nature communications

Article

https://doi.org/10.1038/s41467-025-60124-6

hpCasMINI: An engineered hypercompact CRISPR-Cas12f system with boosted gene editing activity

Received: 24 June 2024

Accepted: 13 May 2025

Published online: 29 May 2025

Check for updates

Shufeng Ma[1,2,5], Kaitong Liao[1,5], Kechen Chen[1], Tong Cheng[3,4], Xiaofeng Yang[1], Peihan Chen[1], Sijie Li[1], Mengrao Li[1], Xin Zhang[1,2], Yanqun Zhang[1], Tao Huang[1], Xiaobo Wang[1], Lanfeng Wang[3,4], Ying Lin[1] & Zhili Rong[1]

Compact CRISPR-Cas systems have demonstrated potential for effective packaging into adeno-associated viruses (AAVs) for use in gene therapy. However, their applications are currently limited due to modest gene-editing activity. Here we introduce an engineered compact CRISPR-Cas12f (hpCasMINI, 554 aa), with hyper editing efficiency in mammalian cells via adding an α-helix structure to the N-terminus of an Un1Cas12f1 variant CasMINI (529 aa). The hpCasMINI system boosts gene activation and DNA cleavage activity with about 1.4-3.0-fold and 1.1-19.5-fold, respectively, and maintains the high specificity when compared to CasMINI. In addition, the system can activate luciferase reporter gene and endogenous *Fgf21* gene in adult mouse liver, as well as construct liver tumorigenesis model via disrupting *Trp53* and *Pten* genes and inserting oncogenic $Kras^{G12D}$ into the *Trp53* locus. When compared to *Sp*Cas9 and *Lb*Cas12a, hpCasMINI displays higher gene activation and exhibits higher DNA cleavage specificity, although with lower activity, at the tested sites. Moreover, with a similar strategy, we engineer compact versions of hp*Os*Cas12f1 (458 aa) from en*Os*Cas12f1 and hp*As*Cas12f1 (447 aa) from *As*Cas12f1-HKRA, both of which display increased DNA cleavage activity, with hp*As*Cas12f1 also showing improved gene activation capability. Therefore, we develop activity-increased miniature hpCasMINI, hp*Os*Cas12f1 and hp*As*-Cas12f1 nucleases, which hold great potential for gene therapy in the future.

The CRISPR-Cas system, which serves as an RNA-guided defense mechanism in bacteria and archaea, has garnered extensive applications in basic research and gene therapy[1,2]. Nonetheless, its substantial size, typically over 1000 amino acids, surpasses the packaging threshold of viral vectors, thereby limiting its effective delivery into target cells[3].

Recently, several compact CRISPR or CRISPR-like nucleases are discovered, including CasMINI (a protein engineered from Un1Cas12f1)[4,5], *Syntrophomonas palmitatica* Cas12f1 (*Spa*Cas12f1)[6], enhanced *Oscillibacter sp*. Cas12f1 (en*Os*Cas12f1, a protein engineered from *Os*Cas12f1)[7], *Acidibacillus sulfuroxidans* Cas12f1 (*As*Cas12f1)[8,9],

[1]Cancer Research Institute, School of Basic Medical Sciences, State Key Laboratory of Multi-organ Injury Prevention and Treatment, Guangdong Province Key Laboratory of Immune Regulation and Immunotherapy, Southern Medical University, Guangzhou, China. [2]State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, National Center for Respiratory Medicine, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China. [3]The Center for Microbes, Development and Health, CAS Key Laboratory of Molecular Virology & Immunology, Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China. [4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. [5]These authors contributed equally: Shufeng Ma, Kaitong Liao. e-mail: rongzhili@smu.edu.cn

enhanced *Ruminiclostridium herbifermentans* Cas12f1 (en*Rh*Cas12f1, a variant from *Rh*Cas12f1)[7], ISDra2 TnpB from *Deinococcus radiodurans* (an evolutionary ancestor of Cas12)[10–13] and OgeuIscB from the human gut metagenome (an ancestor of Cas9)[14,15], consisting of 529, 497, 433, 422, 415, 408, and 350 amino acid residues, respectively. These nucleases are used as gene editing tools in eukaryotic cells, and some of them are demonstrated to be efficiently delivered in vivo with single adeno-associated virus (AAV) and enable gene editing[11,12,16], including CasMINI and en*Os*Cas12f1. However, the editing efficiency of these miniature nucleases is much lower than that of the widely used large *Sp*Cas9, *As*Cas12a, and *Lb*Cas12a.

There are several strategies to enhance Cas nuclease activity, including protein engineering and gRNA modification. For example, amino acid mutation into positively charged arginine within the interface of Cas protein is able to enhance the interaction with the negatively charged target DNA and/or sgRNA[17,18], fusion of exo-endonuclease can trim the ends of the double-strand break cleaved by Cas nucleases[19,20], circularization of gRNA is capable of increasing the abundance of Cas/gRNA complexes[21], and deletion of the gRNA regions that do not directly interact with Cas nuclease or addition of a uridinylate-rich 3′-overhang on the crRNA can increase indel frequencies[5,22].

As most of the above-mentioned strategies have been employed to engineer Un1Cas12f1, we aim to develop a new one. The Un1Cas12f1 nuclease operates as an asymmetric homodimer, where each protomer of Cas12f1 adopts a unique conformation and carries out distinct functions in DNA recognition and cleavage[23,24]. Similar to Cas12a, Cas12b, and Cas12e, Un1Cas12f1 forms a bilobed architecture with NUC and REC lobes, that accommodates the guide RNA-target DNA heteroduplex in the central channel, demonstrating the conserved mechanisms among type V Cas12 nucleases (Supplementary Fig. 1)[24]. Since *Lb*Cas12a, *As*Cas12a, and *Fn*Cas12a display much higher nuclease activity than Un1Cas12f1[25], we try to identify the differential secondary α-helix or β-sheet structures between Cas12a and Un1Cas12f1, which might affect the nuclease activity. Fortunately, an α-helix structure in the very N-terminal of Cas12a is found to be lost in Un1Cas12f1 and the α-helix structure is essential for the nuclease activity of Cas12a (Supplementary Fig. 1 and 2).

In this study, by adding an α-helix structure in the N-terminus of Un1Cas12f1, we construct the hpCasMINI editor, which exhibits efficient gene activation and DNA cleavage in mammalian cells and mice. Moreover, we also successfully apply this strategy to two more compact *Os*Cas12f1 and *As*Cas12f1 to generate hp*Os*Cas12f1 and hp*As*-Cas12f1 with increased editing activity. These findings expand the toolbox of the miniature gene editors and might pave the way for the therapeutic application of Cas12f1.

Results

Adding of an α-helix to the N-terminus of dCasMINI-VPR improves gene activation

In order to develop a strategy to improve the gene editing efficiency of Un1Cas12f1, we compared the structures of type V CRISPR nucleases and noticed that the WED domain, which was reported to stabilize *Lb*Cas12a-crRNA-target DNA complex[26,27], lost an α-helix secondary structure in the N-terminus of Un1Cas12f1 (Fig. 1a and Supplementary Fig. 1). Subsequent analysis of seven subtypes of Cas12a (*Lb*, *As*, *Fn*, *Pe*, *Ee*, *Mb*, and *Li*Cas12a) revealed that four subtypes (*Lb*, *As*, *Fn*, and *Pe*Cas12a) had an α-helix structure at the N-terminus (Supplementary Fig. 2a, b). It was worth noting that, among Cas12a subtypes, *Lb*Cas12a, *As*Cas12a, and *Fn*Cas12a were most widely used for gene editing with high activity in eukaryotic cells, including plant cells and mammalian cells[28,29], and the three Cas12as exhibited higher editing efficiency than

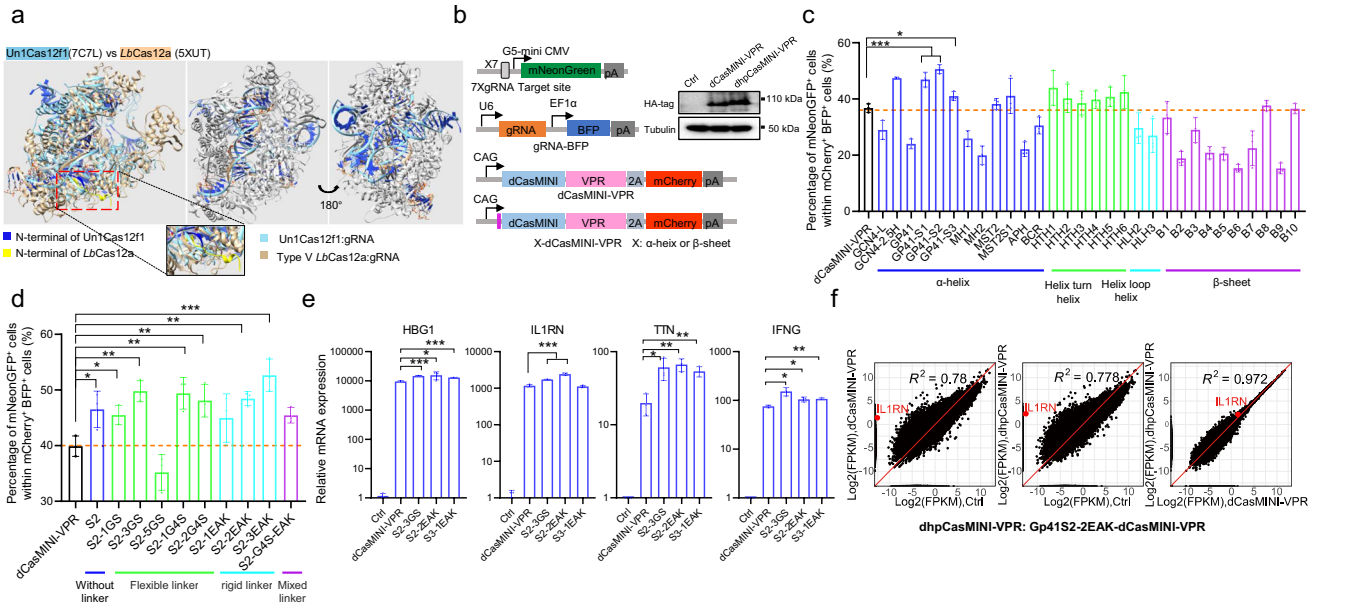


Fig. 1 | Adding of an α-helix to the N-terminus of dCasMINI-VPR improves gene activation efficiency. a Structure comparison between *Lb*Cas12a (orange, PDB: 5XUT) and Un1Cas12f1 (light blue, PDB: 7C7L). The inset showed the α-helix domain (highlighted by the red dashed box) in the N-terminal region. The middle and right panels showed the sgRNAs in the *Lb*Cas12a/sgRNA and Un1Cas12f1/gRNA complexes in cartoon presentation. **b** Left, schematic of the engineered dCasMINI-VPR variants with a different α-helix or β-sheet at each N-terminus of dCasMINI. The gRNA targeted the G5 promoter (7 repeated binding sites). Right, Western blot analysis confirmed the protein expression of CasMINI and hpCasMINI. *One experimental repetition was conducted due to the well-established design of the plasmid expression experiment*. **c** Gene activation with different dCasMINI-VPR variants in the G5-mNeonGreen reporter cells. **d** Gene activation with different linkers between Gp41S2 and dCasMINI-VPR. **e** RT–qPCR revealed relative mRNA expression of *HBG1*, *IL1RN*, *TTN*, and *IFNγ* in the human HEK293T cells transfected with the dCasMINI-VPR variants (S2-3GS, S2-2EAK and S3-1EAK) targeting the promoter region of each gene. **f** The specificity of gene activation with the hpCasMINI-VPR (S2-2EAK) system targeting *IL1RN* was revealed by RNA-seq in the HEK293T cells. For (**c**–**f**), mean values were presented with SD (n = 3 biological replicates). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, one-way ANOVA test versus dCasMINI-VPR. Ctrl, the cells transfected with the plasmid encoding the gRNA alone. *Source data are provided as a Source Data file.*

*Mb*Cas12a without an α-helix at the N-terminus (Supplementary Fig. 2b)[30]. Based on the above observations, we hypothesized that the α-helix at the N-terminus of Cas12a might affect its activity.

