIX antibody (CEDARLANE). After blocking with 5% casein, diluted plasma samples were incubated for 1 h at 37°C. The antigen binding was detected with the anti-human FIX antibody conjugated with horseradish peroxidase (Affinity Biologicals) and ABTS microwell peroxidase substrate (Seracare). Plasma transthyretin was measured with a Prealbumin ELISA Kit (Aviva Systems, San Diego, CA), according to the manufacturer's recommendations.

Immunohistochemistry

Anesthetized mice were perfused with phosphate buffered saline (PBS). The isolated tissues were fixed with 4% paraformaldehyde, incubated with PBS containing sucrose, and then frozen in optimal cutting temperature compound (Sakura Fintek Japan, Tokyo, Japan). The tissue sections were blocked with 5% donkey serum and then incubated overnight at 4°C with an anti-GFP antibody (MBL Co., Aichi, Japan) and an anti-CD146 antibody. The sections were incubated with anti-rabbit IgG conjugated with AlexaFluor 488 (Thermo Fisher Scientific) and anti-rat IgG conjugated with AlexaFluor 594 (Thermo Fisher Scientific) for 2 h at 4°C. The sections were mounted with VECTASHIELD Mounting Medium with DAPI (Vector Laboratories, Burlingame, CA, USA). Immunofluorescence staining was observed and photographed using an all-in-one microscope (BZ-X700, Keyence, Tokyo, Japan). The EGFP-positive cells were quantified with the BZ-X 700 imaging software (Keyence).

Measurement of luciferase activity

The *luciferase* gene, driven by the minimal cytomegalovirus promoter conjugated with two *HEXA* sgRNA-binding sequences, was cloned into the pcDNA3 vector (Thermo Fisher Scientific). The linearized plasmid was transduced into Huh-7 cells, using Lipofectamine 3000 (Thermo Fisher Scientific). To select the transfected cell clones, G418 (Nacalai Tesque) was added to the culture medium. For the luciferase activity measurements, the cells were lysed with 100 μl of lysis reagent (Promega), and then aliquots (10 μL) were added into the wells of a 96-well plate. The 96-well plate was placed in a luminometer (Centro LB 960, BERTHOLD Technologies), and 50 μL of the Luciferase Assay Reagent (Promega) was injected into each well, using the automatic injector.

Measurement of luciferase activity by *in vivo* imaging in mice

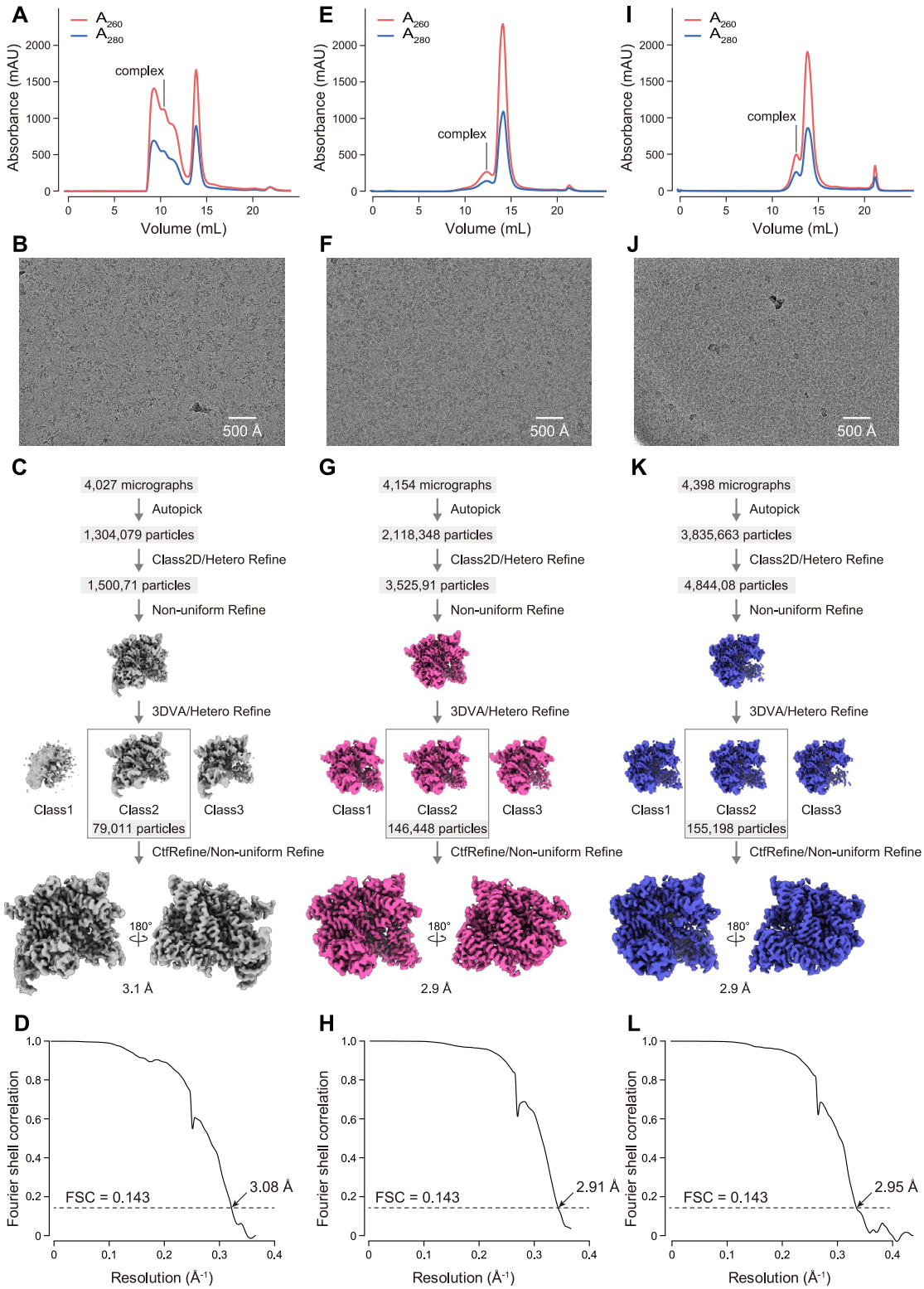
The anesthetized mice were injected intraperitoneally with the luciferin substrate (3 mg/body). We injected two AAV serotype 8 vectors, one encoding dAsCas12f or dAsCas12f-HKRA conjugated with VP64, MS2-p65-HSF1, and sgRNA, and the other encoding luciferase with a *HEXA* sgRNA binding site. Photons derived from luciferase activity were measured by an IVIS® Imaging System and Living Image software (Xenogen Corp., Alameda, CA). Quantitative data were expressed as photon units (photons/second).

