GENETICS

A highly specific CRISPR-Cas12j nuclease enables allele-specific genome editing

Yao Wang[1], Tao Qi[1], Jingtong Liu[1], Yuan Yang[1], Ziwen Wang[2], Ying Wang[3], Tianyi Wang[1], Miaomiao Li[1], Mingqing Li[4], Daru Lu[1,5], Alex Chia Yu Chang[6], Li Yang[7], Song Gao[2], Yongming Wang[1,8]*, Feng Lan[1,9]*

The CRISPR-Cas system can treat autosomal dominant diseases by nonhomologous end joining (NHEJ) gene disruption of mutant alleles. However, many single-nucleotide mutations cannot be discriminated from wild-type alleles by current CRISPR-Cas systems. Here, we functionally screened six Cas12j nucleases and determined Cas12j-8 as an ideal genome editor with a hypercompact size. Cas12j-8 displayed comparable activity to AsCas12a and Un1Cas12f1. Cas12j-8 is a highly specific nuclease sensitive to single-nucleotide mismatches in the protospacer adjacent motif (PAM)–proximal region. We experimentally proved that Cas12j-8 enabled allele-specific disruption of genes with a single-nucleotide polymorphism (SNP). Cas12j-8 recognizes a simple TTN PAM that provides for high target site density. In silico analysis reveals that Cas12j-8 enables allele-specific disruption of 25,931 clinically relevant variants in the ClinVar database, and 485,130,147 SNPs in the dbSNP database. Therefore, Cas12j-8 would be particularly suitable for therapeutic applications.

INTRODUCTION

The RNA-guided CRISPR-Cas system is a prokaryotic adaptive immune system that has been used for mammalian genome editing (*1*, *2*). CRISPR arrays comprise multiple repeats and spacers that are transcribed and processed into multiple mature RNA species [CRISPR RNAs (crRNAs)], which direct the Cas nuclease to target foreign DNAs through complementary base pairing (*3*). Some Cas nucleases require an additional trans-activating crRNA (tracrRNA), which forms stem loops for Cas recognition (*4*). Three types of Cas nucleases have been developed as genome editors: Cas9 (type II) (*5*–*7*), Cas12 (type V) (*4*, *8*–*11*), and Cascade-Cas3 (type I) (*12*, *13*). In addition to performing conventional genome editing, the fusion protein of nuclease-deficient Cas and effector domain enables a variety of applications, including base editing (*14*, *15*), prime editing (*16*), transcriptional silencing/activation (*17*), epigenetic modulation (*18*), and genome imaging (*19*).

Several studies have demonstrated that CRISPR-Cas systems can treat autosomal dominant diseases by disrupting mutant alleles via nonhomologous end joining (NHEJ) repair pathway (*20*, *21*). The autosomal dominant disease is predominantly caused by single-nucleotide missense mutations (*22*). If the missense mutation forms a novel protospacer adjacent motif (PAM), then CRISPR-Cas nucleases can disrupt the mutant allele by a PAM-specific approach (*21*, *23*). Alternatively, the mutant allele can be disrupted by a guide-specific approach, where the missense mutation is located within the guide sequence (*20*, *24*). However, many single-nucleotide mutations cannot be discriminated from wild-type alleles by the guide-specific approach because of the low specificity of the Cas nuclease (*25*). Therefore, it is crucial to develop a highly specific Cas nuclease that can discriminate single-nucleotide mutations.

In addition to specificity, the size of Cas nucleases poses another hurdle when considering translational capability. Adeno-associated virus (AAV) vectors, the most popular method of gene delivery for gene therapy, have a packaging capacity of up to 4.7 kb (*26*). Although several small Cas nucleases together with a guide RNA (gRNA) can be codelivered by a single AAV (*27*–*29*), Cas effector fusion proteins exceed the packaging capacity of AAVs. Fortunately, a family of 10 Cas12j (CasΦ) nucleases with a length of ~700 to 800 amino acids has recently been identified (*30*). In addition to its favorable protein size, Cas12j does not require a tracrRNA, leaving more space for AAVs to deliver other components. Pausch *et al.* (*30*) characterized three Cas12j nucleases (Cas12j-1, Cas12j-2, and Cas12j- 3), of which two displayed minimal activity in mammalian and plant cells. In this study, we tested six Cas12j orthologs reported by Pausch *et al.* (*30*) (fig. S1) and identified Cas12j-8 as an efficient and highly specific genome editor. We demonstrated that Cas12j-8 can discriminate single-nucleotide mutations. Leveraging Cas12j-8's small size, we also generated a Cas12j-8–based adenine base editor (Cas12j-8ABE8e) consisting of 938 amino acids.

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Zhongshan Hospital, Fudan University, Shanghai 200438, China. [2]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China. [3]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [4]Laboratory for Reproductive Immunology, NHC Key Lab of Reproduction Regulation (Shanghai Institute of Planned Parenthood Research), Hospital of Obstetrics and Gynecology, Shanghai Medical School, Fudan University, Shanghai 200080, China. [5]NHC Key Laboratory of Birth Defects and Reproductive Health, Chongqing Key Laboratory of Birth Defects and Reproductive Health, Chongqing Population and Family Planning, Science and Technology Research Institute, Chongqing 400020, China. [6]Department of Cardiology and Shanghai Institute Precision Medicine, Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200125, China. [7]Center for Molecular Medicine, Children's Hospital, Fudan University and Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University, Shanghai 201102, China. [8]Shanghai Engineering Research Center of Industrial Microorganisms, Shanghai 200438, China. [9]State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China.
*Corresponding author. Email: fenglan@ccmu.edu.cn (F.L.); ymw@fudan.edu.cn (Y. W.)

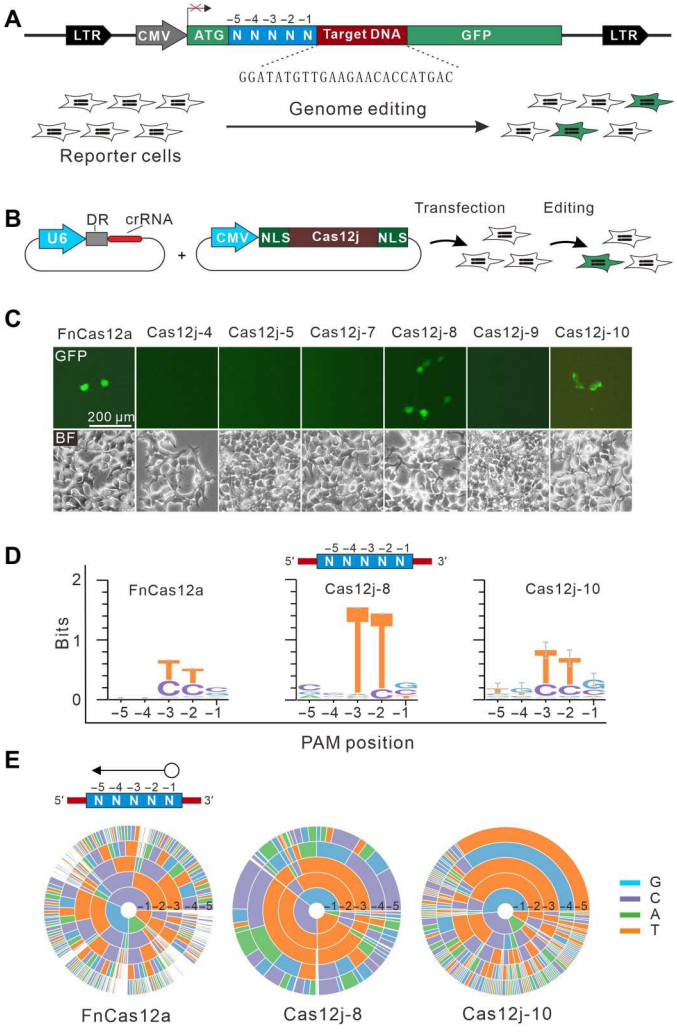


Fig. 1. Test of Cas12j activity for mammalian genome editing. (**A**) A green fluorescent protein (GFP) activation assay for testing of Cas12j activity. A target DNA with a random 5–base pair (bp) sequence at the 5′ end is inserted between the translation initiation codon (ATG) and GFP coding sequence to prevent GFP expression. The reporter library is stably integrated into human embryonic kidney (HEK) 293T cells. Genome editing results in the reestablishment of the open reading frame in a portion of cells, resulting in GFP expression. CMV, cytomegalovirus. LTR, long terminal repeated. (**B**) A Cas12j expression plasmid was cotransfected with a crRNA expression plasmid into reporter cells for genome editing. DR, direct repeat. NLS, nuclear localization sequence.(**C**) Transfection of Cas12j with crRNA induces GFP expression. FnCas12a is used as a positive control. Cas12j-9's crRNA is used for Cas12j-10. BF, bright field.(**D**) WebLogo diagrams for FnCas12a, Cas12j-8, and Cas12j-10 are generated on the basis of deep sequencing data. (**E**) PAM wheels for FnCas12a, Cas12j-8, and Cas12j-10 are generated on the basis of deep sequencing data.