To test this hypothesis, we deleted the N-terminal α-helix (SKLEKF) or replaced it with variant peptide linkers, used these variants to construct different d*Lb*Cas12a-VPR activation systems to activate the endogenous gene *HBG1* in human HEK293T cells (Supplementary Fig. 2c). The results showed that the d*Lb*Cas12a-VPR variants lost gene activation ability, indicating that the α-helix at the N-terminal was crucial for the activity of *Lb*Cas12a (Supplementary Fig. 2c, d).

Next, we speculated that fusing an α-helix to the N-terminus of Un1Cas12f1 might enhance its editing efficiency. Since CasMINI was reported as an activity-enhanced Un1Cas12f1 variant[4], we performed engineering on CasMINI. To easily quantify the editing efficiency of the CasMINI variants, we first developed a G5-mNeonGreen gene activation HEK293T reporter cell line (Supplementary Fig. 3 a-c). When the plasmids encoding the gRNA targeting the G5 promoter (7 repeated targeting sites, also encoding BFP) and the VPR-fused dCasMINI (also encoding mCherry) were co-transfected into the reporter cells, the expression of mNeonGreen was activated and the percentage of mNeogreen-positive cells within mCherry-BFP-double positive cells measured the activation efficiency (Supplementary Fig. 3d). Then, we selected a suite of α-helical structures, including GCN4, Gp41, MH1, MST2, APH, BCR, HTH, HLH, and truncated versions of GCN4 and Gp41, as well as a series of β-sheet structures (Supplementary Fig. 4a). These peptide fragments were each fused to the N-terminus of CasMINI protein to explore their potential to enhance editing efficiency. Compared to dCasMINI-VPR, fusions with GCN4-2.5H and truncated Gp41 (Gp41S1-3) significantly increased the proportion of mNeonGreen-positive cells, from 36.7% to 47.4%, 46.9%, 50.6%, 41.0%, respectively (Fig. 1b, c). With the most substantial enhancement, Gp41S2-dCasMINI-VPR was selected for further investigation. Next, we optimized the length, flexibility (-GS-$_n$), and rigidity (-EAAAK-$_n$) of the linker between Gp41S2 and dCasMINI-VPR and found that Gp41S2 with most linkers performed better than dCasMINI-VPR in the mNeonGreen reporter cells (Fig. 1d and Supplementary Fig. 4b). Further, we tested the ability of two Gp41S2-dCasMINI-VPR variants with a flexible linker (S2-3GS) or a rigid linker (S2-2EAK) and a Gp41S3-dCasMINI-VPR variant (S3-1EAK, with more compact size, about 10 aa less) to activate endogenous *HBG1*, *IL1RN*, *TTN*, and *IFN-γ* genes in human HEK293T cells. The results showed that S2-2EAK performed better than S2-3GS and S3-1EAK to activate three out of the four tested genes, except for *IFN-γ* (Fig. 1e). Thus, the hyperactive S2-2EAK was termed dhpCasMINI-VPR and selected for subsequent experiments.

Finally, we performed RNA-seq in HEK293T cells co-transfected with the dCasMINI-VPR or dhpCasMINI-VPR targeting *IL1RN* for assessing the specificity of gene activation. As shown in Fig. 1f, both the dCasMINI-VPR and dhpCasMINI-VPR significantly and obviously activated *IL1RN* and only a few genes showed expression change, indicating high genome-wide targeted transcriptional specificity.

In conclusion, we successfully established a dhpCasMINI-VPR system with enhanced targeted gene activation via adding an α-helix to the N-terminus of CasMINI.

The hpCasMINI increases the DNA cleavage efficiency

To investigate the potential of hpCasMINI in genomic DNA cleavage, we transfected hpCasMINI into cells with gRNA targeting ten different genomic sites (Fig. 2a). The results indicated that hpCasMINI significantly improved editing efficiency at all the tested sites compared to CasMINI (Fig. 2b). At *SITE1*, the most notable enhancement in editing efficiency was observed, surging from 10.2% to 75.2%. This was followed by substantial increments at the *NLRC4*, *KLHL29*, *DMNT1*-g8, and *AAVS1*-g1 sites, where the efficiency elevated from 6.4%, 6.5%, 5.3%, and 5.2% to 34.1%, 34.7%, 23.7%, and 31.1%, respectively.

Next, we analyzed the indel pattern formed by hpCasMINI and CasMINI, finding that both hpCasMINI and CasMINI predominantly produced deletions (from 1-20 bps) (Fig. 2c). We also quantified the indel formation frequency downstream of the PAM region and found that significant editing occurred in the PAM-distal region outside the sgRNA binding sequence, with peak activity around positions 20-30 bp from the PAM, consistent with the previous report[31]. At the same time, the deletions produced by hpCasMINI were more numerous and of greater length compared to CasMINI (Fig. 2c, d). Hence, we demonstrated that the fusion of an α-helix to the N-terminus of CasMINI could indeed enhance the editing efficiency of CasMINI in mammalian cells.

The hpCasMINI exhibits high specificity for gene editing

We then evaluated the specificity of hpCasMINI-mediated gene editing in human cells using Tag-seq[32]. Within the 10 tested sites, hpCasMINI exhibited about 2.5-fold increased activity and similar specificity compared to CasMINI (Fig. 3a–c). We calculated the specificity index (the ratio of total on-target reads to the on-target reads plus the off-target reads) and counted total off-target sites within the 10 sites, found no significant difference between CasMINI and hpCasMINI (Fig. 3b). Furthermore, Tag-seq results confirmed that double-strand breaks (DSBs) induced by hpCasMINI occurred at positions around 12 and 25 downstream of the PAM (Fig. 3c), consistent with previous reports[31,33].

Next, we examined the mismatched tolerance of hpCasMINI in the editing capability with gRNA mismatches by tiling 1- or 2-nt mismatches and gRNA length. For the *SITE1* and *NLRC4* loci, both CasMINI and hpCasMINI exhibited intolerance to gRNA mismatches as the presence of even a single mismatch, let alone two, led to a significant reduction or complete loss of editing activity (Fig. 3d, e). Proximal mismatches displayed more impact than distal ones, and mismatches at positions 19 and 20 downstream of the PAM showed much less interference for editing ability (Fig. 3d, e). Shortening of the gRNA length also notably inhibited the editing efficiency of both hpCasMINI and CasMINI (Fig. 3f, g). Compared to a 20-nt gRNA, an 18-nt gRNA induced a significantly reduced indel formation efficiency, with a 16nt gRNA still capable of inducing minimal DSBs at the *SITE1* locus, while almost no DSBs was detected at the *NLRC4* site. In summary, the impact of gRNA mismatches and length alterations on hpCasMINI was similar to that on CasMINI.

dhpCasMINI-VPR enhances gene activation in vivo

Next, we assessed the in vivo applicability of the dCasMINI-VPR variants. In addition to the dhpCasMINI-VPR (S2-2EAK), S3-1EAK-dCasMINI-VPR (S3-1EAK) was also employed for the in vivo assays, as the latter contained a different α-helix structure and 11 less amino acid residues and showed comparable gene activation capability in cell culture. We delivered the plasmids encoding the systems (a gRNA targeting the TRE-promoter and dCasMINI-VPR, dhpCasMINI-VPR, or S3-1EAK) and a TRE-driven luciferase reporter into the livers of adult mice using hydrodynamic tail vein injection (Fig. 4a). Following injection, bioluminescence was measured at 12, 24, 48, and 72 hours, revealing a progressive decline in signal intensity over time (Fig. 4b). At the 12-hour and 24-hour time points, significantly higher luciferase activity was detected in mice receiving the dhpCasMINI-VPR and S3-1EAK systems compared to those receiving the dCasMINI-VPR system (Fig. 4b, c). Last, we tested whether the dhpCasMINI-VPR system could efficiently activate endogenous target genes in adult mice. Similar to the luciferase assay, the plasmids encoding a gRNA targeting the promoter region of *Fgf21* gene and dCasMINI-VPR, dhpCasMINI-VPR, or S3-1EAK were delivered to adult mouse liver via hydrodynamic tail vein injection. Quantitative RT-PCR assay at 24 h post-injection revealed that dhpCasMINI-VPR and S3-1EAK demonstrated a 3.9- and 4.7-fold increase in gene activation efficiency, compared to the dCasMINI-VPR control (Fig. 4d). Western Blot assay also confirmed that

https://doi.org/10.1038/s41467-025-60124-6

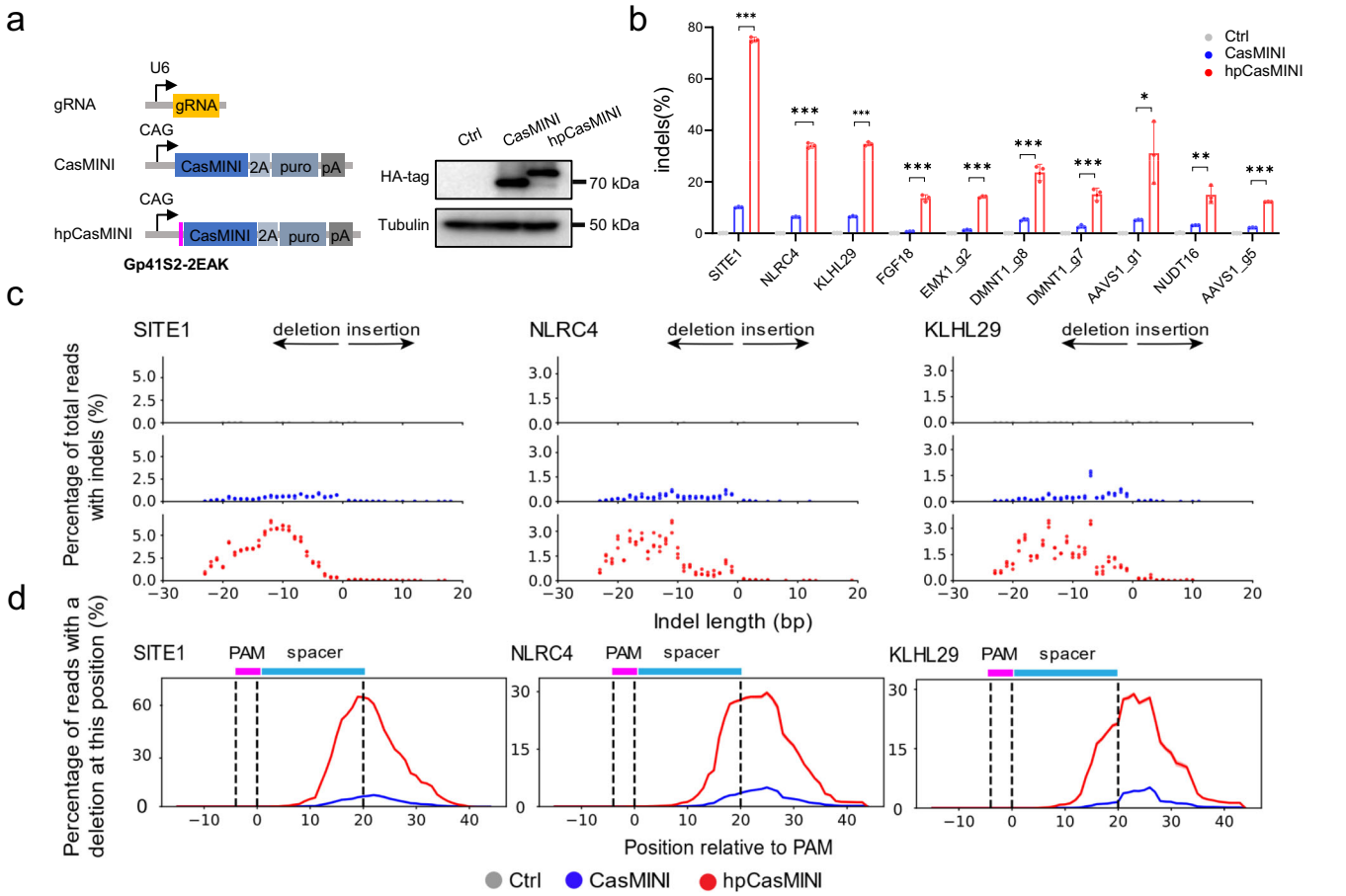


Fig. 2 | The hpCasMINI increases the DNA cleavage efficiency. a Left, Schematic of CasMINI and hpCasMINI systems for gene cleavage. Right, Western blot analysis confirmed the protein expression of CasMINI and hpCasMINI. *One experimental repetition was conducted due to the well-established design of the plasmid expression experiment.* **b** The editing activity at ten genomic target sites were revealed by Deep-seq in the human MCF7 Cells. **c** The indel patterns induced by CasMINI and hpCasMINI effectors at the *SITE1, NLRC4* and *KLHL29* sites. **d** The percentage of reads with deletions at each nucleotide position downstream of the PAM sequence showed the indel position and editing efficiency. For (**b**–**d**), Ctrl, the cells transfected with the plasmid encoding the gRNA alone. Mean values were presented with SD ($n = 3$ biological replicates), $^{***}p < 0.001$, one-way ANOVA test. *Source data are provided as a Source Data file.*

Fgf21 was successfully activated (Fig. 4e). In conclusion, the dhpCasMINI-VPR system was able to achieve better gene activation in vivo than dCasMINI-VPR.