QUANTIFICATION AND STATISTICAL ANALYSIS

All data are expressed as mean ± SD. No statistical methods were used to predetermine sample size. Sample size was based on experimental feasibility and sample availability. Samples were processed in random order. Statistical analyses were performed using GraphPad Prism 10 (Graph Pad Software, San Diego, CA). All data are presented as the mean ± standard deviation (SD). Statistical significance was analyzed by two-tailed Student's t test, one-way ANOVA with post hoc Tukey's multiple comparison test, or two-way ANOVA with post post hoc Sidak's multiple comparison test. $P<0.05$ was considered as statistically significant.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Single-particle cryo-EM analysis of the AsCas12f–sgRNA–target DNA ternary complex, related to Figures 1 and 4

Cryo-EM analyses of the AsCas12f–sgRNA–target DNA (A–D), AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA (E–H), and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA (I–L) complexes.

(A, E, and I) Size-exclusion chromatography profiles of the AsCas12f–sgRNA–target DNA (A), AsCas12f-YHAM–sgRNA_ΔS3–5_v7–target DNA (E), and AsCas12f-HKRA–sgRNA_ΔS3–5_v7–target DNA (I) complexes. The peak fraction was used for the following cryo-EM analyses.

(B, F, and J) Representative cryo-EM images, recorded on a 300-kV Titan Krios microscope with a K3 camera.

(C, G, and K) Single-particle cryo-EM image processing workflows.

(D, H, and L) Fourier shell correlation (FSC) curves for the 3D reconstruction, with the gold-standard cut-off (FSC = 0.143) marked with a black dashed line.