RESULTS

Activity analysis of six Cas12j orthologs

We first analyzed the homology of CRISPR repeats belonging to different Cas12j orthologs (*30*). These repeats contain a conserved "ATTGC" motif in the middle and a "GAC" motif at the 3′ end (fig. S2A). The secondary structure prediction revealed that these repeats can form conserved 5-nucleotide (nt) stems and tetraloops (fig. S2B).

Next, we tested whether the six Cas12j orthologs (Cas12j-4, Cas12j-5, Cas12j-7, Cas12j-8, Cas12j-9, and Cas12j-10) can function for mammalian genome editing. Cas12j-6 looks like a truncated protein and was excluded from this study. A previously established green fluorescent protein (GFP) activation assay (*6*) was used to test the Cas12j activity (Fig. 1A). In this assay, a target DNA with a random 5–base pair (bp) sequence is inserted between the translation initiation codon (ATG) and GFP coding sequence to prevent GFP expression. If a Cas12j enables genome editing, then insertions and deletions (indels) would occur at the target sequence and GFP expression can be observed in a portion of cells because of inframe mutation. FnCas12a (*8*) was used as a positive control. Each *Cas12j* ortholog was human codon-optimized, synthesized, and cloned into a mammalian SaCas9 expression construct (*27*). The native CRISPR repeat was used as the crRNA scaffold for each Cas12j ortholog except for Cas12j-10, whose CRISPR repeat has not yet been identified. The repeats of Cas12j-2, Cas12j-4, Cas12j-5, Cas12j-7, Cas12j-8, and Cas12j-9 were tested as scaffolds for Cas12j-10.

The individual Cas12j orthologs and the corresponding 24-nt crRNAs were transfected into reporter cells (Fig. 1B). Three days after transfection, GFP-positive cells were observed for FnCas12a and Cas12j-8 (Fig. 1C). When Cas12j-9's repeat was used as a scaffold, GFP-positive cells were observed for Cas12j-10 (Fig. 1C). No GFP-positive cells were observed when other repeats were used (fig. S3). These data demonstrated that Cas12j-8 and Cas12j-10 can potentially be used for mammalian genome editing.

PAM analysis

Next, we investigated the PAMs required for these active nucleases. GFP-positive cells were isolated by flow cytometry, and the target DNA was polymerase chain reaction (PCR)–amplified for deep sequencing. The sequencing results revealed that indels occurred at the target sites for FnCas12a, Cas12j-8, and Cas12j-10 (fig. S4). Both the WebLogo diagram and PAM wheel revealed that FnCas12a preferentially recognized a TTN PAM (Fig. 1, D and E), consistent with a previous study (*8*). The results revealed that Cas12j-8 and Cas12j-10 also preferentially recognized a TTN PAM (Fig. 1, D and E). The PAM requirements of FnCas12a and Cas12j-8 were confirmed by editing another target sequence (fig. S5, A to C). To further investigate the PAM requirement for Cas12j-8, we incubated purified Cas12j with crRNA and a library of plasmids containing randomized PAMs (fig. S6A). After incubation, the cleaved DNA containing randomized PAMs was isolated for deep sequencing. The sequencing results confirmed that Cas12j-8 preferentially recognized a TTN PAM (fig. S6B). We focused on Cas12j-8 in the following study because it displayed higher activity in the GFP activation assay.

Analysis of Cas12j-8 specificity

Next, we analyzed the specificity of Cas12j-8 by using a previously developed GFP activation assay (*6*). FnCas12a and Cas12j-2 were used for comparison. We generated a panel of 12 crRNAs with dinucleotide mismatches. FnCas12a and Cas12j-2 displayed substantial off-target cleavage among these crRNAs, while Cas12j-8 only generated background levels of GFP-positive cells when mismatches occurred at PAM-proximal positions 1 to 14 (fig. S7A). We further generated a panel of 24 crRNAs with a single mismatch. Cas12j-8 generated minimal levels of GFP-positive cells when a single

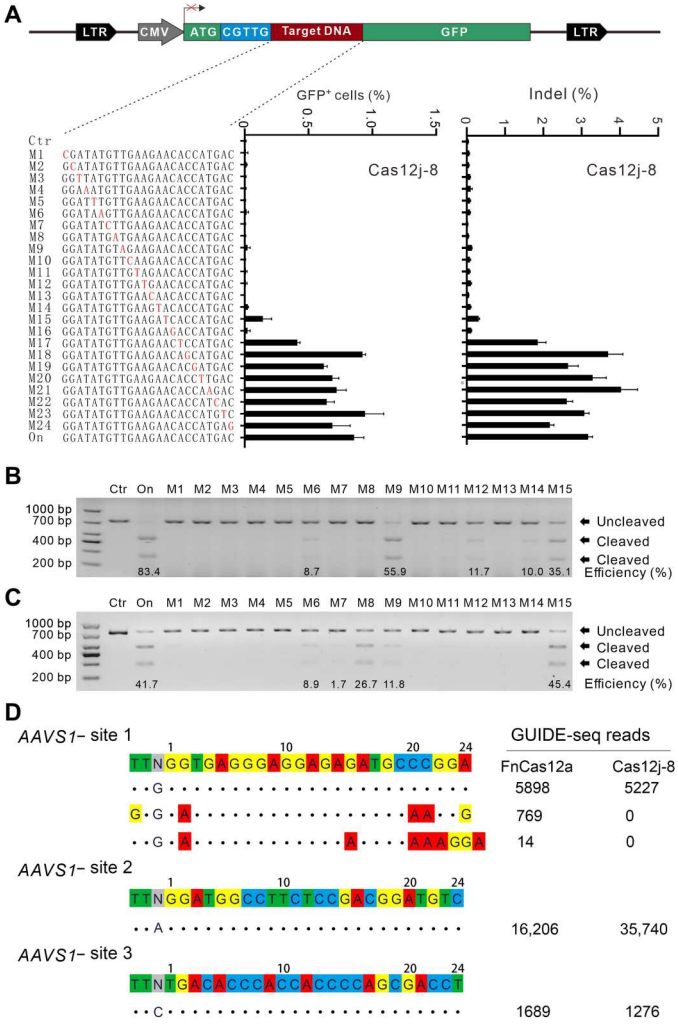


Fig. 2. Analysis of the Cas12j-8 specificity. (**A**) Schematic of the GFP activation assay for specificity analysis is shown on the top. A panel of crRNAs with a single-nucleotide mutation is shown below. Editing efficiencies are quantified as the percentage of GFP-positive cells. Mismatches are shown in red (n = 3). Ctr: Reporter cells without transfection are used as a negative control. (**B** and **C**) An in vitro cleavage assay is used to test Cas12j-8 specificity. The purified Cas12j-8, individual crRNAs, and the target DNAs are incubated for 2 hours and then analyzed on an agarose gel. The same crRNAs and target DNA are used in (A) and (B). (**D**) The genome-wide off-target effects of FnCas12a and Cas12j-8 are analyzed by GUIDE-seq. On-target and off-target sequences are shown on the left. Read numbers are shown on the right. Mismatches compared to the on-target site are shown and highlighted in color.

mismatch occurred at PAM-proximal positions 1 to 14 and 16 (Fig. 2A). Targeted deep sequencing results revealed that Cas12j-8 generated minimal levels of indels at PAM-proximal positions 1 to 14 and 16, consistent with the GFP activation assay (Fig. 2A).