Tumorigenesis modeling in mice with hpCasMINI

Subsequently, we assessed the in vivo genomic DNA cleavage capability of hpCasMINI. Since *KRAS*, *TP53*, and *PTEN* mutants are major drivers of intrahepatic cholangiocarcinoma (ICC)[34], our group have established a tumorigenesis model to examine the in vivo DNA cleavage capability of the CRISPR systems via simultaneously inserting $Kras^{G12D}$ into the *Rosa26* locus and disrupting *Trp53* and *Pten* in the mouse liver[21,35]. However, we failed to screen out a CasMINI gRNA targeting *Rosa26* loci with high efficiency. Thus, we modified our strategy by inserting $Kras^{G12D}$ into the *Trp53* locus instead of the *Rosa26* locus via a HITI method[36]. Similarly, the plasmids encoding a DNA donor carrying $KRAS^{G12D}$, the sgRNAs targeting the*Trp53* and *Pten* loci, and hpCasMINI or CasMINI were delivered into the mouse liver via hydrodynamic tail vein injection (Fig. 5a). After three months, the mice were sacrificed and their livers were analyzed. Morphologically, the livers of the control mice (injected with the plasmids encoding a DNA donor carrying $KRAS^{G12D}$ and the sgRNAs targeting the*Trp53* and *Pten* loci) appeared normal, with no tumor nodule present. In contrast, livers from the CasMINI- and hpCasMINI-injected mice developed an average of 1.8 and 5.3 tumor nodules, respectively (Fig. 5b, c). And the liver weights of mice injected with control, CasMINI, and hpCasMINI were about 1.1, 1.4, and 2.0 g, respectively (Fig. 5c). H&E and immunohistochemical staining for the cholangiocarcinoma marker CK19 showed slight hepatocyte swelling in the CasMINI group but preserved liver architecture; disorganized cellular arrangement and strong positive CK19 staining were observed in the hpCasMINI group (Fig. 5d). Genomic DNA analysis from the liver tissues revealed higher editing efficiency for *Trp53* and *Pten* in the hpCasMINI group compared to the CasMINI group (Fig. 5e, f). Both hpCasMINI and CasMINI groups achieved integration of the $KRAS^{G12D}$ DNA donor at the *Trp53* locus in the murine hepatocytes (Fig. 5g, h). All the above results that the hpCasMINI achieved a higher editing efficiency in vivo than CasMINI and it successfully established a murine model of cholangiocarcinoma.

Performance comparison of *Sp*Cas9, *Lb*Cas12a and hpCasMINI

We then compared the efficiency of gene activation and DNA cleavage between hpCasMINI and the widely used *Sp*Cas9 and *Lb*Cas12a. The gRNAs for hpCasMINI, *Sp*Cas9, and *Lb*Cas12a were designed at the same or adjacent location within the promoter regions of the endogenous *HBG1* and *IL1RN* genes (Fig. 6a). The results in human HEK293T and MCF7 cells showed that the gene activation efficiency of dhpCasMINI-VPR on these two sites was significantly higher than that of *Sp*Cas9 and *Lb*Cas12a (Fig. 6b). Next, we compared the DNA cleavage activity of *Sp*Cas9, *Lb*Cas12a, and hpCasMINI across four genomic sites at the same location and five sites at adjacent locations (Fig. 6c).

https://doi.org/10.1038/s41467-025-60124-6

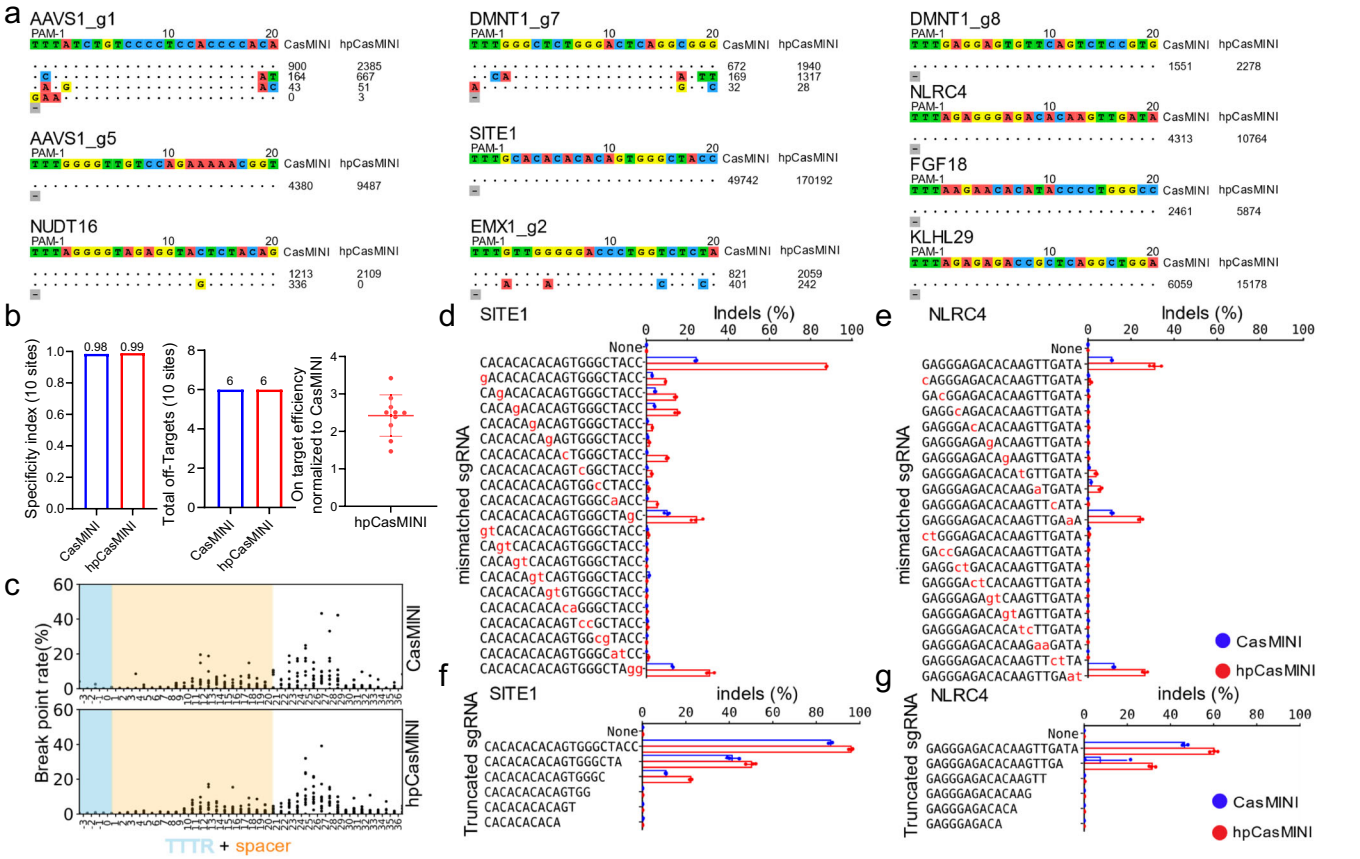


Fig. 3 | The high specificity of hpCasMINI-mediated gene editing. a Tag-seq-based comparative analyses of CasMINI and hpCasMINI in human MCF7 cells. **b** Tag-seq-based comparative analysis illustrated the Specificity index (value was calculated by the ratio of total on-target reads to the on-target reads plus the off-target reads within the twenty-one sites), the total amount of off-target sites, and editing efficiency between CasMINI and hpCasMINI for the tested ten genomic sites. **c** Mapping of double-strand breaks (DSBs) with Tag-seq induced by CasMINI and hpCasMINI. Each dot represented a DSB event, plotted against its genomic position relative to the TTTR + Spacer sequence. **d**, **e** Effects of mismatched sgRNA with 1 or 2 bp on activity of CasMINI and hpCasMINI at the *SITE1* and *NLRC4* loci. (**f**, **g**) The editing efficiency of CasMINI and hpCasMINI with gRNAs of different lengths at the *SITE1* and *NLRC4* sites. For (**d**, **g**), mean values were presented with SD ($n = 3$ biological replicates). *Source data are provided as a Source Data file.*

The editing efficiency of *Sp*Cas9 and *Lb*Cas12a at these nine sites ranged from 30% to 60%, and hpCasMINI ranged from 10% to 50% (Fig. 6d). It was worth noting that, compared to *Sp*Cas9, the editing efficiency of hpCasMINI was even higher at the *SITE1* site and comparable at the *NLRC4* and *FGF18* sites, and compared to *Lb*Cas12a, the editing efficiency of hpCasMINI was comparable at the *KLHL29*, *NUDT16*, and *SITE1* sites (Fig. 6d). For the remaining sites, the editing efficiency of hpCasMINI was lower than that of *Sp*Cas9 and *Lb*Cas12a (Fig. 6d). The indel patterns induced by *Sp*Cas9, *Lb*Cas12a, and hpCasMINI were examined. The results showed that the location of indels produced by *Sp*Cas9, *Lb*Cas12a, and hpCasMINI was within 5 bps upstream of NGG PAM or 10–30 bps downstream of TTTR PAM (Supplementary Fig. 5a). *Sp*Cas9 produced both insertion and deletion of small fragments, while the other two mainly produced deletion, ranging from 1 to 25 bps (Supplementary Fig. 5b). We also used Tag-seq to detect the specificity of these three effectors within the nine sites. *Sp*Cas9 showed the strongest off-target activity, with a Specificity Index value of 0.16 and 951 off-target sites, which was significantly higher than *Lb*Cas12a and hpCasMINI, with Specificity Index value of 0.89 and 0.98 and 23 and 13 off-target sites, respectively (Fig. 6e, f and Supplementary Fig. 6a). Tag-seq analysis further revealed the cleavage characteristics of *Sp*Cas9, *Lb*Cas12a, and hpCasMINI, which were consistent with previous reports (Supplementary Fig. 6b)[4,33]. In summary, compared to *Sp*Cas9 and *Lb*Cas12a, hpCasMINI displayed higher efficiency for gene activation and exhibited higher specificity and lower activity for genomic DNA cleavage at the tested sites.