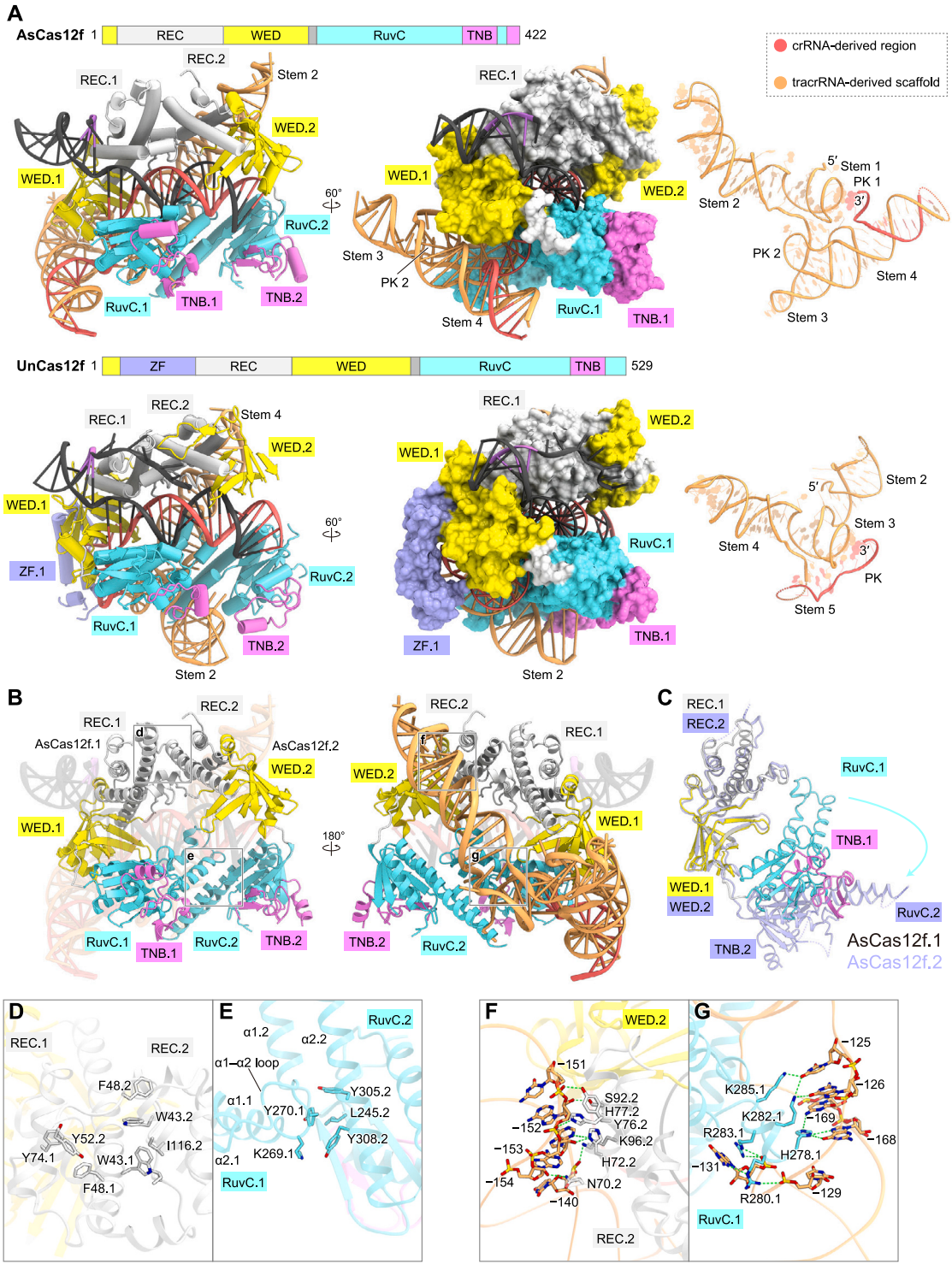


Figure S2. Structural comparison and dimer interfaces of AsCas12f, related to Figures 1 and 2

(A) Structural comparison of AsCas12f with UnCas12f (from an uncultured archaeon) (PDB: 7C7L).

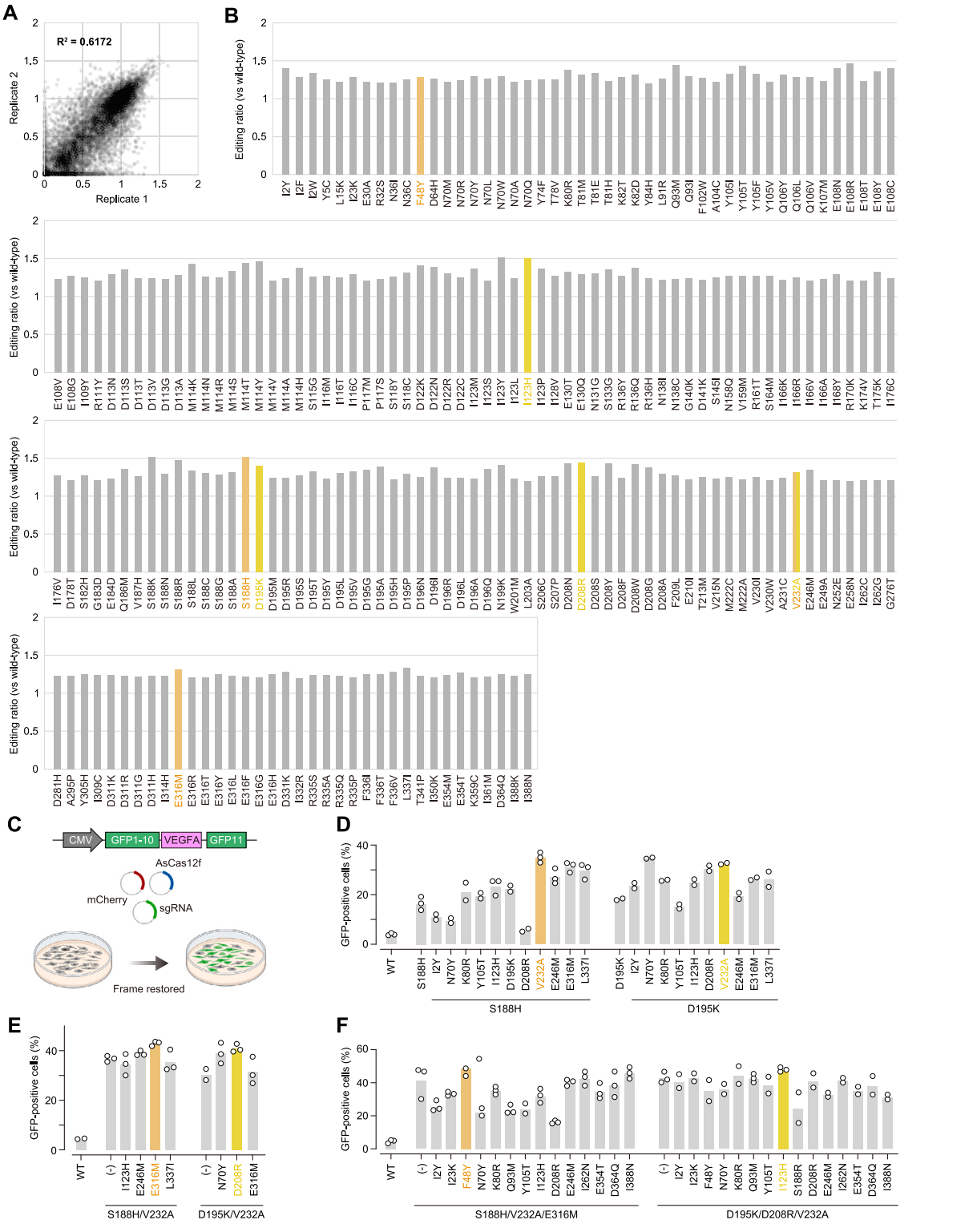
(B) Dimer interfaces between AsCas12f.1 and AsCas12f.2 and recognition sites of the sgRNA scaffold.