During our manuscript revision, a Cas12f system has been developed for genome editing (*9*–*11*). This system contains a small Cas effector protein (400 to 700 amino acids) and requires a tracrRNA for genome editing. We further investigated the dinucleotide and single-mismatch tolerance of Un1Cas12f1 by using the GFP activation assay. Un1Cas12f1 generated background levels of GFP-positive cells when dinucleotide mismatches occurred at PAM-proximal positions 1 to 17; Un1Cas12f1 generated minimal levels of GFP-positive cells when a single mismatch occurred at PAM-proximal positions 1 and 4 to 17 (fig. S7B).

To further investigate Cas12j-8 specificity, we performed the in vitro digestion assay by incubating the purified Cas12j-8 with individual crRNAs bearing a single mismatch, and the target DNA was PCR-amplified from the GFP activation. Cas12j-8 was highly sensitive to the single mismatch at PAM-proximal positions 1 to 5, 7 to 8, 10 to 11, and 13 and relatively less sensitive at other positions (Fig. 2B). We performed the in vitro digestion assay with another target, and a similar specificity was observed (Fig. 2C). It was expected that higher off-target effects could be detected by the in vitro cleavage assay because the in vitro cleavage activity was much higher than the in vivo cleavage activity.

Next, we analyzed the genome-wide off-target effects of Cas12j-8 by using genome-wide, unbiased identification of DSBs enabled by sequencing (GUIDE-seq) (*31*), and FnCas12a was used as a control. We designed three crRNAs targeting *AAVS1* locus. After transfection of the Cas12/crRNA plasmid and the GUIDE-seq oligos into human embryonic kidney (HEK) 293T cells for 5 days, we prepared libraries for deep sequencing. The sequencing results revealed that on-target cleavage occurred for FnCas12a and Cas12j-8 at all targets, as revealed by the high GUIDE-seq read counts (Fig. 2D). Two off-target sites were identified for FnCas12a at *AAVS1* site 1 locus, but no off-target sites were identified for Cas12j-8. Next, we compared the specificity of Cas12j-8 to FnCas12a, Un1Cas12f1, and Cas12j-2 using GUIDE-seq targeting *EMX1*, *NLRC4*, *P2RX5-TAX1BP3*, *CLIC4*, and *Intergene*. FnCas12a, Un1Cas12f1, Cas12j-2, and Cas12j-8 generated off-target sites of 2, 2, 0, and 1 at *EMX1* locus (fig. S8). No off-targets were detected for the remaining targets. No on-target reads were detected for Cas12j-2 for the remaining targets. These data indicated that these Cas12 nucleases were highly specific for genome editing.

Investigation of Cas12j-8 activity

First, we designed two panels of crRNAs varied from 14 to 24 nt to test the Cas12j-8 activity with the GFP activation assay. The results revealed that 18-nt crRNAs achieved the most efficient editing (fig. S9, A and B). Next, we compared the activity of Cas12j-8 to FnCas12a and Cas12j-2 because they recognized the same TTN PAM. These nucleases were expressed from the same construct backbone (Fig. 3A), and similar gene expression levels were observed by reverse transcription quantitative PCR (RT-qPCR) analysis (Fig. 3B). We tested the activities of these nucleases at 12 endogenous target loci containing TTN PAMs in HEK293T cells. All three Cas12 nucleases generated indels at the selected sites (Fig. 3C). Cas12j-8 displayed higher activity than FnCas12a and Cas12j-2 (Fig. 3D). We tested Cas12j-8 in additional cell types, including A375, HeLa, SH-SY5Y, and C33A cells. Cas12j-8 generated indels in all these cell types with varying efficiencies (fig. S10, A to D). Cas12j-8 also could be delivered by AAV for genome editing (Fig. 3, C and D). These data demonstrated that Cas12j-8 can function in a variety of cell types.

Next, we compared the activity of Cas12j-8 to the commonly used SpCas9 and AsCas12a. We cloned SpCas9 and AsCas12a to the Cas12j-8 expression construct (fig. S11A), and similar gene expression levels were confirmed by RT-qPCR (fig. S11B). We tested the activities of these nucleases at 11 endogenous target loci containing TTTN PAMs at the 5′ end and NGG PAM at the 3′ end

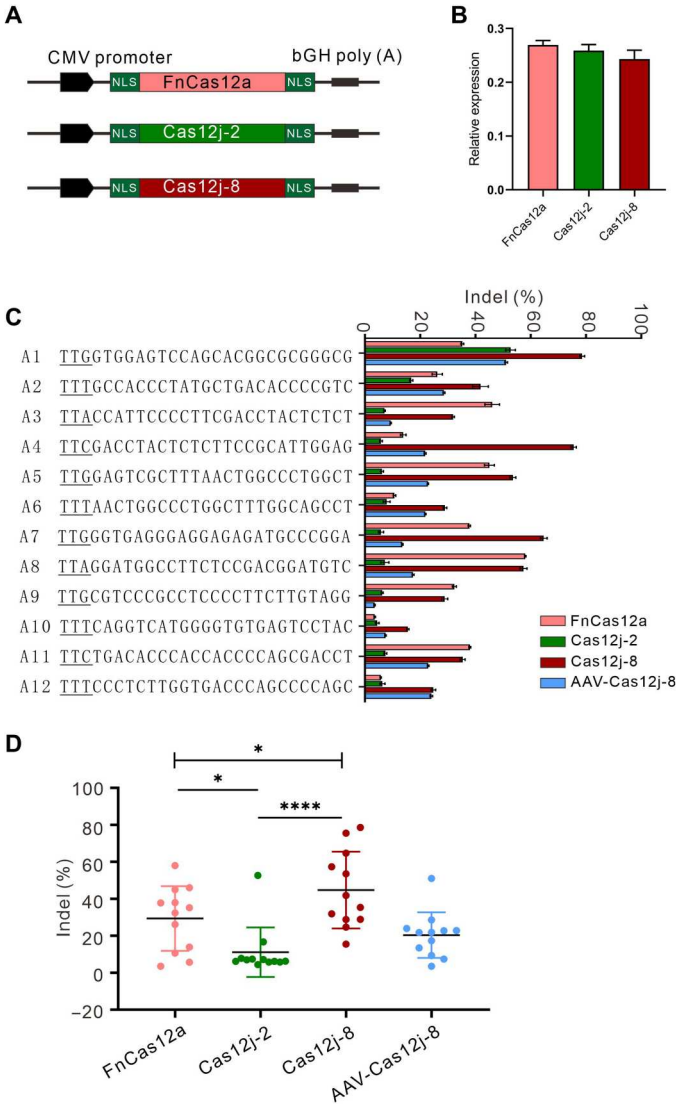


Fig. 3. Analysis of Cas12j-8 activity. (**A**) Schematic of the Cas12 expression constructs. All three Cas12 expression constructs have the same backbone. Poly (A), polyadenylated. (**B**) Expression levels of FnCas12a, Cas12j-2, and Cas12j-8 relative to glyceraldehyde-3-phosphate dehydrogenase were measured by reverse transcription quantitative PCR (RT-qPCR). (**C**) Editing efficiency of FnCas12a, Cas12j-2, and Cas12j-8 in HEK293T was measured by targeted deep sequencing (n = 3). PAMs are underlined. Sites A1 to A12 are located in *AAVS1* loci. AAV-Cas12j-8: Cas12j-8 is delivered by AAV. (**D**) Quantification of editing efficiencies. For post hoc analysis, repeated-measures (RM) one-way analysis of variance (ANOVA) test followed by Fisher's least significant difference test was used. A value of $P < 0.05$ was considered to be statistically significant (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$).