Miniature hp*Os*Cas12f1 increases the DNA cleavage efficiency

Then, we speculated that the strategy to enhance the gene editing efficiency of Cas nucleases via adding an α-helix secondary structure to the N-terminus might be applied to other miniature Cas proteins. Therefore, we tested this hypothesis using en*Os*Cas12f1 (433 aa) and en*Rh*Cas12f1 (415 aa) with more compact size. As the structures of these two Cas12f nucleases have not been resolved by wet experiments, we utilized AlphaFold to predict the structures of these two proteins and found that the structure of en*Os*Cas12f1 was more similar to CasMINI and that neither N-terminus of these two nucleases contained an α-helix (Fig. 7a and Supplementary Fig. 7a, b). Next, we fused the Gp41S2-2EAK α-helix to each N-terminus of these two nucleases and found that this strategy could enhance the editing efficiency of en*Os*Cas12f1, not en*Rh*Cas12f1 (Fig. 7). Gp41S2-2EAK-en*Os*Cas12f1 (hp*Os*Cas121 thereafter) outperformed en*Os*Cas12f1 to induce indels across seven sites in human MCF7 cells with 1.2-2.5-fold increase (Fig. 7b, c). We also analyzed the type, length, and position of indels (Supplementary Fig. 8). Similar to previous observations with CasMINI and hpCasMINI, hp*Os*Cas12f1 generated deletions at a higher frequency than en*Os*Cas12f1 at each position (Fig. 2d and Supplementary Fig. 8a). The result showed that both hp*Os*Cas12f1 and en*Os*Cas12f1 primarily induced deletions from 1–25 bps in length, with peaks at around 10 bps (Supplementary Fig. 8b). Tag-seq analysis of the editing specificity revealed that DNA cleavage occurred at around the 14th and 24th positions downstream of the YTTH PAM (Fig. 7d), with both variants exhibiting comparable specificity with a similar Specificity

https://doi.org/10.1038/s41467-025-60124-6

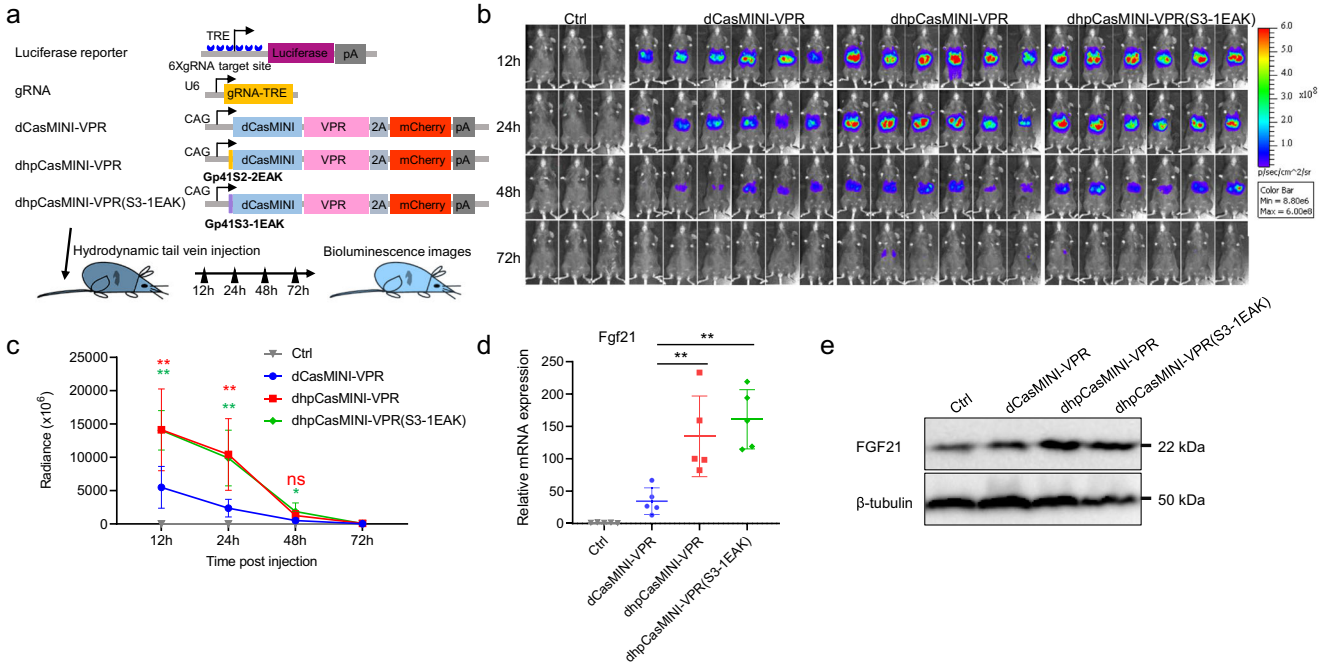
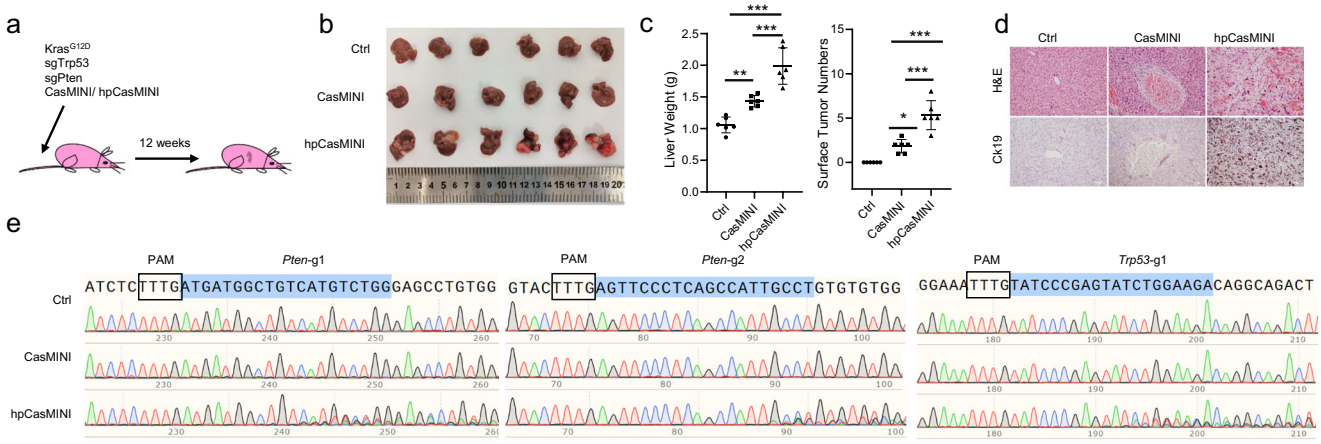


Fig. 4 | dhpCasMINI-VPR enhances gene activation in vivo. a Schematic of dhpCasMINI-VPR to activate luciferase reporter in mouse liver. **b** Bioluminescence imaging of mice at different time points after plasmid injection (12, 24, 48, and 72 h). **c** Quantitative analysis of bioluminescence intensity at various time points post-injection. **d** Endogenous *Fgf21* activation in vivo at 24 h post-injection. **e** Western blot analysis of Fgf21 expression. For (**b**–**e**), Ctrl, the mouse injected with the plasmids encoding the gRNA and Luciferase reporter (**b**, **c**) or the gRNA alone (**d**, **e**), $n = 3$ mice for the Ctrl group and $n = 6$ mice per other group (**c**, **d**), and $n = 5$ mice (**d**). Mean values were presented with SD, One-way ANOVA test versus dCasMINI-VPR, $^{*}p < 0.05$, $^{**}p < 0.01$. *Source data are provided as a Source Data file.*



TIDE	*Pten*-g1	*Pten*-g2	*Trp53*-g1
CasMINI	7.9%	4%	2.9%
hpCasMINI	30.3%	20.3	17.3%

Fig. 5 | Tumorigenesis modeling in mice with hpCasMINI. a Experimental design for liver cancer modeling in vivo. **b** Image analysis of mouse liver harvested 12 weeks post-injection. **c** Quantitative analysis of liver weights and surface tumor nodules, $n = 6$ mice per group. **d** H&E and IHC staining of Ck19 in mouse liver tumors. **e**, **f** Assessment of genome editing of *Pten* and *Trp53* loci by Sanger sequencing and tracking of indels by decomposition (TIDE). **g** Genomic PCR confirmed targeted integration of $Kras^{G12D}$ at the *Trp53* site. Primers for the 5′-KI, 3′-KI and WT were P53-TIDE-g2-F/KI-5-R, KI-3-F/P53-TIDE-g2-R and P53-g2-F/ P53-g2-R, respectively. M, Marker; WT, wild-type. Ctrl, mouse injected with the plasmids encoding the gRNA and the $Kras^{G12D}$ donor. **h** Sequencing validation of the $Kras^{G12D}$ mutation. For c, one-way ANOVA test, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. For (**d**, **g**), $n = 6$ mice per group, representative images are shown. *Source data are provided as a Source Data file.*

https://doi.org/10.1038/s41467-025-60124-6

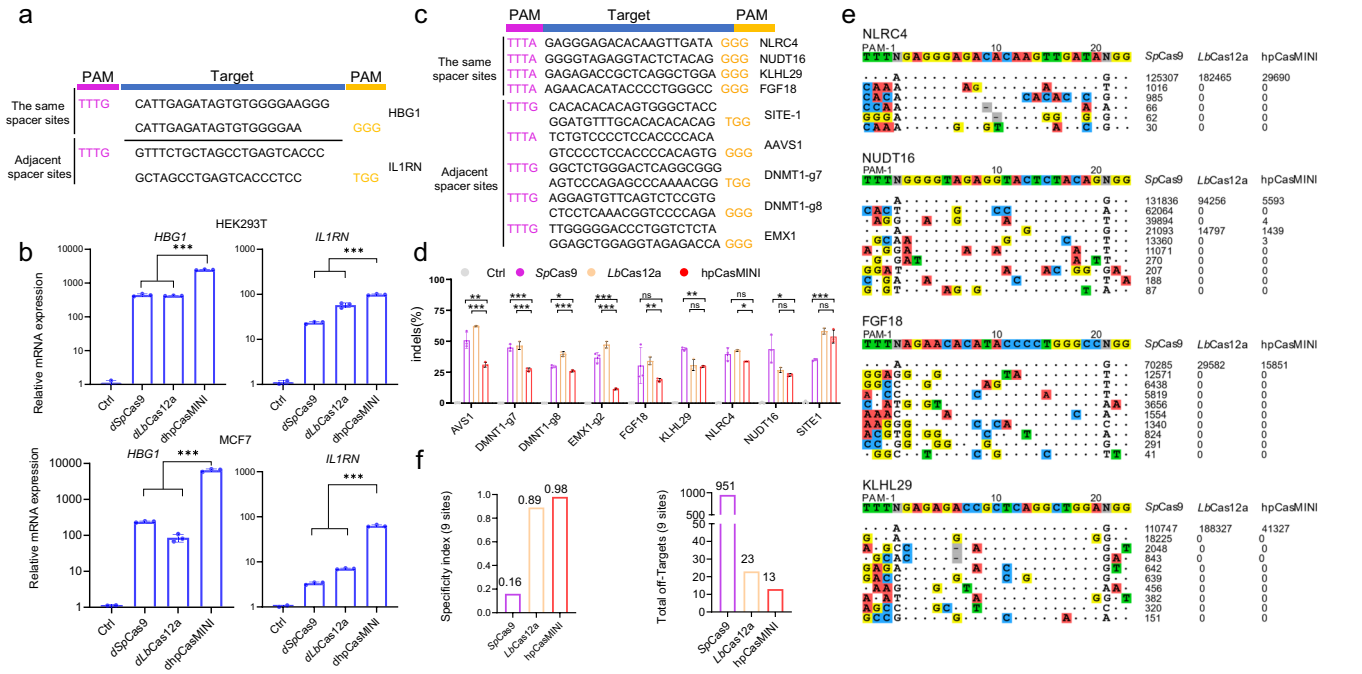


Fig. 6 | Performance comparison of *Sp*Cas9, *Lb*Cas12a and hpCasMINI. a The sequence information of the target sites for gene activation. d*Lb*Cas12a-VPR and dhpCasMINI-VPR shared a common target sequence with TTTR PAM (pink). d*Sp*Cas9-VPR recognized a target sequence with NGG PAM (orange). **b** Relative mRNA expression of *HBG1* and *IL1RN* in human HEK293T and MCF7 cells activated by different effectors. **c** The sequence information of the target sites for DNA cleavage. **d** Deep-seq quantified the indels induced by *Sp*Cas9, *Lb*Cas12a, and hpCasMINI in MCF7 cells. **e** Tag-seq revealed the specificity of different Cas nucleases in MCF7 cells at the genomic *NLRC4*, *NUDT16*, *FGF18*, and *KLHL29* sites. **f** Summary of Specificity index and total amount of off-target sites across the examined nine genomic sites. Ctrl, the cells without transfection. For (**b**, **d**), mean values were presented with SD, $n = 3$ biological triplicates. One-way ANOVA test, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *Source data are provided as a Source Data file.*

Index value of 0.50 and around 35 off-target sites across the tested seven sites (Fig. 7e, f). Overall, compared to en*Os*Cas12f1, hp*Os*Cas12f1 effectively increased the editing efficiency without compromising specificity.