(C) Superimposition of AsCas12f.1 and AsCas12f.2, based on their REC lobes.

(D) Dimer interface between REC.1 and REC.2.

(E) Dimer interface between RuvC.1 and RuvC.2.

(F and G) Recognition of stem 2 by AsCas12f.2 (F) and AsCas12f.1 (G).



(legend on next page)

Figure S3. Engineering AsCas12f variants, related to Figure 3

(A) Correlation in single-mutant effects on GFP deletion, as determined from independent mutant library replicates.

(B) List of single amino acid substitutions that resulted in a more than 20% increase in editing efficiency compared with the WT AsCas12f. The mutations contained in AsCas12f variants are highlighted in yellow and orange colors.

(C) Split-GFP-reframing assay to evaluate the genome-editing activity against the VEGFA gene in HEK293T cells. The plasmid-encoded mCherry was co-transfected to identify plasmid-transfected cells.

(D–F) The genome-editing activities of double (D), triple (E), and quadruple (F) AsCas12f mutants.

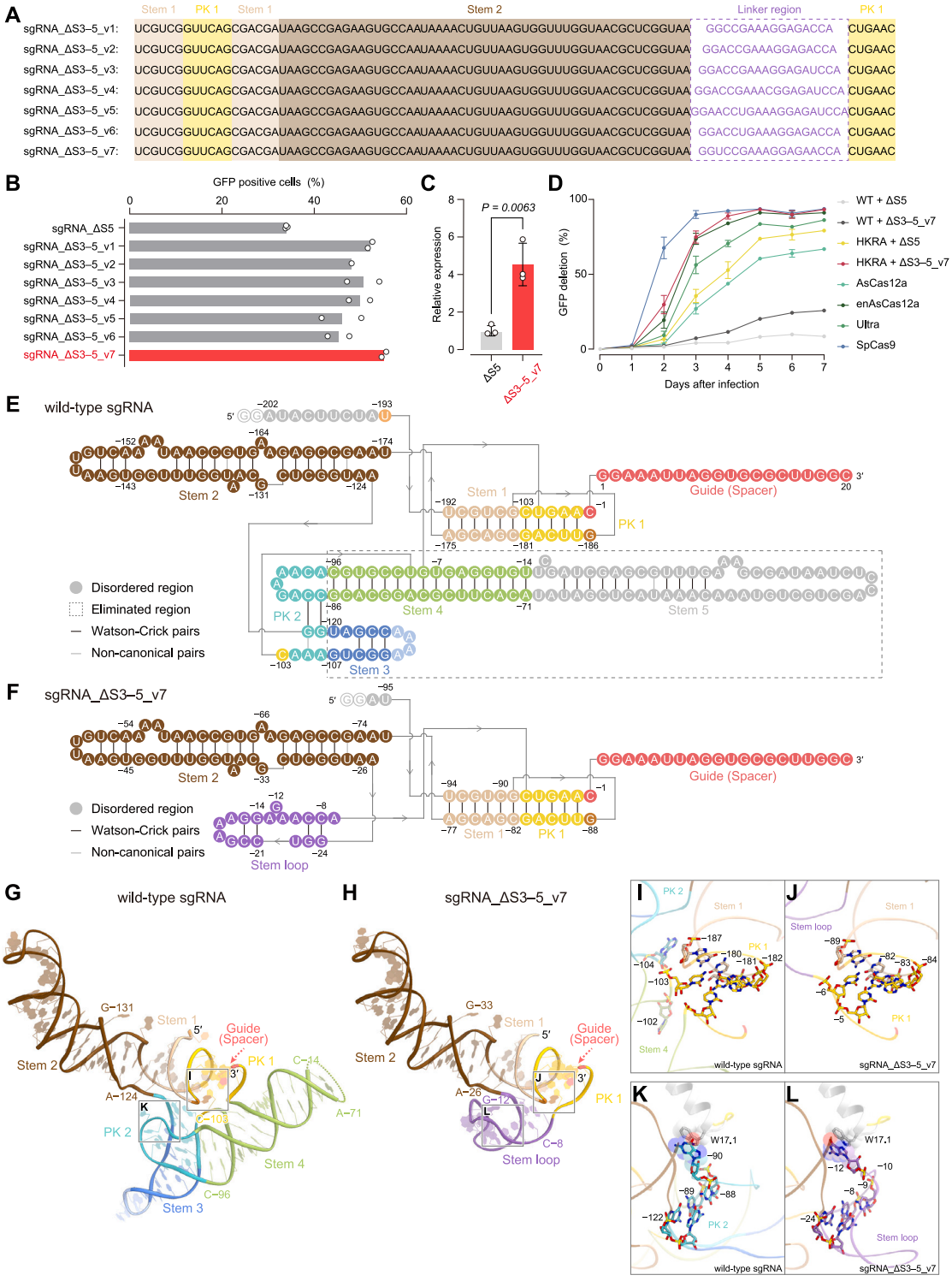


Figure S4. Optimization of sgRNA, related to Figures 3 and 4

(A) Sequences of sgRNA variants (sgRNA_ΔS3–5_v1–v7), designed by eliminating stems 3, 4, and 5 and modifying the length and sequence of the linker region to retain the original architecture of the remaining regions. The linker regions are enclosed in a purple dashed box.

(B) Genome-editing activities of sgRNA variants, determined by flow cytometry.

(C) Expression of sgRNA quantified by qPCR for the common sequence site (n = 3, mean ± SD).

(legend continued on next page)

(D) Time course of GFP deletions induced by various Cas enzymes in HEK293T cells expressing d2EGFP (n = 3, mean ± SD). Each sgRNA was transduced by single-copy lentivirus infection. Although the GFP deletion by AsCas12f-HKRA with sgRNA_ΔS3–5_ v7 is slightly lower than that by SpCas9, it is faster than the deletion rates observed for AsCas12a, Ultra, and enAsCas12a.

(E and F) Schematics of the wild-type sgRNA (E) and sgRNA_ΔS3–5_v7 (F). The disordered regions are colored gray. Our present structure revealed that, although PK1, PK2, stem 1, and stem 2 form extensive interactions with the AsCas12f protein, stems 3 and 4 form few interactions with the protein and stem 5 is disordered. Thus, we designed sgRNA_ΔS3–5_v7 by eliminating stems 3–5, which are enclosed in the dashed box, while retaining the original architecture of the remaining regions.

(G and H) Structures of the wild-type sgRNA scaffold (G) and the sgRNA_ΔS3–5_v7 scaffold (H). The disordered regions are indicated as dotted lines.

(I and J) Close-up views of the PK 1 region of wild-type sgRNA (I) and sgRNA_ΔS3–5_v7 (J). C(−6) in the sgRNA_ΔS3–5_v7, which corresponds to the flipped C(−103) in the wild-type sgRNA (I), base pairs with G(−83) and connects PK 1/stem 1 to form a continuous helix (J).

(K and L) Close-up views of the PK 2 region of wild-type sgRNA (K) and the stem 3 region of sgRNA_ΔS3–5_v7 (L). G(−12) in the sgRNA_ΔS3–5_v7 is flipped out and stacks with Trp17 (L), akin to the interaction between G(−90) and Trp17 in the wild type (K).