(fig. S11C). All three nucleases could generate indels at these targets (fig. S11D). Overall, Cas12j-8 displayed higher activity than AsCas12a and lower activity than SpCas9, although not significant (fig. S11E).

We further compared the activity of Cas12j-8 to the optimized Un1Cas12f1 that recognized a TTTR (R=A or G) PAM. We cloned Un1Cas12f1 into the Cas12j-8 expression construct (fig. S12A), and similar gene expression levels were confirmed by RT-qPCR (fig. S12B). Our results demonstrated that Cas12j-8 displayed similar activity to Un1Cas12f1 (fig. S12, C to E).

Base editing with Cas12j-8

Next, we investigated whether Cas12j-8 can be used for base editing. We constructed a SWISS-MODEL of Cas12j-8 based on the LbCas12a crystal structure (Fig. 4, A and B) (*32*). The structure revealed that the Cas12j-8 residue E567 corresponds to the LbCas12a residue E925, which is essential for DNA cleavage activity. We mutated E567 to alanine to generate a catalytically dead Cas12j-8 (dCas12j-8). dCas12j-8 could not edit endogenous loci, while wild-type Cas12j-8 could edit these loci (Fig. 4C). In vitro digestion assay revealed that dCas12j-8 could not cleave a target DNA (Fig. 4D). These results demonstrated that dCas12j-8 was inactivated.

dCas12j-8 was fused to an engineered adenine deaminase enzyme (TadA8e) (*33*), resulting in a base editor that we named Cas12j-8ABE8e. Cas12j-8ABE8e contained only 938 amino acids. We tested the activity of Cas12j-8ABE8e at 20 endogenous loci, and we observed editing at four loci (Fig. 4E). Recent studies have developed adenine base editors, which can be delivered by a single AAV (*34*, *35*). Inspired by their work, we also delivered Cas12j-8ABE8e by a single AAV, but we failed to detect editing events probably due to the low editing efficiency. It would be interesting to generate a Cas12j-8 nickase to improve Cas12j-8ABE8e efficiency in the future work. We also generated a cytosine base editor (Cas12j-8CBE-hA3A) by fusing dCas12j-8 to an hAPOBEC3A-hA3A (*36*) and two monomers of uracil glycosylase inhibitors. However, we observed no editing activity at any of the 50 endogenous loci in HEK293T cells (fig. S13). The cytidine base editor required fusion on proteins at both sites of Cas12j-8, which may disrupt the targeting ability of Cas12j-8.

Allele-specific genome editing with Cas12j-8

A recent study has shown that if a natural variant lies in cis with the causative mutation, then this variant can be used to disrupt the mutant allele (*37*). As a proof of concept, we used Cas12j-8 to allele-specifically disrupt four loci containing single-nucleotide polymorphisms (SNPs). These SNPs were identified from the National Center for Biotechnology Information (NCBI) SNP database and confirmed by Sanger sequencing. We selected loci containing two SNPs so that we can differentiate individual alleles after genome editing. Target alleles could be efficiently disrupted with low off-target effects at nontarget alleles (Fig. 5, A to D).

In silico analysis revealed that Cas12j-8 could allele-specifically disrupt 25,931 clinically relevant variants in the ClinVar database (*38*) and 485,130,147 SNPs in the dbSNP database (Fig. 5E) (*39*). Un1Cas12f1 can also be potentially used to disrupt pathogenic mutations. In silico analysis revealed that Un1Cas12f1 could allele-specifically disrupt 10,443 clinically relevant variants in the ClinVar database (*38*) and 253,229,256 SNPs in the dbSNP database (Fig. 5E) (*39*). Cas12j-8 has approximately twice the number of editing sites as Un1Cas12f1. The crRNAs designed for clinically relevant variants were deposited on a website (www.deephf.com/#/resources/snp-edit).

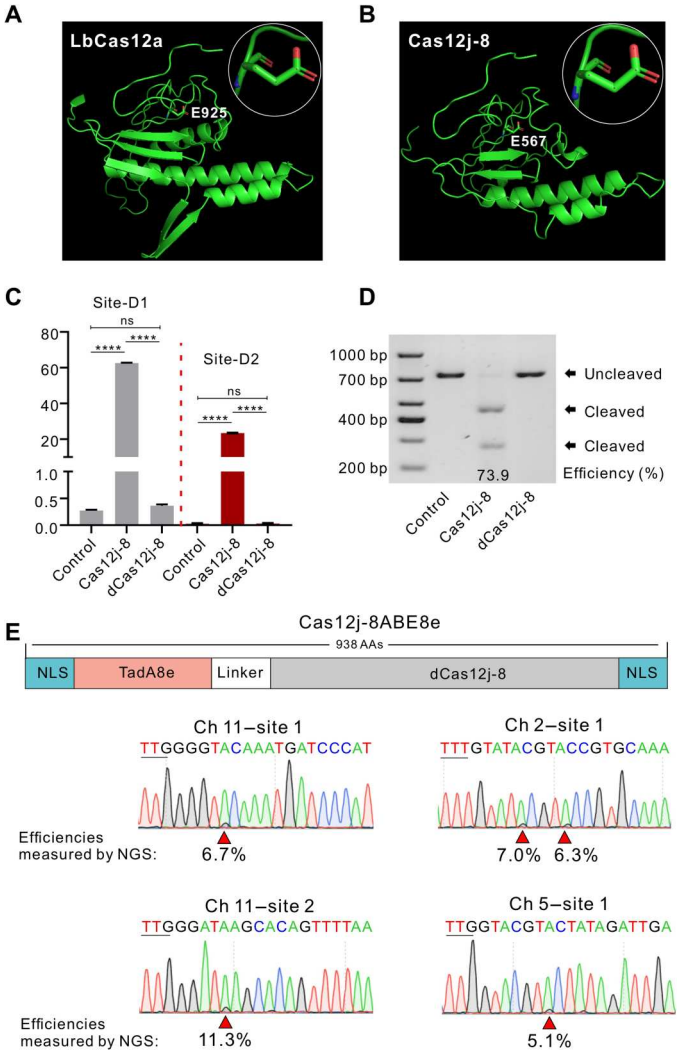


Fig. 4. Cas12j-8 enables base editing. (**A**) LbCas12a crystal structure. (**B**) SWISS-MODEL of Cas12j-8 structure built according to the LbCas12a crystal structure [SWISS-MODEL Template Library (SMTL) ID: 6kl9.1]. Residue E567 in Cas12j-8 was mutated to alanine, generating a dead Cas12j-8 (dCas12j-8). (**C**) Test of dCas12j-8 activity at two endogenous loci. A value of $P < 0.05$ was considered to be statistically significant (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$). Sites A1 and A10 are located in *AAVS1* loci. Cells without transfection were used as negative controls. The wild-type Cas12j-8 was used as a positive control ($n = 3$). ns, not significant. (**D**) Test of dCas12j-8 activity by an in vitro cleavage assay. Target DNA without digestion was used as a negative control. The wild-type Cas12j-8 was used as a positive control. Digestion efficiency is shown below. (**E**) Base editing with Cas12j-8ABE8e. The schematic of the Cas12j-8ABE8e is shown above. A-to-G conversions are indicated by the red triangles. Conversion efficiencies were quantified by deep sequencing and are shown below. NGS, next-generation sequencing; AAs, amino acids.

DISCUSSION

Allele-specific gene disruption through NHEJ is a potential strategy to treat autosomal dominant diseases (*20*, *21*), where the causative gene is haploinsufficient. Compared to gene correction by homology-directed repair, base editing, or prime editing, NHEJ is more efficient and feasible for in vivo editing. The guide-specific approach relies on a highly specific CRISPR-Cas system to differentiate single-nucleotide mutations. Many efforts have been made to improve the specificity of Cas9, but these high-fidelity CRISPR-Cas9 variants substantially tolerate single-nucleotide mutations (*40*, *41*). In this study, we demonstrated that Cas12j-8 is sensitive to single-nucleotide mutations and enables allele-specific genome editing, holding great promise for therapeutic applications.