Miniature dhp*As*Cas12f1-VPR improves gene activation efficiency

Lastly, we explored enhancing the efficiency of another miniature *As*Cas12f1 protein (422aa) with the similar strategy. Structural comparisons between Un1Cas12f1 and *As*Cas12f1 revealed resemblances to CasMINI, with both lacking an α-helix in their N-termini (Fig. 8a and Supplementary Fig. 9). To validate our hypothesis, we initially developed another G3-mNeonGreen gene activation HEK293T reporter cell line (Supplementary Fig. 3 a-d), with a distinct target site from the G5-mNeonGreen HEK293T reporter cell line. Next, we fused the Gp41S2-2EAK α-helix to the N-terminus of d*As*Cas12f1_HKRA-VPR (d*As*Cas12f1-VPR)[37] to create Gp41S2-2EAK-d*As*Cas12f1-VPR (dhp*As*Cas12f1-VPR thereafter), and found that this strategy could enhance the gene activation efficiency in both G3 and G5 reporter cell lines. dhp*As*Cas12f1-VPR significantly increased the proportion of mNeonGreen-positive cells by co-transfecting with three different gRNAs targeting G3-, G5-promoter, from 2.5%, 14.0%, and 19.9% to 28.2%, 40.8%, and 39.0%, respectively (Fig. 8b–d). Further, we tested the ability of dhp*As*Cas12f1-VPR to activate endogenous *HBG1*, *IL1RN* genes with 5 gRNAs targeting diverse sites of each promoter in human HEK293T cells. The activation efficiency of dhp*As*Cas12f-VPR was 6.7-fold higher in *HBG1* and 5.3-fold higher in *IL1RN* than that of d*As*CAs12f- VPR for the best-performed gRNA for each gene (Fig. 8e and Supplementary Fig. 10). Similar results were found in the MCF7 cell line (Supplementary Fig. 10). Finally, we performed RNA-seq analysis on HEK293T cells transfected with d*As*Cas12f-VPR or dhp*As*Cas12f-VPR targeting *IL1RN* to assess the specificity of gene activation. Both activated *IL1RN* specifically, with minimal off-target effects (Fig. 8f). Overall, dhp*As*Cas12f-VPR enhanced activation efficiency while maintaining specificity over d*As*Cas12f-VPR.

Miniature hp*As*Cas12f1 improves DNA cleavage efficiency

We also constructed Gp41S2-2EAK-*As*Cas12f1 (hp*As*Cas12f1) to achieve efficient gene cleavage. To investigate the potential of hp*As*Cas12f in genomic DNA cleavage, we transfected hp*As*Cas12f into cells with gRNA targeting 11 different genomic sites (Fig. 9a, b). The results indicated that hp*As*Cas12f substantially enhanced editing efficiency at 9 of the sites tested, with the exception of *PDCD1* and *APOB* (Fig. 9b). At *NLRC4*, the most notable enhancement in editing efficiency was observed, surging from 35.7% to 62.4%. This was followed by substantial increments at the *IL1RN*-g2, *HEXA*, *HBG* and *PSCK9* sites, where the efficiency elevated from 48.4%, 62.5%, 47.6%, and 54.1% to 72.2%, 81.0%, 65.2%, and 70.8%, respectively. We also analyzed the type, length, and position of indels (Fig. 9c, d and Supplementary Fig. 11). Notably, both hp*As*Cas12f1 and *As*Cas12f1 primarily induced deletions from 1–30 bps in length, with peaks at around 10 bps (Fig. 9c and Supplementary Fig. 11a). In line with CasMINI and hpCasMINI, hp*As*Cas12f1 generated deletions at a higher frequency than *As*Cas12f1 at each position (Fig. 9d and Supplementary Fig. 11b). Tag-seq analysis indicated DNA cleavage peaks near the 12th and 23rd bases downstream of the NTTR PAM (Fig. 9e), showing a Specificity Index of 0.82 for *As*Cas12f1 and 0.83 for hp*As*Cas12f1, and 51 off-target sites for *As*Cas12f1 and 58 for hp*As*Cas12f1 across 11 sites (Fig. 9f, g). In conclusion, hp*As*Cas12f1 increased editing efficiency without obvious compromising the specificity of *As*Cas12f1.

Discussion

In this study, we engineered a compact and efficient CRISPR-Cas effector, termed hpCasMINI. We added an α-helix secondary structure to the N-terminus of Cas12f1 and enhanced the editing efficiency. This

https://doi.org/10.1038/s41467-025-60124-6

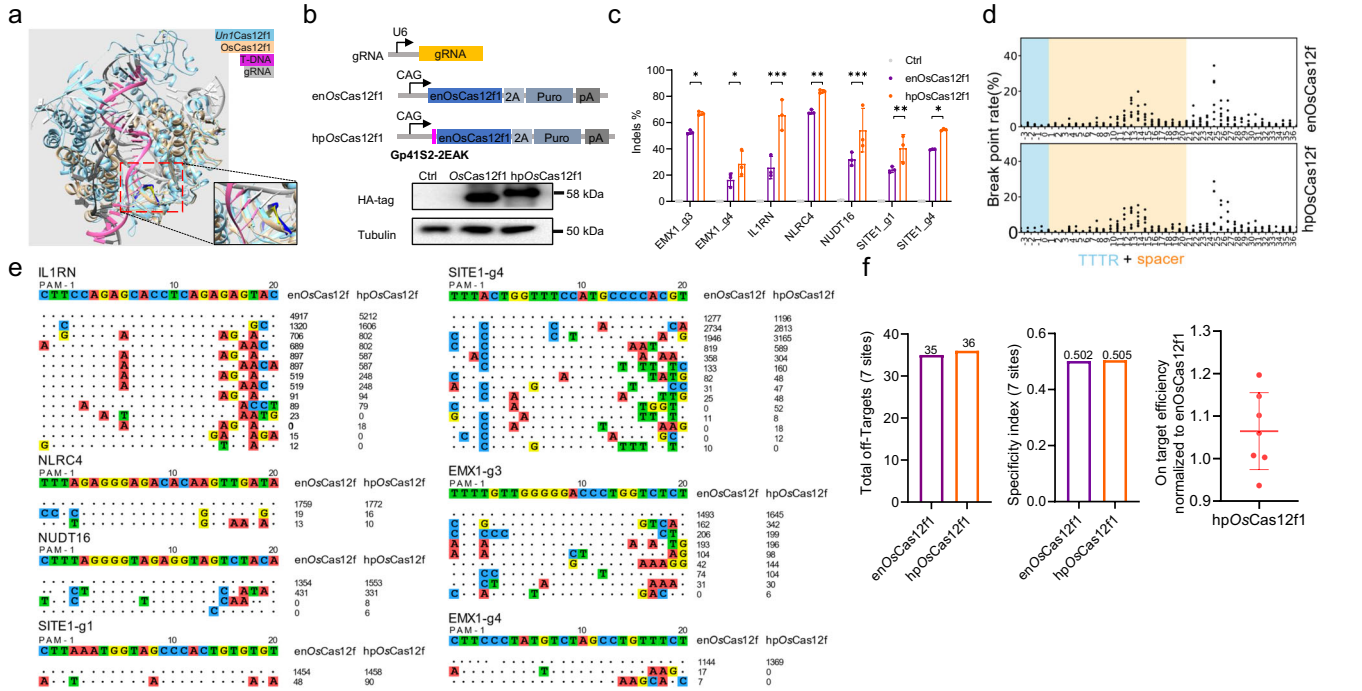


Fig. 7 | Miniature hp*Os*Cas12f1 increases the DNA cleavage efficiency. **a** Structural comparison of the Un1Cas12f1 (light blue) with en*Os*Cas12f1 (orange) bound to target DNA (pink), non-target DNA and gRNA (gray), with an inset showing the N-terminal structure. **b** Top, schematic of the hp*Os*Cas12f1 system for DNA cleavage. Bottom, Western blot analysis confirmed the protein expression of en*Os*Cas12f1 and hp*Os*Cas12f1. *One experimental repetition was conducted due to the well-established design of the plasmid expression experiment*. **c** Deep-seq quantified the indels induced by en*Os*Cas12f1 and hp*Os*Cas12f1 in MCF7 cells. **d** The distribution of break points induced by en*Os*Cas12f1 and hp*Os*Cas12f1 across the tested sites. **e** Tag-seq revealed the specificity of en*Os*Cas12f1 and hp*Os*Cas12f1 in MCF7 cells at the indicated genomic sites. **f** The total amount of off-target sites, Specificity index, and relative efficiency of en*Os*Cas12f1 and hp*Os*Cas12f1. Ctrl, the cells transfected with the plasmids encoding the gRNAs alone. For (**c**–**f**), mean values were presented with SD, $n = 3$ biological triplicates. One-way ANOVA test, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *Source data are provided as a Source Data file.*

provides a method to develop compact and efficient CRISPR-Cas effectors that can be valuable for basic biological research and therapeutic applications.

A structure-based optimization strategy for the CRIPSR system is proposed in this study, which we term "Spatial-structure optimization". The conventional structure-based optimization methods generally involve amino acid mutation at particular residues to change the interaction between Cas protein and gRNA or target DNA[17,18]. Here, through comparing the secondary α-helix and β-sheet structures between high-activity and low-activity Cas nucleases with similar overall structures, we can select out differential secondary structures and test their impact on gene editing activity. With luck, we have identified a differential α-helix at the N-terminus of Cas12a (with high activity), which is missing in Cas12f (with low activity). Disrupting the α-helix abolishes d*Lb*Cas12a-VPR-mediated gene activation (Supplementary Fig. 2c), demonstrating the necessity of this α-helix for *Lb*Cas12a activity. On the other hand, adding an α-helix to the N-terminus of Cas12f enhances both gene activation and DNA cleavage (Figs. 1–9), demonstrating the sufficiency of this α-helix for Cas12f activity. Since the N-terminal α-helix is a component of the WED domain and this domain is crucial to stabilize *Lb*Cas12a-crRNA-target DNA complex[26,27] and thus critical for *Lb*Cas12a activity, adding an exogenous α-helix back into Cas12f might stabilize Cas12f-crRNA-target DNA complex as well and thus enhance editing activity. The "Spatial-structure optimization" is a promising approach for the optimization of the CRISPR/Cas systems (both CasMINI, *Os*Cas12f and *As*Cas12f benefit from this strategy) and thus deserves more and further exploration, such as used to optimize the hypercompact and inefficient IscB and TnpB nucleases in the future.

We have developed the hpCasMINI effector, a compact and efficient tool suitable for in vivo gene editing. In general, delivery is a bottleneck to hinder fulfilling the promise of therapeutic gene editing[38]. And the large size of the most widely used Cas9 and Cas12a (both about 3.6 kb) is one of the major factors that limit delivery[3]. Although several compact nucleases have been efficiently delivered by AAV vectors and enable in vivo gene editing, the editing efficiency remains much lower than that of Cas9 and Cas12a[7,11,12,16]. The AAV vectors, a widely used and relatively safe in-body delivery tool, have a packaging capacity of about 4.7-5.0 kb and are unable to efficiently deliver the large Cas9 and Cas12a. While hpCasMINI is less than half the size of Cas9 and Cas12a, and thus provides extra space for functional modules, such as VP64 and VPR transcription activators for gene activation, AID and Tad deaminases for base editing, reverse transcriptase domains for prime editing, to realize intended gene editing in vivo. We have demonstrated that hpCasMINI and dhpCasMINI-VPR are more efficient to disrupt genes and activate gene expression than the corresponding CasMINI systems in the liver of adult mice, respectively (Figs. 4, 5), suggesting that they hold great promise for in vivo gene therapy. However, whether hpCasMINI performs better than CasMINI in the context of base editing or prime editing needs further investigation.