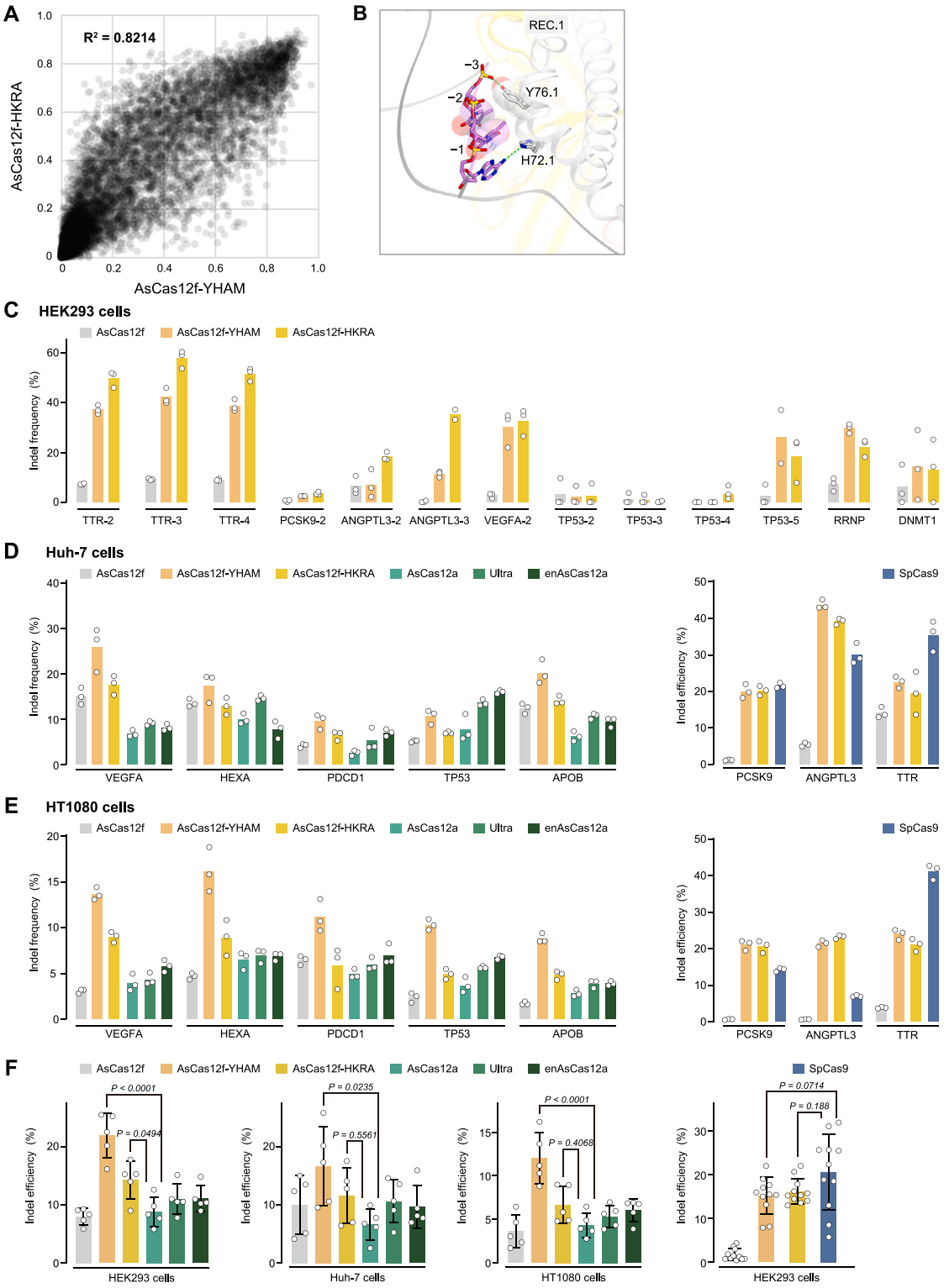


Figure S5. Genome-editing activities of AsCas12f variants in other cells, related to Figure 5

(A) Comparison of target editing efficiencies of AsCas12f-YHAM and -HKRA.

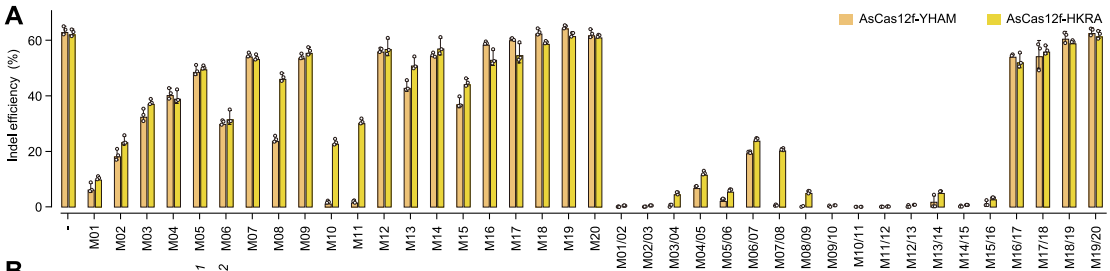
(B) Recognition of the PAM duplex.

(C) Indel efficiencies by AsCas12f, AsCas12f-YHAM, and AsCas12f-HKRA at 13 different target sites in HEK293T cells (n = 3).

(legend continued on next page)

(D and E) Indel efficiencies by AsCas12f, AsCas12f-YHAM, AsCas12f-HKRA, AsCas12a, Ultra, enAsCas12a, and SpCas9 in Huh-7 cells (D) and HT1080 cells (E) (n = 3).

(F) Average indel efficiencies for all tested loci in each cell. White circles indicate the average indel efficiencies of each locus. Statistical significance was analyzed by one-way ANOVA with post hoc Tukey's multiple comparison test.