Two compact Cas12f nucleases (Un1Cas12f1 and AsCas12f) have recently been developed as genome editors (*10*, *11*). Un1Cas12f1 is particularly interesting due to its high efficiency and specificity. We compared Cas12j-8 to Un1Cas12f1 and demonstrated that they displayed similar editing efficiency. Cas12j-8 exhibited less tolerance for single-nucleotide mismatches in the PAM-proximal regions (positions 1 to 13), while Un1Cas12f1 exhibited less tolerance for single-nucleotide mismatches in the middle region (positions 6 to 17). Because of its simple PAM requirement, Cas12j-8 has approximately twice the number of editing sites than Un1Cas12f1 in the ClinVar and dbSNP databases. We expect that these two Cas12s can complement each other for clinical applications.

MATERIALS AND METHODS

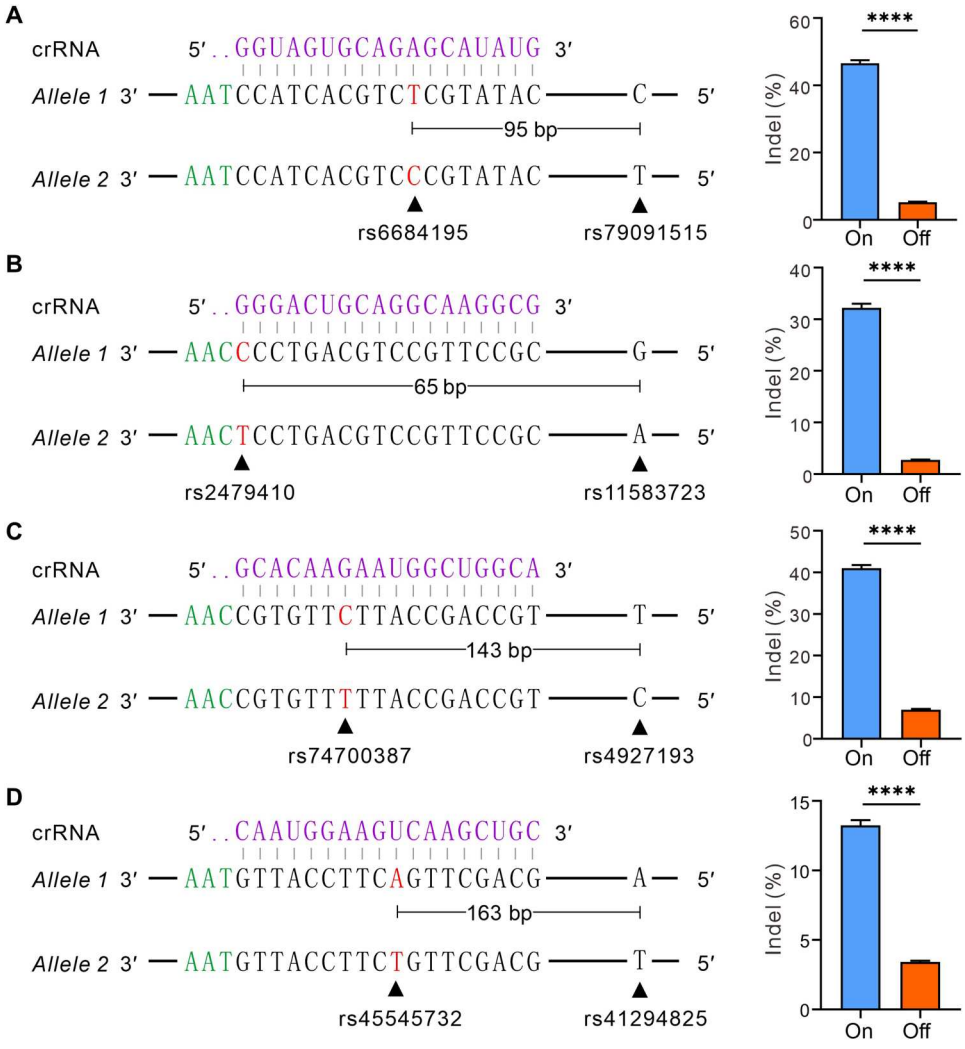
Cell culture and transfection

Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and streptomycin (100 U/ml) was used to culture HEK293T, C33A, SH-SY5Y, and HeLa cells at 37°C and 5% CO_2. A375 cells were maintained in RPMI 1640 medium supplemented with 10% FBS (Gibco). GFP reporter cells were plated in 10-cm dishes and transfected using Lipofectamine 2000 (Life Technologies) at approximately 60% confluence with Cas12 (10 μg) + crRNA plasmid (5 μg) for PAM screening of Cas12j-8, Cas12j-10, and FnCas12a. To test the editing capability of Cas12j-8 at the endogenous sites, HEK293T cells were seeded in 48-well plates and transfected with the Cas12 plasmid (300 ng) + crRNA (200 ng) plasmid or the Cas12j-8ABE–crRNA expression plasmid (500 ng) using Lipofectamine 2000 (Life Technologies). To test the specificity, GFP reporter cells were cultured in 48-well plates and transfected with the Cas12 plasmid (300 ng) + crRNA (200 ng) plasmid using Lipofectamine 2000 (Life Technologies).

Plasmid construction

For pAAV-CMV-Cas-puro, fragments of *SpCas9*, *AsCas12a*, and *FnCas12a* were PCR-amplified using pSpCas9(BB)-2A-puro (PX459) V2.0 (Addgene no. 62988), pCMV-T7-AsCas12a-P2A-EGFP (Addgene no. 160140), and FnoCas12a-2C-NLS (Nuclear localization sequence) in pCSDest (Addgene no. 126639) as templates. pAAV backbone was PCR-amplified with primers pAAV-F and pAAV-R using pAAV-CMV-SauriCas9-puro (Addgene no. 135965) as a template. Fragments of *Un1Cas12f1* and *Cas12j* orthologs were human codon-optimized and synthesized into pAAV backbone by HuaGene (Shanghai, China). These fragments of *SpCas9*, *AsCas12a*, and *FnCas12a* were recombined with pAAV backbone using the NEBuilder HiFi DNA Assembly Cloning Kit [New England Biolabs (NEB), E5520] to generate pAAV-CMV-Cas-puro vectors. In this study, all primers were listed in table S1. In this study, all *Cas* ortholog sequences are listed in table S2.

For PSK (pBluescriptSK)-mU6-Cas-gRNA, a fragment containing a mU6 promoter and a Cas12j-8 repeat was synthesized by



E

	Un1Cas12f1	Cas12j-8
Pathogenic sites in ClinVar database (35,964)	10,443	25,931
dbSNP (597,146,516)	253,229,256	485,130,147

Fig. 5. Cas12j-8 enables allele-specific genome editing. (**A** to **D**) Cas12j-8 enables allele-specific disruption of SNPs. PAMs are shown in green; target SNPs are shown in red; crRNAs are shown in purple; the distance between two SNPs is indicated below the target allele; indel efficiencies are shown on the right. A value of $P < 0.05$ was considered to be statistically significant (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$). (**E**) Number of pathogenic sites or SNPs are potentially disrupted by Cas12j-8 and Un1Cas12f1.

HuaGene (Shanghai, China) and inserted into Xho 1 and Cla 1 digested pSKB- vector (Addgene no. 62540) using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB, E5520) to generate PSK-mU6-Cas12j-8-gRNA vector, which was used to express crRNA of Cas12j-8 in eukaryotes. To obtain vectors that express gRNA of SpCas9, AsCas12a, FnCas12a, Un1Cas12f1, and Cas12j-2, the gRNA scaffolds were synthesized by HuaGene (Shanghai, China) and recombined with PSK-mU6 backbone, which was PCR-amplified with primers PSK-F and PSK-R using PSK-mU6-Cas12j-8-gRNA vector as a template. The sequences of gRNA scaffolds were listed in table S2. gRNAs were inserted into the PSK-mU6-Cas-gRNA plasmid between two Bbs 1 restriction sites. All target sites were listed in table S3.