Compared to the widely used *Sp*Cas9 and *Lb*Cas12a, hpCasMINI displays a better gene activation ability at the tested *HBG1* and *IL1RN* sites (Fig. 6a, b). It is noteworthy that though hpCasMINI is more potent for gene activation, this phenomenon could reflect a suboptimal gRNA for Cas9 and Cas12a. The reason is that, based on our experience, the binding location in the promoter region of the gRNA for a given Cas protein plays an essential role for the gene activation extent, and for different Cas proteins, the optimal binding locations are different[39–41]. In this study, the binding locations could be better for hpCasMINI than for *Sp*Cas9 and *Lb*Cas12a. The gene activation extent induced by hpCasMINI at the tested sites (hundreds to thousands) is

https://doi.org/10.1038/s41467-025-60124-6

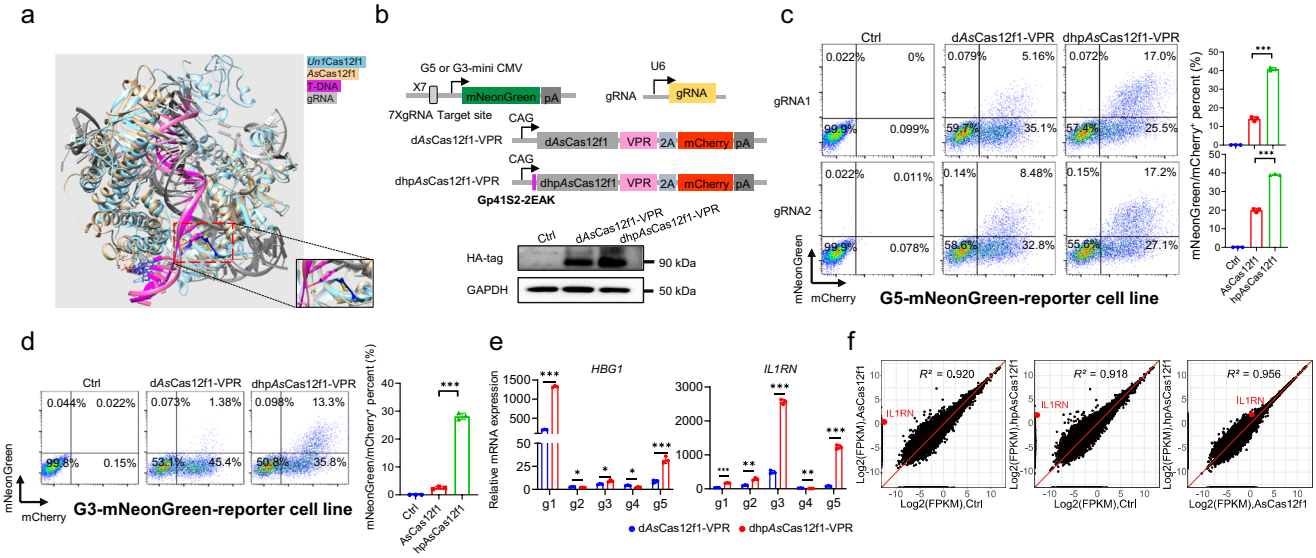


Fig. 8 | Miniature dhp*As*Cas12f1-VPR improves gene activation efficiency. **a** Structural comparison of the Un1Cas12f1 (light blue) with *As*Cas12f1 (orange) bound to target DNA (pink), non-target DNA and gRNA (gray), with an inset showing the N-terminal structure. **b** Top, schematic of the dhp*As*Cas12f1-VPR system for gene activation. The gRNA targeted the G3 or G5-mini CMV promoter (7 repeated binding sites). Bottom, Western blot analysis confirmed the protein expression of d*As*Cas12f1-VPR and dhp*As*Cas12f1-VPR. *One experimental repetition was conducted due to the well-established design of the plasmid expression experiment*. **c**, **d** Gene activation with d*As*Cas12f1-VPR variants in the G5-mNeonGreen (**c**) and G3-mNeonGreen (**d**) reporter cells. Left, the representative results of FACS analysis of the frequencies of mNeonGreen$^+$ cell populations. Right, quantification of the percentage of mNeonGreen$^+$ cells within mCherry$^+$ cells. **e** RT-qPCR revealed the relative mRNA expression of *HBG1* and *IL1RN* transfected with different gRNAs targeting the promoter region of *HBG1* and *IL1RN* in the human HEK293T cells. **f** The specificity of gene activation with the dhp*As*Cas12f-VPR system targeting *IL1RN* was revealed by RNA-seq in the HEK293T cells. Ctrl, the cells transfected with the plasmid encoding the gRNA alone. For (**c**–**f**), mean values were presented with SD, $n = 3$ biological triplicates. One-way ANOVA test, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *Source data are provided as a Source Data file.*

also comparable to that of *Sp*Cas9 and *Lb*Cas12a as reported[25,39–42]. In addition, as a broad promoter region can be used for screening out an optimal gRNA binding site, it is relatively easy to activate gene expression to a high extent using a given Cas protein. Therefore, in theory and based on our data, hpCasMINI can activate gene expression to a better or comparable extent with *Sp*Cas9 and *Lb*Cas12a at more other sites, and because of the much more compact size of hpCasMINI for easy delivery, it could be better than *Sp*Cas9 and *Lb*Cas12a for therapeutic gene activation in the future.

Compared to *Sp*Cas9 and *Lb*Cas12a, hpCasMINI exhibits a higher specificity across the nine tested sites for DNA cleavage (Fig. 6c–e and Supplementary Fig. 6). Compared to *Sp*Cas9, hpCasMINI displays a higher activity at one site and a comparable activity at two sites, and compared to *Lb*Cas12a, hpCasMINI displays a comparable activity at three sites (Fig. 6d). These observations suggest that, with higher specificity, more compact size, and comparable activity, hpCasMINI might perform as well as, even better than, *Sp*Cas9 and *Lb*Cas12a for a given interested genomic editing site. We also noticed that hpCasMINI exhibits a lower activity at two thirds tested sites, six out of nine, and therefore further improvement of activity via other strategies could broaden its applications.

AI can help to engineer the CRISPR systems. In this study, AlphaFold is used to predict the 3D structures of some Cas12a and Cas12f proteins to make it possible for the "Spatial-structure optimization". In particular, the 3D structures of the more compact en*Os*Cas12f1 (433 aa) and en*Rh*Cas12f1 (415 aa) have been predicted by AlphaFold and revealed that en*Os*Cas12f1 displays an overall similar structure to CasMINI and loses an α-helix at the very N-terminus (Fig. 7a and Supplementary Fig. 7a, b). However, the predicted 3D structure of en*Rh*Cas12f1 is more different from that of CasMINI. Consistently, hp*Os*Cas12f1, not hp*Rh*Cas12f1 demonstrates enhanced activity compared to the corresponding Cas12f proteins (Fig. 7).

In conclusion, we have constructed the hpCasMINI, hp*Os*Cas12f1 and hp*As*Cas12f1 systems via the "Spatial-structure optimization" strategy, which are compact and efficient tools for in vitro and in vivo gene editing.

Methods

Ethics statement

Our research complies with all relevant ethical regulations, and mice experiments were approved by the Scientific Investigation Board of Southern Medical University, Guangzhou, China.

Plasmid construction

● **Constructs for the DNA fragment.** DNA fragments (GCN4-L/GCN4-2.5H/Gp41/Gp41S1 ~ 3/MH1/MH2/MST1 /MST2/MSTS1/APH/BCR/HTH1 ~ 6/B1 ~ 10/Gp41S2-GS/3GS/5GS/G4S/2G4S/1EAK/2EAK/3EAK/G4S-EAK) were synthesized (RUIBO, China) as complementary paired ssDNA, then were annealed as dsDNA and was amplified, used for subsequent cloning.

pCAG-dCasMINI(dCasMINI-V4)-VPR-P2A-mCherry-PA[4], pCAG-CasMINI (CasMINI-V3.1)-P2A-puro-PA[4] and pCAG-en*Os*Cas12f1-P2A-puro-PA[7], pCAG-*As*Cas12f1 (HKRA)-P2A-puro-PA, pCAG-d*As*Cas12f1(HKRA)-VPR-P2A-mCherry-PA[37] were synthesized (Genewiz, China) and used for subsequent cloning.

● **Constructs for the X-dCasMINI-VPR/CasMINI fusion proteins.** pCAG-X-dCasMINI-VPR-P2A-mCherry-PA and pCAG-X-CasMINI-P2A-puro-PA: the DNA fragment X (GCN4-L/GCN4-2.5H/Gp41/Gp41-S1 ~ 3/MH1/MH2/MST1/MST2/MSTS1/APH/BCR/HTH1 ~ 6/B1 ~ 10) were inserted into the EcoRI and XmaI sites in the pCAG-dCasMINI-VPR-P2A-mCherry-PA or pCAG-CasMINI-P2A-puro-PA plasmid.

pCAG-Linker-Gp41S2/S3-dCasMINI-VPR-P2A-mCherry-PA: the DNA fragment Gp41S2/3-1GS/ 3GS/ 5GS/ G4S/ 2G4S/ 1EAK/ 2EAK/ 3EAK/ G4S-