B

	APOB-1	*APOB-2*	*PDGFRA-1*	*PDGFRA-2*	*KDR-1*	*KDR-2*	*SCNN1D*	*VEGFA*	*TTR*
SpCas9	4	73	2	1	14	3	7	1	12
AsCas12f-HKRA	7	15	9	4	3	17	32	3	4

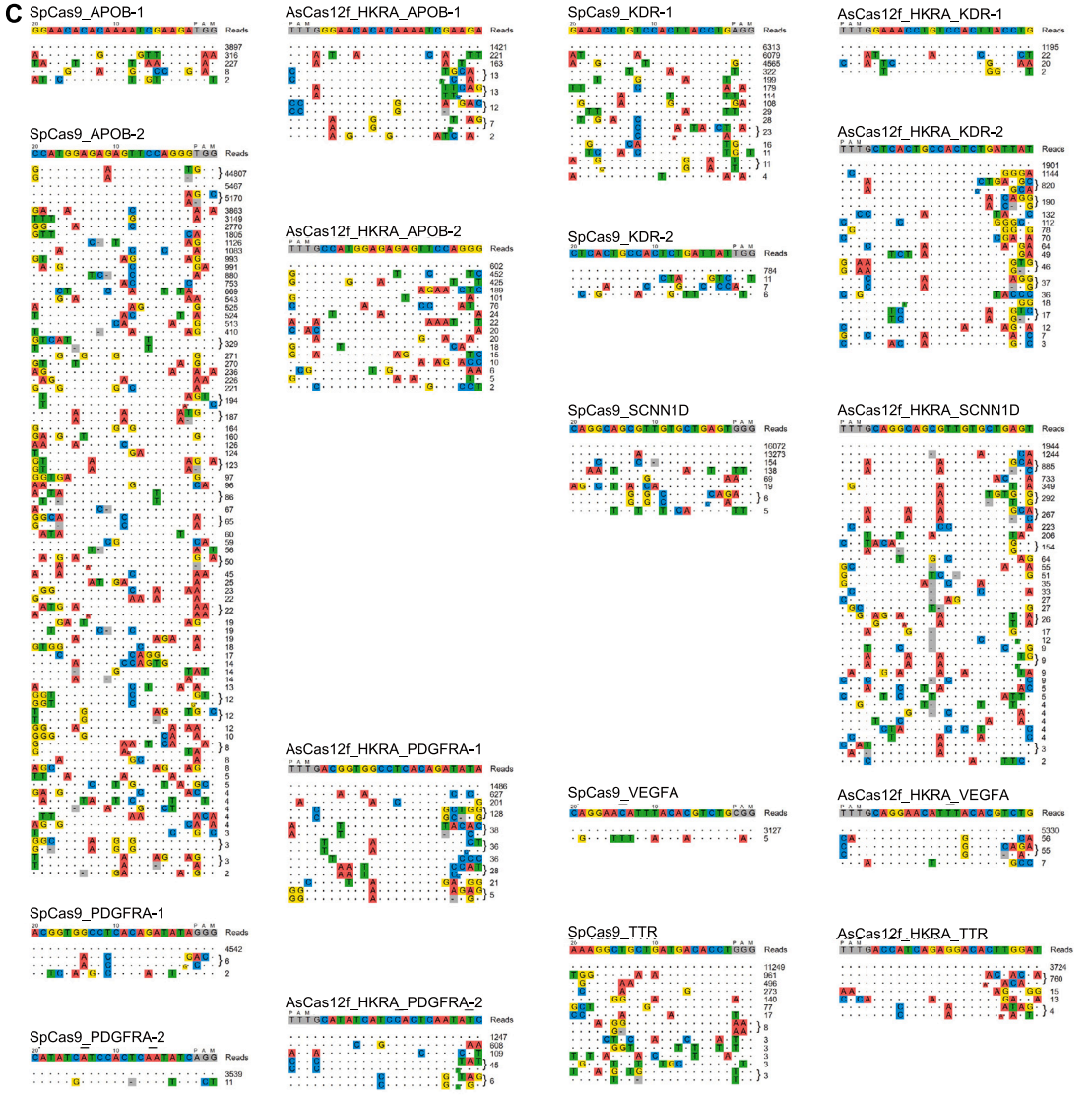


Figure S6. Specificity of AsCas12f variants, related to Figure 5

(A) Impacts of single and paired mismatches on AsCas12f-YHAM- and AsCas12f-HKRA-mediated EGFP activation activities in a Split-GFP-reframing assay with the *TTR* linker sequence (n = 3).

(B) The number of GUIDE-seq detected off-target sites for AsCas12f-HKRA and SpCas9 on the same spacer sites as in Figure 5D.

(C) Off-target sites identified by GUIDE-seq.