For pAAV-Cas12j-8-gRNA-puro, pAAV-CMV-Cas12j-8-puro and PSK-mU6-Cas12j-8-gRNA vectors were digested with Kpn 1 and Not 1, and two target fragments were assembled with T4 DNA ligase (NEB), resulting in pAAV-Cas12j-8-gRNA-puro vector for AAV delivery.

For pAAV-dCas12j-8-ABE8e-puro, primers pAAV-Cas12j-8-F and pAAV-Cas12j-8-R were used to generate pAAV-Cas12j-8-puro backbone. The envTadA fragment was PCR-amplified using envTadA-F and envTadA-R with ABE8e (Addgene no. 138489) as a template. The two fragments were cloned using the NEBuilder HiFi DNA Assembly Cloning Kit. Primers E567-F and E567-R were used to construct pAAV-dCas12j-8-ABE8e-puro for introducing E567A mutation to the plasmid with blunt-end ligation.

For pCMV-Cas12j-8-CBE-hA3A, primers CBE-F and CBE-R were used to generate pCMV-CBE backbone with pCMV-AncBE4-max (Addgene no. 354456) as a template, and primers cas-CBE-F and cas-CBE-R were used to generate dCas12j-8 fragment with pAAV-dCas12j-8-ABE8e-puro as a template. The two fragments were cloned using the NEBuilder HiFi DNA Assembly Cloning Kit. Primers CBE-hA3A-F and CBE-hA3A-R were used to generate the backbone, and the hA3A fragment was PCR-amplified using hA3A-F and hA3A-R with pCMV-hA3A-BE3 (Addgene no. 113410) as a template. The two fragments were cloned using the NEBuilder HiFi DNA Assembly Cloning Kit, resulting in the pCMV-Cas12j-8-CBE-hA3A vector.

For pET28a-HRV 3C-Cas, to obtain pET28a-HRV 3C-Cas12j-8 and pET28a-HRV 3C-dCas12j-8 plasmids, the protein-coding sequences were PCR-amplified with primers Cas12j-8-pskb-F and Cas12j-8-pskb-R. The protein-coding fragments and pET28a-HRV 3C vector were digested with the restriction enzymes Bam H1 and Xho l1 (NEB). Then, the linearized fragments were ligated with T4 DNA ligase (NEB).

Expression and purification of Cas12j-8 and dCas12j-8 protein

For purification of Cas12j-8 and dCas12j-8, competent *Escherichia coli* Rosetta cells were transformed with pET28a-HRV 3C-Cas12j-8 or pET28a-HRV 3C-dCas12j-8 plasmid and cultured until OD_{600} (optical density at 600 nm) = 0.4 to 0.6 in 2-liter LB medium supplemented with kanamycin (10 mg/ml). The target protein synthesis was induced by the addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside. After 16 hours of growth at 16°C, cells were centrifugated at 4000*g*; the pellet was resuspended in lysis buffer containing 300 mM NaCl, 50 mM Hepes, 1 mM β-mercaptoethanol, and 30 mM imidazole; and cells were lysed by sonication. The cell lysate was centrifuged at 16,000*g* (4°C) and filtered through 0.45-μm filters. The lysate was applied to 1-ml HisTrap HP column, and Cas12j-8 or dCas12j-8 was eluted by 300 mM imidazole in the same buffer. The fractions containing the protein of interest were pulled, and the HRV 3C-tag was cleaved by overnight incubation with PreScission Protease (PSP) at 4°C. To remove the cleaved HRV 3C-tag and PSP, reaction mixtures were loaded onto a HisTrap HP column and a GSTrap HP column, respectively. The Cas12j-8 and dCas12j-8 proteins were collected in the flow-through, and the collected fractions were further purified with gel filtration chromatography. The purity of the nucleases was assessed by denaturing 8% polyacrylamide gel electrophoresis. Sequences of Cas12j-8 and dCas12j-8 proteins are listed in table S2.

RNA synthesis

Cas12j-8 crRNAs were produced by in vitro transcription using the HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB) and purified using phenol-chloroform extraction and ethanol precipitation. Templates for T7 transcription were generated by PCR, containing a T7 promoter at the proximal end followed by the crRNA sequence. The sequences of $T7^{-}$ crRNAs used in this study are listed in table S2.

In vitro cleavage assays

Before the cleavages, crRNAs were dissolved in diethyl pyrocarbonate H_2O and heated for 5 min at 95°C before cooldown on the ice. Active Cas12j-8 ribonucleoprotein (RNP) complexes were assembled in a 1:1.2 molar ratio by diluting Cas12j-8 protein to 4 μM and crRNA to 5 μM in RNP assembly buffer (10 mM Hepes-K, 150 mM KCl, 5 mM $MgCl_2$, and 0.5 mM TCEP [Tris (2-carboxyethyl) phosphine]) and incubated for 30 min at room temperature. Cleavage reactions were initiated by the addition of DNA (10 nM) to preformed RNP (1 μM) in reaction buffer (10 mM Hepes-K, 150 mM KCl, 5 mM $MgCl_2$, and 0.5 mM TCEP). The samples were incubated at 37°C for 2 hours and then treated with 0.8 U of proteinase K (NEB) for 20 min at 37°C. Reaction products were analyzed by electrophoresis in 1.5% agarose gels. ImageJ was used to analyze the gray level of the strip and calculate the editing efficiency.

Construction of GFP activation system

We synthesized a lentiviral plasmid library, in which a target DNA with a random 5-bp sequence was inserted between the translation initiation codon (ATG) and GFP coding sequence to prevent GFP expression. The plasmid library was packed into the lentivirus, and the titration of the lentivirus library was detected with qPCR. HEK293T was infected at a multiplicity of infection ≤ 1, and the cells were cultured with puromycin for 5 to 7 days after infection. The GFP-positive cells induced by mutations were removed by MoFlo XDP machine. The GFP reporter cell library was cultured in 10-cm dishes to keep the integrity of the library.

Analysis of PAM sequence in the GFP activation assay

The plasmids of Cas12 and crRNA were transfected into the GFP reporter cell library. After the genome editing of the target site, random indel resulted in GFP inframe mutation, leading to the positive GFP of a portion of cells. The GFP-positive cells were sorted out by flow cytometry, and the genome was extracted, which would be used in the following deep sequencing. While in the analysis of the PAM sequence, 15-bp sequences (TTGTTTGCCACCATG/GT GAGCAAGGGCGAG) flanking the target sequence were used to locate the target sequence (library 1: NNNNNGGATATGTTGAA GAACACCATGAC or library 2: NNNNNGGGCTTCAAGCAAC TTGTAGTGGG). TTGTTTGCCACCATG and GGA (library 1) or GGG (library 2) was used to locate the 5-bp random sequence. Target sequences with in-frame mutations were used for PAM analysis. The 5-bp random sequence was extracted and visualized by WebLog3 and PAM wheel charts to display PAMs. Library 1 was used in Fig. 1, and library 2 was used in fig. S6.

Analysis of PAM in vitro

CRISPR-Cas12j-8 system was modified to target a 5-bp randomized PAM plasmid library. The negative control was set as the same condition without a crRNA. After incubation at 37°C, to efficiently capture the blunt ends of the plasmid library generated by Cas12-gRNA complex cleavage, a 3′ dA (deoxyadenosine monophosphate) was added to the cleaved DNA fragments. Reaction products were purified and linked to adapters with a 3′ dT (deoxythymidine monophosphate) overhang. Next, to selectively enrich cleaved products

containing the PAM sequence, the target sequence was PCR-amplified by nested PCR and purified with the Gel Extraction Kit (QIAGEN) for deep sequencing. Sequences (TTGTTTGCCAC CATG/GGATAT) flanking the target sequence were used to locate the 5-bp random sequence. The frequency of the extracted PAM sequences normalized to the original PAM library to account for inherent biases using the following formula

$$\text{Normalized frequency} = \frac{\text{Treatment frequency}}{\text{Control frequency}/\text{Average control frequency}}$$

After normalization, the biases of every PAM between the negative control group and the experimental group were analyzed. Biases were considered significant if they deviated by >2.5-fold from the negative control. Analyses were limited to the top 10% of the most frequent PAMs to reduce the impact of background noise.