https://doi.org/10.1038/s41467-025-60124-6

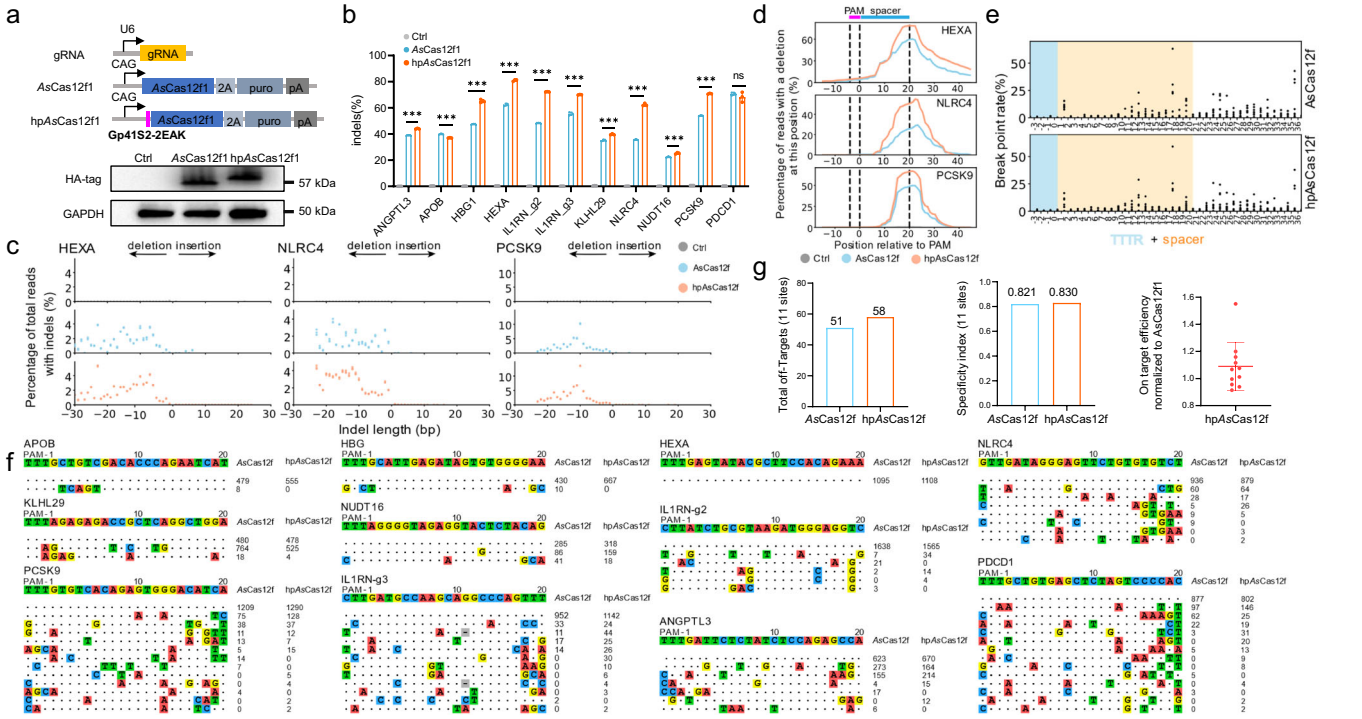


Fig. 9 | Miniature hp*As*Cas12f1 improves DNA cleavage efficiency. a Top, schematic of the hp*As*Cas12f1 system for DNA cleavage. Bottom, Western blot analysis confirmed the protein expression of *As*Cas12f1 and hp*As*Cas12f1.One experimental repetition was conducted due to the well-established design of the plasmid expression experiment. **b** Deep-seq quantified the indels induced by *As*Cas12f1 and hp*As*Cas12f1 in MCF7 cells. **c** The indel patterns induced by *As*Cas12f1 and hp*As*-Cas12f1 effectors at the HEXA, NLRC4 and PCSK9 sites. **d** The percentage of reads with deletions at each nucleotide position downstream of the PAM sequence showed the indel position and editing efficiency. **e** The distribution of break points induced by *As*Cas12f1 and hp*As*Cas12f1 across the tested sites. **f** Tag-seq revealed the specificity of *As*Cas12f1 and hp*As*Cas12f1 in MCF7 cells at the indicated genomic sites. **g** The total amount of off-target sites, Specificity index, and relative efficiency of *As*Cas12f1 and hp*As*Cas12f1. Ctrl, the cells transfected with the plasmids encoding the gRNAs alone. For b-g, mean values were presented with SD, $n = 3$ biological triplicates. One-way ANOVA test, $^{**}p < 0.01$, $^{***}p < 0.001$. *Source data are provided as a Source Data file.*

EAK were inserted into the EcoRI and XmaI sites in the CAG-dCasMINI-VPR-P2A-mCherry-PA.

● **Constructs for the SpCas9-puro/ LbCas12a-puro fusion proteins.** pCAG-*Sp*Cas9/*Lb*Cas12a-P2A-puro-PA: DNA fragment *Sp*Cas9 or *Lb*Cas12a was amplified from *Sp*Cas9 or *Lb*Cas12a template and inserted into EcoRI and BamHI sites in pCAG-CasMINI-P2A-puro-PA plasmid.

● **Constructs for G3- mNeonGreen reporter and G5-mNeonGreen reporter.** PbluSK-gACTB-IR-IRES-puro-G5/G3-minCMV-mNeonGreen-PA donor: A DNA fragment containing G5/G3-minCMV was synthesized (Genewiz, China) and inserted into the Bgl II and BsrGI sites in the PbluSK-gACTB-IR-IRES-puro-pCAG-d2mNeonGreen plasmid.

● **Constructs for the hpOsCas12f proteins.** pCAG-Gp41S2-2EAK-en*Os*Cas12f1-P2A-puro-PA: A DNA fragment containing en*Os*Cas12f1 was amplified from the pCAG-en*Os*Cas12f1-P2A-puro-PA plasmid and inserted into XmaI and AgeI sites in the pCAG-S2-2EAK-CasMINI-P2A-puro-PA plasmid.

● **Constructs for the hpAsCas12f proteins.** pCAG-Gp41S2-2EAK-*As*Cas12f1-P2A-puro-PA: A DNA fragment containing *As*Cas12f1 was amplified from the pCAG-*As*Cas12f1-P2A-puro-PA plasmid and inserted into XmaI and AgeI sites in the pCAG-S2-2EAK-CasMINI-P2A-puro-PA plasmid.

pCAG- Gp41S2-2EAK -d*As*Cas12f1(HKRA)-VPR-P2A-mCherry-PA: A DNA fragment containing d*As*Cas12f1 was amplified from the pCAG-d*As*Cas12f1(HKRA)-VPR-P2A-mCherry-PA plasmid and inserted into XmaI and NheI sites in the pCAG-dCasMINI(dCasMINI-V4)-VPR-P2A-mCherry-PA plasmid.

● **Constructs for the tumorigenesis modeling.** pBlue-gTP53-IR-T2A-Puro-pEF1α-KrasG12D-IR-gTP53 donor: DNA fragment containing IR-T2A-Puro-pEF1α-KrasG12D-IR was amplified from pBlue-Rosa26-IR-T2A-Puro-pEF1α-KrasG12D-IR-Rosa26 template and inserted into the XhoI and NotI sites in the pBluSKP plasmid. Primers were designed containing the sgTp53 sequence, sequence of amplifing DNA fragment, and overlap of the backbone plasmid.

●**Other plasmids in this article.** The plasmid pEF1α-d*Lb*Cas12a-VPR-mCherry, pCAG-d*Sp*Cas9-VPR-mCherry, pCAG-d*As*Cas12a-VPR-mCherry, pBluSKP-TRE-luciferase, and pBluSKP-Rosa26-IR-T2A-Puro-pEF1α-KrasG12D-IR-Rosa26 were described previously[21,41].

● **Single guide RNA (sgRNA) constructs.** sgRNA target sites were designed through https://www.benchling.com/. The basic gRNA plasmids, CasMINI gRNA (Cas12f1_ge4.0 and Cas12f1_ge4.1)[5], en*Os*Cas12f1 gRNA[7], and *As*Cas12f1 gRNA(Δ3-5 v7) were synthesized under the basic plasmid after the human U6 promoter (Genewiz of China). The Cas12f1_ge4.0 was used for gene activation and Cas12f1_ge4.1 were used for gene cleavage. The sgRNA expression plasmid was constructed by ligating the corresponding annealed oligos to the basic plasmid. The sequences of all sgRNAs are listed in Supplementary Table 1.

Cell culture and transfection

HEK293T cells (ATCC, CRL-3216 ™) were maintained in Dulbecco' s modified Eagle' s medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. MCF7 cells (ATCC, HTB-22) were cultured in RPMI-1640 medium supplemented with 10% FBS and 1% penicillin/streptomycin. All cells were cultured at 37 °C under 5% CO2.

Cells were transfected with polyethyleneimine (PEI) (Polysciences, 23966). The total amount of DNA was 500 ng per well for a 24-well plate, and plasmids encoding the Cas12f1 protein and sgRNA were transfected at a 1:49 ratio. Approximately 1×10^5 cells were plated in 24-well plates 1 day before transfection. For endogenous gene or G3/G5-mNeogreen reporter activation, the transfected cells were analyzed 2 days post-transfection. For gene cleavage, the transfected cells were selected for 2 days with media containing 0.8 and 2 μg/ml puromycin for HEK293T and MCF7 after 2 days post-transfection, respectively, and were analyzed in the 5 days post-transfection. For large-scale off-target cleavages detection of Tag-seq analysis, the total amount of DNA was 2 μg with 10 nM Tag per well for a 6-well plate, and plasmids encoding the Cas12f1 protein and sgRNA were transfected at a 1:49 ratio. Each gRNA were added equally.

RT-qPCR

RNA was isolated using Trizol Reagent according to the manufacturer's protocol. Briefly, the total RNA was extracted from cells by adding 500 μl TRIzol and 100 μl chloroform, after centrifugation at 13,400 g for 10 min at 4 °C, the supernatant was transferred to a 1.5 mL RNase-free tube. RNA was purified by precipitation with isopropanol and 75% ethanol. 500 ng RNA was reversetranscribed into cDNA with the HiScript IV RT SuperMix for qPCR kit (Vazyme, R423-01). RT–qPCR was performed by the LightCycler 96 System (Roche, Switzerland) using HiScript II One Step qRT-PCR SYBR Green Kit (Vazyme, Q221-01). The ΔΔCt method was applied to show relative gene expression after normalizing to glyceraldehyde phosphate dehydrogenase (GAPDH) expression. All qPCR primers are listed in Supplementary Table 2.

Mice

C57BL/6 mice (Guangdong Animal Center) aged 6 weeks were housed under controlled conditions: 12-hour light/dark cycles, ambient temperature maintained at 25 ± 2 °C, and relative humidity between 50–60%.

Gene activation in vivo

Female BALB/c mice aged 6 weeks were hydrodynamically injected with plasmids encoding the dCasMINI-VPR or dhpCasMINI-VPR, sgTET targeting TRE promoter and pBluSKP-TRE-luciferase reporter. A total of 50 μg of dhpCasMINI-VPR or dCasMINI-VPR, 10 μg of pBluSKP-TRE-luciferase and 49 μg of sgRNA plasmids were mixed with 1.5-1.8 ml of 0.9% sterile saline (10% of body weight) and then delivered to each mouse by hydrodynamic tail vein injection lasting 7 s. For bioluminescence detection, mice were injected intraperitoneally with 100 μl of 20 mg/ml D-Luciferin potassium salt (Beyotime, ST196) and then anesthetized with isoflurane. Bioluminescence images were captured using the IVIS Lumina II In Vivo Imaging System within 10 min of D-luciferin injection. Radiance values were calculated for the ROI using Living Image 3.1 software. For endogenous Fgf21 activation, male mice aged 8–10 weeks were hydrodynamically injected with plasmids encoding the dCas12fMINI-VPR or dhpCasMINI-VPR and sgFgf21 targeting the endogenous gene Fgf21. A total of 4 μg of hpCasMINI-VPR or hpCasMINI-VPR, and 196 μg of sgRNA plasmids were mixed with 1.8–2.0 ml of 0.9% sterile saline (10% of body weight) and then delivered to each mouse by hydrodynamic tail vein injection lasting 7 s. 24 h after injection, mice were euthanized and their liver were harvested.

Tumorigenesis modeling in mice

Female BALB/c mice aged 6 weeks were hydrodynamically injected with plasmids encoding the CasMINI or hpCasMINI, gRNA targeting the tumor suppressor genes (*Pten* and *Trp53*) and $Kras^{G12D}$ DNA donor. A total of 4 μg of CasMINI or hpCasMINI plasmids, 60 μg of $Kras^{G12D}$ DNA donor and 40 μg of sgRNA (sgTrp53-g2, sgPten-g1 and sgPten-g2) were mixed with 1.5-1.8 ml of 0.9% sterile saline (10% of body weight) and then delivered to each mouse by hydrodynamic tail vein injection lasting 7 s. After 12 weeks, the mice were euthanized for liver assessment. The maximal tumor size permitted by the ethics committee is a single tumor diameter less than 15 mm and a single tumor volume less than 2000 mm^3. These limits were not exceeded in our study.

Western blotting

Briefly, HEK293T cells transfected with sgRNA, Cas proteins or its variants were harvested at 48 h after transfection. Cells were lysed with RIPA buffer, boiled with 4X SDS sample loading buffer for 10 minutes at 95 °C, and then separated on an 8% or 10% sodium dodecyl sulfate-polyacrylamide gel. The proteins were transferred to a nitrocellulose membrane and detected by sequential blotting with primary antibodies against GAPDH (60004-1, proteintech) or β-tubulin (66240, proteintech), HA-tag (M180-3, MBL) and anti-mouse IgG HRP secondary antibody (7076S, Cell Signaling Technology). Protein expression was detected by chemiluminescence according to the manufacturer's instructions (WBKLS0500, Millipore, MA, USA).

For the in vivo endogenous Fgf21 activation, the liver tissues were cut into tiny pieces, then the lysate was added at a ratio of 100 μl per 10 mg of tissue. The liver tissue were ground until it is fully cracked and were centrifuged at 10000–14000 g for 5 minutes, were boiled with 4X SDS-sample loading buffer for 10 min at 95 °C and were separated on a 10% sodium dodecyl sulphate-polyacrylamide gel, transferred to a nitrocellulose membrane and detected with sequential blotting with primary antibodies against β-tubulin, Fgf21 (ab171941, abcam) and anti-mouse IgG HRP secondary antibody, anti-rabbit IgG HRP secondary antibody (7074S, Cell Signaling Technology). Protein expression was detected by chemiluminescence according to the manufacturer's instructions.