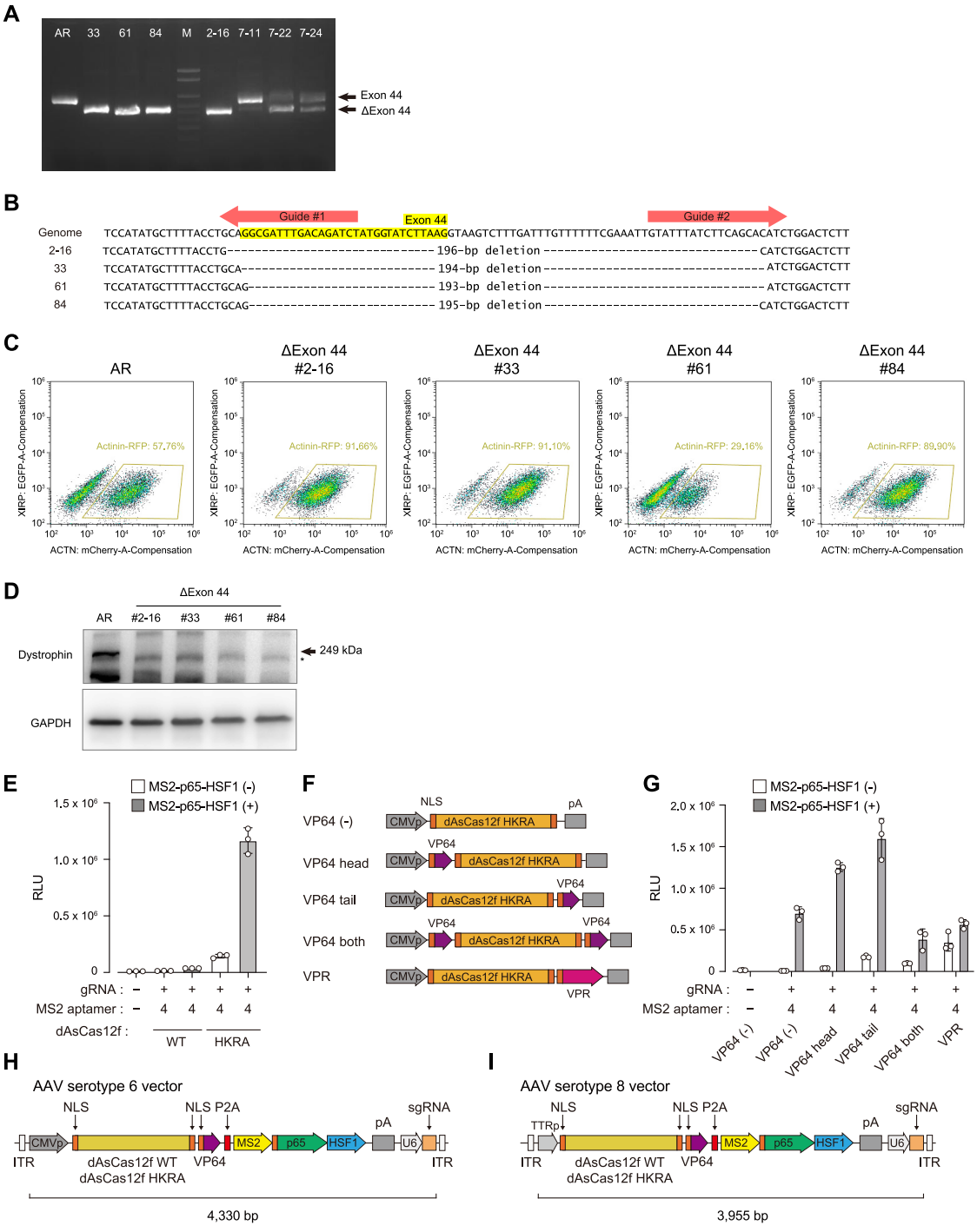


Figure S7. Application of compact enAsCas12f enzyme, related to Figures 6 and 7

(A) PCR-confirmed complete deletion of *DMD* exon 44.

(B) Genomic sequences of *DMD* exon 44-deleted iPS cell lines.

(C) Flow cytometry analysis of alpha-actinin-mCherry at day 14 of cardiac differentiation. All DMD exon 44-deleted lines differentiated toward cardiomyocytes efficiently without purification.

(D) Western blot analysis of AR and DMD iPSC cardiomyocytes (#2–16, 33, 66, and 84). The arrow indicates the immunoreactive bands of dystrophin, and the asterisk indicates a non-specific band.

(E) Comparison of luciferase activities between dAsCas12f and dAsCas12f-HKRA (n = 3, mean ± SD).

(legend continued on next page)

(F) A schematic representation of the dAsCas12f plasmid construct conjugated with VP64 or VPR at different sites. The plasmid and sgRNA targeting *HEXA* were transduced into Huh-7 cells stably expressing luciferase under the control of the minimal CMV promoter with *HEXA* sgRNA-binding sites.
(G) Increase in luciferase activities in cells transduced with the plasmid vector (n = 3, mean ± SD).
(H) A schematic representation of the AAV6 vector for the transcriptional activation in Huh-7 cells. NLS, SV40; pA, poly A sequence.
(I) A schematic representation of the AAV8 vector for the transcriptional activation in mice. NLS, SV40; pA, poly A sequence.