Immunofluorescence staining, H&E staining and immunohistochemistry

Livers were fixed overnight in 4% paraformaldehyde at 4 °C, embedded with paraffin, and then sliced into 5-μm sections. For H&E staining, the sections were rehydrated with gradient ethanol and stained with hematoxylin and eosin. For immunohistochemistry, the rehydrated sections were boiled for 15 min to retrieve antigen. After the endogenous peroxidase was blocked for 15 min, the sections were sequentially incubated with an anti-CK19 primary antibody (Abcam, ab133496) overnight at 4 °C, a secondary antibody (HRP-conjugated anti-rabbit IgG secondary antibody, 7074S, CST) for 1 h at room temperature, and the chromogenic substrate for 20 min. Finally, the sections were counterstained with hematoxylin, dehydrated, and sealed with neutral resins.

FACS analysis

In the G3/G5-mNeonGreen reporter assay, cells were harvested 2 days after transfection and resuspended in 400 μl FACS buffer (1× DPBS, 0.2% BSA, 2 mM EDTA), then loaded onto a flow cytometer (BD Fortessa, CA, USA) to detect BFP-, mCherry- and mNeonGreen- positive cells. The activation efficiency of X-dCasMINI-VPR or d*As*Cas12f1-VPR was calculated as the proportion of mNeonGreen-positive cells within the transfected cells (mCherry-BFP-double positive cells).

RNA-seq and bioinformatics analysis

RNA was isolated using Trizol Reagent according to the manufacturer's protocol. The constructed sequencing libraries were sequenced on the Illumina HiSeq Platform with 150 bp paired-end reads. The raw paired-end reads were preprocessed to remove the adapters with cutadapt (v1.15)[43]. Subsequently, the reads were mapped to the human genome (hg38) using HISAT2 aligner (v2.2.1)[44]. The uniquely mapped reads that passed the "markdup –r" filter by sambamba (v1.0.0)[45] were kept to quantify gene expression. The generation of gene counts

(reads aligned to each gene of each sample) was performed using gencode.v43.primary_assembly.basic.annotation GTF file (available at https://www.gencodegenes.org/human/) and FeatureCounts (v2.0.3)[46] with the following parameters: "-p --countReadPairs -B". Genes with no counts in any sample were filtered out. For each gene, Fragments per Kilobase per Million (FPKM) was calculated based on its length and the read count mapped to it. The Pearson correlation coefficient was calculated between the average FPKM of three replicates for each sample.

Tag-seq and bioinformatics analysis

The detailed procedures for Tag-seq library construction and analyses were described in our previous work[32]. The library was constructed with nested PCR and sequenced with Illumina HiSeq instrument with a 150-bp paired-end module. The bioinformatic analysis pipeline could be accessed at https://github.com/zhoujj2013/Tag-seq. The parameter in the "confg. docker.text" file was configured as follows: "MinSupportReadCount, 1; MinCuttingEventCount, 2; MaxMismatch, 6; MaxGap, 2; MaxGapMismatch, 4". The reference genome used in the analysis was the human genome (hg38). The specificity index was calculated as the ratio of on-target reads to the sum of on-target reads and off-target reads across all tested sites. On-target efficiency was calculated as the mean ratio of on-target reads across all tested sites, normalized to CasMINI. CRISPR/Cas-induced DSBs were quantified within a range spanning from 4 bp upstream of the gRNA target region to 16 bp downstream of the gRNA target region. The breaking rates were calculated as the ratio of the read count at each break site to the total read count within this gRNA quantification region. The primers are listed in Supplementary Table 2.

Deep-seq and bioinformatics analysis

The gDNA were purified using the TIANamp Genomic DNA Kit. Forward and reverse primers were used to amplify the target site of the first round of PCR, and then an equal amount of the first PCR product was mixed with primers containing P5 and p7 for a second round of PCR to generate the library. The amplicon deep sequencing libraries were sequenced in PE150 mode. Raw paired-end reads were subjected to sublibrary barcode removal (first 6 bp) using cutadapt (v1.15). Reads were then aligned to their respective reference sequences using Bowtie 2 (v2.5.1)[47] with the following parameters: "--no-unal -N 1 -L 10 --no-mixed --very-sensitive". Only one of the paired-end reads, positioned nearest to the gRNA target site, was retained for subsequent analysis. Editing events were quantified using CRISPResso2 (v2.2.12)[48], executed with the following parameters: "--exclude_bp_from_left 0 --exclude_bp_from_right 0 --quantification_window_center -12 –quantification_window_size 27 --ignore_substitutions".

Statistics & Reproducibility

The sample size was not predetermined using statistical methods. In cell line experiments, cells were randomly plated in 24-well or 6-well plates before plasmid transfection, with no programmed allocation of plasmids to individual cells. For animal studies, mice were randomly assigned to experimental groups. Researchers were not blinded to group allocations during both experimentation and outcome evaluation. Complete descriptions of statistical methods and data quantification are provided in the corresponding figure legends or Methods section. All collected data were included in the analyses without exclusion.

Reporting summary

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

Data availability

All sgRNAs and primers sequences are available in the Supplementary table. The raw and processed data of RNA-seq generated in this study have been deposited in the National Center for Biotechnology Information database under the accession numbers. GSE270586 The raw and processed data of Deep-seq are available in the National Center for Biotechnology Information database under accession numbers GSE270587. The raw and processed data of Tag-seq are available in the National Center for Biotechnology Information database under accession numbers GSE270588. Source data are provided in this paper.

Code availability

The code for the Tag-seq analysis pipeline used in this study can be accessed at https://github.com/zhoujj2013/Tag-seq and is also available through https://doi.org/10.5281/zenodo.4679460. Any future updates will be posted on both Zenodo and GitHub.

References

1. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
2. Chen, K., Wang, Y., Zhang, R., Zhang, H. & Gao, C. CRISPR/Cas Genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* **70**, 667–697 (2019).
3. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
4. Xu, X. et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* **81**, 4333–4345 (2021).
5. Kim, D. Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2021).
6. Wang, Y. et al. Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease. *Cell Rep.* **40**, 111418 (2022).
7. Kong, X. et al. Engineered CRISPR-OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. *Nat. Commun.* **14**, 2046 (2023).
8. Wu, Z. et al. Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.* **17**, 1132–1138 (2021).
9. Fan, P. et al. Targeted mutagenesis in mice via an engineered AsCas12f1 system. *Cell. Mol. Life Sci.* **81**, 63 (2024).
10. Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
11. Li, Z. et al. Engineering a transposon-associated TnpB-ωRNA system for efficient gene editing and phenotypic correction of a tyrosinaemia mouse model. *Nat. Commun.* **15**, 831 (2024).
12. Wang, M. et al. Hypercompact TnpB and truncated TnpB systems enable efficient genome editing in vitro and in vivo. *Cell Discov.* **10**, 31 (2024).
13. Xiang, G. et al. Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. *Nat. Biotechnol.* **42**, 745–757 (2024).
14. Han, D. et al. Development of miniature base editors using engineered IscB nickase. *Nat. Methods* **20**, 1029–1036 (2023).
15. Hirano, S. et al. Structure of the OMEGA nickase IsrB in complex with ωRNA and target DNA. *Nature* **610**, 575–581 (2022).
16. Guo, R. et al. In vivo treatment of tyrosinaemia with hypercompact Cas12f1. *Cell Discov.* **9**, 73 (2023).
17. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
18. Strecker, J. et al. Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).

19. Yin, J. et al. Cas9 exo-endonuclease eliminates chromosomal translocations during genome editing. *Nat. Commun.* **13**, 1204 (2022).
20. Lainšček, D. et al. Coiled-coil heterodimer-based recruitment of an exonuclease to CRISPR/Cas for enhanced gene editing. *Nat. Commun.* **13**, 3604 (2022).
21. Zhang, X. & Wang, X. Engineered circular guide RNAs boost CRISPR/Cas12a- and CRISPR/Cas13d-based DNA and RNA editing. *Genome Biol.* **24**, 145 (2023).
22. Bin Moon, S. et al. Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3′-overhang. *Nat. Commun.* **9**, 3651 (2018).
23. Xiao, R., Li, Z., Wang, S., Han, R. & Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res.* **49**, 4120–4128 (2021).
24. Takeda, S. N. et al. Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol. Cell* **81**, 558–570.e553 (2021).
25. Huang, H. et al. Comparison of DNA targeting CRISPR editors in human cells. *Cell Biosci.* **13**, 11 (2023).
26. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
27. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).
28. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
29. Tang, X. et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* **3**, 17018 (2017).
30. Tóth, E. et al. Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. *Nucleic Acids Res.* **46**, 10272–10285 (2018).
31. Karvelis, T. et al. PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res.* **48**, 5016–5023 (2020).
32. Huang, H. et al. Tag-seq: a convenient and scalable method for genome-wide specificity assessment of CRISPR/Cas nucleases. *Commun. Biol.* **4**, 830 (2021).
33. Xin, C. et al. Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. *Nat. Commun.* **13**, 5623 (2022).
34. Ong, C. K. et al. Exome sequencing of liver fluke-associated cholangiocarcinoma. *Nat. Genet.* **44**, 690–693 (2012).
35. Ma, S. et al. Enhancing site-specific DNA integration by a Cas9 nuclease fused with a DNA donor-binding domain. *Nucleic Acids Res.* **48**, 10590–10601 (2020).
36. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
37. Hino, T. et al. An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* **186**, 4920–4935 (2023).
38. van Haasteren, J., Li, J., Scheideler, O. J., Murthy, N. & Schaffer, D. V. The delivery challenge: fulfilling the promise of therapeutic genome editing. *Nat. Biotechnol.* **38**, 845–855 (2020).
39. Zhang, X. et al. Gene activation in human cells using CRISPR/Cpf1-p300 and CRISPR/Cpf1-SunTag systems. *Protein Cell* **9**, 380–383 (2018).
40. Zhang, X. et al. MiniCAFE, a CRISPR/Cas9-based compact and potent transcriptional activator, elicits gene expression in vivo. *Nucleic Acids Res.* **49**, 4171–4185 (2021).
41. Ma, S. et al. Phase-separated DropCRISPRa platform for efficient gene activation in mammalian cells and mice. *Nucleic Acids Res.* **51**, 5271–5284 (2023).
42. Liu, J. et al. CRISPR-assisted transcription activation by phase-separation proteins. *Protein Cell* **14**, 874–887 (2023).
43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 3 (2011).
44. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
45. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
46. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
47. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
48. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

Acknowledgements

We would like to thank Weiqi Lv for lab coordination and support. We also thank every member of Dr Zhili Rong's and Dr Ying Lin's lab for their helpful discussion and advice. This work was funded by the National Key R&D Program of China (2022YFA0806303 to Z.R.), the Major Project of Guangzhou National Laboratory (GZNL2023A03004 to Z.R.), the National Natural Science Foundation of China (82370078 to Z.R., 82370003 to Y.L., 82200072 to S.M, 82304008 to Z.X., and 82303995 to X.W.), the GuangDong Basic and Applied Basic Research Foundation (2023B1111050005 and 2024A1515012753 to Z.R., 2023A1515012269 to Y.L., and 2022A1515111046 to X.Z.), the grant of State Key Laboratory of Respiratory Disease (SKLRD-Z-202502 to S.M.).

Author contributions

Z.R., Y.L., and S.M. conceived the idea, designed the experiments, analyzed the data and wrote the manuscript. S.M. and K.L. performed most experiments. K.C. performed bioinformatics analyses. X.Y., P.C., S.L., Y.Z., M.L., X.Z., X.W., and T.H. constructed plasmids, harvested RNA and performed some RT-PCR experiments. T.C. and L.W. conducted the structural predictions and analyses. Z.R., Y.L., S.M., X.Z., and X.W. provided funding support.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-60124-6.

Correspondence and requests for materials should be addressed to Zhili Rong.

Peer review information *Nature Communications* thanks Yong-Sam Kim and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.