Genome editing at endogenous sites

HEK293T cells were cultured into a 48-well plate and transfected with the Cas12 plasmid (300 ng) + crRNA plasmids (200 ng) with Lipofectamine 2000 (Life Technologies). Cells were selected with puromycin and collected 5 to 7 days after transfection. For HeLa, A375, SH-SY5Y, and C33A cells, we transfected Cas12 plasmid (300 ng) + crRNA plasmids (200 ng) with Lipofectamine 3000 (Life Technologies). Cells were selected with puromycin and collected 5 to 7 days after transfection. Genomic DNA was isolated, and the target sites were PCR-amplified by nested PCR and purified with the Gel Extraction Kit (QIAGEN) for deep sequencing.

RT-qPCR

Total RNAs were extracted using TRIzol reagent (Invitrogen), and RT was performed using RT SuperMix for qPCR (APExBIO). qPCR was performed to measure the expression of Cas proteins relative to glyceraldehyde-3-phosphate dehydrogenase expression using 2X SYBR Green qPCR Master Mix (APExBIO). Primers used for qPCR were listed in table S1.

AAV production

HEK293T cells were seeded at approximately 40% confluency in a 10-cm dish the day before transfection. For each well, 2 μg of pAAV-Cas12j-8-gRNA-puro plasmid, 2 μg of pAAV-RC plasmid (GeneBank, AF369963), and 4 μg of pAAV helper plasmid (GeneBank, AF369965.1) were transfected using 80 μl of PEI [Poly (ethylenimine)] [0.1% (m/v); Polysciences, catalog no. 23966]. The pAAV helper plasmid contains the subset of adenovirus genes VA, E2A, and E4, which are necessary for high-titer AAV production. The medium was changed 8 hours after transfection. After 72 hours, cells were scraped and pelleted by centrifugation, and the supernatant was collected for later use. The cell pellet was resuspended in phosphate-buffered saline (PBS). The PBS containing cell pellet was frozen and melted repeatedly three times. After centrifugation, the supernatant was added to the supernatant above. The mixture was filtered with a 0.45-μm polyvinylidene fluoride filter. A mixed solution (1 M NaCl + 10% PEG-8000) was added, and the solution was incubated at 4°C overnight. After centrifugation at 4°C for 2 hours at 12,000*g*, the supernatant was discarded, and the AAV viral particles were resuspended in 500 μl of chilled PBS. The qPCR reveals that AAV titration is 4.0×10^8 copies/μl. One hundred microliters of the virus was added into a 12-well plate with ~80% confluency of HEK293T cells. The transduced cells were maintained in DMEM with puromycin for up to 2 weeks and then were collected for the detection of editing efficiency at target loci.

Base editing with Cas12j-8ABE

HEK293T cells were seeded in 48-well plates and transfected with the Cas12j-8ABE–crRNA expression plasmid (500 ng). Cells were selected with puromycin and collected 5 to 7 days after transfection. The genomic DNA was isolated, and the target sites were PCR-amplified by nested PCR and purified with the Gel Extraction Kit (QIAGEN) for deep sequencing.

Test of Cas specificity with the GFP activation assay

The plasmids containing the TTN PAM sequence were isolated from the several of PAM vector library. The plasmid was packaged into lentiviruses to generate a stable cell line. After removing the background mutations that induced a positive GFP signal, the GFP activation cell line was used to reveal the editing ability of Cas and the corresponding single-guide RNA. The GFP activation cell line 1 (CATTGCGCTGGATCGTGAGCAAGGGCGAG) and line 2 (CCTTGCTCACGATCCAGCGCAATGATGAT) were used to analyze the optimum spacer length of Cas12j-8. Line 3 (TTGGGATATGTTGAAGAACACCATGAC) was used to analyze the specificity of FnCas12a, Cas12j-2, and Cas12j-8. Line 4 (TTGGATATGTTGAAGAACACCAT) was used to analyze the specificity of Un1Cas12f1. Three days after transfection, random indel caused by editing led to GFP inframe mutation, so a portion of cells emitted GFP fluorescent signals. Then, the GFP-positive cells were analyzed on a FACSCalibur instrument (BD Biosciences), and the data were analyzed using FlowJo.

GUIDE-seq

GUIDE-seq experiments were performed with FnCas12a, Cas12j-2, and Cas12j-8 at three loci, respectively. HEK293T cells (2×10^5) were transfected with the Cas12 plasmid (1 μg) + crRNA plasmids (500 ng) + 10 pmol of annealed GUIDE-seq oligonucleotides by electroporation and then cultured into 12 wells. Electroporation voltage, width, and the number of pulses were 1150 V, 30 ms, and 1 pulse, respectively. Cells were selected with puromycin and collected 7 days after transfection. Genomic DNA was isolated with the DNeasy Blood and Tissue Kit (QIAGEN), and the target sites were PCR-amplified by nested PCR and purified with the Gel Extraction Kit (QIAGEN) for deep sequencing. Two-tailed, paired Student's *t* tests were used to determine statistical significance when comparing two groups, whereas analyses of variance (ANOVAs) are used when comparing more than two groups.

Quantification and statistical analysis

All data are shown as the means ± SD values. Statistical analyses were conducted using GraphPad Prism 8. Two-tailed, paired Student's *t* tests were used to determine statistical significance when comparing two groups, whereas ANOVA followed by Fisher's least significant difference test are used when comparing more than two groups. A value of $P < 0.05$ was considered to be statistically significant (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$). All data are listed in table S4.

Supplementary Materials

This PDF file includes:

Figs. S1 to S13
Tables S1 to S4

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
2. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, J. van der Oost, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
4. F. Teng, T. Cui, G. Feng, L. Guo, K. Xu, Q. Gao, T. Li, J. Li, Q. Zhou, W. Li, Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discov.* **4**, 63 (2018).
5. N. Gao, C. Zhang, Z. Hu, M. Li, J. Wei, Y. Wang, H. Liu, Characterization of *Brevibacillus laterosporus* Cas9 (BlatCas9) for mammalian genome editing. *Front. Cell Dev. Biol.* **8**, 583164 (2020).
6. Z. Hu, S. Wang, C. Zhang, N. Gao, M. Li, D. Wang, D. Wang, D. Liu, H. Liu, S. G. Ong, H. Wang, Y. Wang, A compact Cas9 ortholog from *Staphylococcus Auricularis* (SauriCas9) expands the DNA targeting scope. *PLOS Biol.* **18**, e3000686 (2020).
7. Z. Hu, C. Zhang, S. Wang, S. Gao, J. Wei, M. Li, L. Hou, H. Mao, Y. Wei, T. Qi, H. Liu, D. Liu, F. Lan, D. Lu, H. Wang, J. Li, Y. Wang, Discovery and engineering of small SlugCas9 with broad targeting range and high specificity and activity. *Nucleic Acids Res.* **49**, 4008–4019 (2021).
8. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
9. X. Xu, A. Chemparathy, L. Zeng, H. R. Kempton, S. Shang, M. Nakamura, L. S. Qi, Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* **81**, 4333–4345.e4 (2021).
10. Z. Wu, Y. Zhang, H. Yu, D. Pan, Y. Wang, Y. Wang, F. Li, C. Liu, H. Nan, W. Chen, Q. Ji, Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.* **17**, 1132–1138 (2021).
11. D. Y. Kim, J. M. Lee, S. B. Moon, H. J. Chin, S. Park, Y. Lim, D. Kim, T. Koo, J. H. Ko, Y. S. Kim, Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2022).
12. H. Morisaka, K. Yoshimi, Y. Okuzaki, P. Gee, Y. Kunihiro, E. Sonpho, H. Xu, N. Sasakawa, Y. Naito, S. Nakada, T. Yamamoto, S. Sano, A. Hotta, J. Takeda, T. Mashimo, CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.* **10**, 5302 (2019).
13. P. Cameron, M. M. Coons, S. E. Klompe, A. M. Lied, S. C. Smith, B. Vidal, P. D. Donohoue, T. Rotstein, B. W. Kohrs, D. B. Nyer, R. Kennedy, L. M. Banh, C. Williams, M. S. Toh, M. J. Irby, L. S. Edwards, C. H. Lin, A. L. G. Owen, T. Kunne, J. van der Oost, S. J. J. Brouns, E. M. Slorach, C. K. Fuller, S. Gradia, S. B. Kanner, A. P. May, S. H. Sternberg, Harnessing type I CRISPR-Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477 (2019).
14. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
15. N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, D. R. Liu, Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
16. A. V. Anzalone, P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, J. M. Levy, P. J. Chen, C. Wilson, G. A. Newby, A. Raguram, D. R. Liu, Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
17. L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, L. S. Qi, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
18. S. Morita, H. Noguchi, T. Horii, K. Nakabayashi, M. Kimura, K. Okamura, A. Sakai, H. Nakashima, K. Hata, K. Nakashima, I. Hatada, Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**, 1060–1065 (2016).
19. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, B. Huang, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
20. X. Gao, Y. Tao, V. Lamas, M. Q. Huang, W. H. Yeh, B. F. Pan, Y. J. Hu, J. H. Hu, D. B. Thompson, Y. L. Shu, Y. M. Li, H. Y. Wang, S. M. Yang, Q. B. Xu, D. B. Polley, M. C. Liberman, W. J. Kong, J. R. Holt, Z. Y. Chen, D. R. Liu, Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).
21. M. Diakatou, G. Dubois, N. Erkilic, C. Sanjurjo-Soriano, I. Meunier, V. Kalatzis, Allele-specific knockout by CRISPR/Cas to treat autosomal dominant retinitis pigmentosa caused by the G56R mutation in NR2E3. *Int. J. Mol. Sci.* **22**, (2021).
22. Y. Guo, X. M. Wei, J. Das, A. Grimson, S. M. Lipkin, A. G. Clark, H. Y. Yu, Dissecting disease inheritance modes in a three-dimensional protein network challenges the "guilt-by-association" principle. *Am. J. Hum. Genet.* **93**, 78–89 (2013).
23. D. G. Courtney, J. E. Moore, S. D. Atkinson, E. Maurizi, E. H. A. Allen, D. M. L. Pedrioli, W. H. I. McLean, M. A. Nesbit, C. B. T. Moore, CRISPR/Cas9 DNA cleavage at SNP-derived PAM enables both in vitro and in vivo KRT12 mutation-specific targeting. *Gene Ther.* **23**, 108–112 (2016).
24. P. J. Li, B. P. Kleinstiver, M. Y. Leon, M. S. Prew, D. Navarro-Gomez, S. H. Greenwald, E. A. Pipcre, J. K. Joung, Q. Liu, Allele-specific CRISPR-Cas9 genome editing of the single-base P23H mutation for rhodopsin-associated dominant retinitis pigmentosa. *CRISPR J.* **1**, 55–64 (2018).
25. K. A. Christie, D. G. Courtney, L. A. DeDionisio, C. C. Shern, S. De Majumdar, L. C. Mairs, M. A. Nesbit, C. B. T. Moore, Towards personalised allele-specific CRISPR gene editing to treat autosomal dominant disorders. *Sci. Rep.* **7**, 16174 (2017).
26. Z. Wu, H. Yang, P. Colosi, Effect of genome size on AAV vector packaging. *Mol. Ther.* **18**, 80–86 (2010).
27. F. A. Ran, L. Cong, W. X. Yan, D. A. Scott, J. S. Gootenberg, A. J. Kriz, B. Zetsche, O. Shalem, X. Wu, K. S. Makarova, E. V. Koonin, P. A. Sharp, F. Zhang, In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
28. A. Edraki, A. Mir, R. Ibraheim, I. Gainetdinov, Y. Yoon, C. Q. Song, Y. Cao, J. Gallant, W. Xue, J. A. Rivera-Perez, E. J. Sontheimer, A compact, high-accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. *Mol. Cell* **73**, 714–726.e4 (2019).
29. E. Kim, T. Koo, S. W. Park, D. Kim, K. Kim, H. Y. Cho, D. W. Song, K. J. Lee, M. H. Jung, S. Kim, J. H. Kim, J. H. Kim, J. S. Kim, In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat. Commun.* **8**, 14500 (2017).
30. P. Pausch, B. Al-Shayeb, E. Bisom-Rapp, C. A. Tsuchida, Z. Li, B. F. Cress, G. J. Knott, S. E. Jacobsen, J. F. Banfield, J. A. Doudna, CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
31. S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, J. K. Joung, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
32. T. Yamano, B. Zetsche, R. Ishitani, F. Zhang, H. Nishimasu, O. Nureki, Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1. *Mol. Cell* **67**, 633–645.e3 (2017).
33. M. F. Richter, K. T. Zhao, E. Eton, A. Lapinaite, G. A. Newby, B. W. Thuronyi, C. Wilson, L. W. Koblan, J. Zeng, D. E. Bauer, J. A. Doudna, D. R. Liu, Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
34. M. T. Nguyen Tran, M. K. N. Mohd Khalid, Q. Wang, J. K. R. Walker, G. E. Lidgerwood, K. L. Dilworth, L. Lisowski, A. Pebay, A. W. Hewitt, Engineering domain-inlaid SaCas9 adenine base editors with reduced RNA off-targets and increased on-target DNA editing. *Nat. Commun.* **11**, 4871 (2020).
35. J. R. Davis, X. Wang, I. P. Witte, T. P. Huang, J. M. Levy, A. Raguram, S. Banskota, N. G. Seidah, K. Musunuru, D. R. Liu, Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base *editors*. *Biomed. Eng.* **6**, 1272–1283 (2022).
36. X. Wang, J. Li, Y. Wang, B. Yang, J. Wei, J. Wu, R. Wang, X. Huang, J. Chen, L. Yang, Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
37. K. A. Christie, L. J. Robertson, C. Conway, K. Blighe, L. A. DeDionisio, C. Chao-Shern, A. M. Kowalczyk, J. Marshall, D. Turnbull, M. A. Nesbit, C. B. T. Moore, Mutation-independent allele-specific editing by CRISPR-Cas9, a novel approach to treat autosomal dominant disease. *Mol. Ther.* **28**, 1846–1857 (2020).
38. M. J. Landrum, J. M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, D. R. Maglott, ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).

39. S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, K. Sirotkin, dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
40. Y. Y. Tan, A. H. Y. Chu, S. Y. Bao, D. A. Hoang, F. T. Kebede, W. J. Xiong, M. F. Ji, J. H. Shi, Z. L. Zheng, Rationally engineered *Staphylococcus aureus* Cas9 nucleases with high genome-wide specificity. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20969–20976 (2019).
41. I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, F. Zhang, Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).

Acknowledgments
Funding: Yongming Wang is funded by the National Key Research and Development Program of China (2021YFA0910602 and 2021YFC2701103); the National Natural Science Foundation of China (82070258, 81870199, and 81630087); Open Research Fund of State Key Laboratory of Genetic Engineering, Fudan University (no. SKLGE-2104); and Science and Technology Research Program of Shanghai (19DZ2282100). F.L. is supported by Shenzhen Fundamental Research Program (ZDSYS20200923172000001). L.Y. is supported by the National Natural Science Foundation of China (NSFC) (31925011) and the Ministry of Science and Technology of China (MoST) (2021YFA1300503 and 2019YFA0802804). Ying Wang is funded by China Postdoctoral Science Foundation (CPSF) (2021TQ0342 and 2021 M700159) and Shanghai Post-Doctoral Excellence Program (2021435). **Author contributions:** Yao Wang, T.Q., J.L., and Miaomiao Li performed experiments. Yuan Yang, Z.W., Ying Wang, and T.W. analyzed the data. A.C.Y.C. and L.Y. revised the manuscript. Mingqing Li, D.L., S.G., F.L., and Yongming Wang provide experimental guidance. Yongming Wang wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The raw sequencing data have been submitted to the NCBI Sequence Read Archive [SRA PRJNA881616 (www.ncbi.nlm.nih.gov/bioproject/881616)]. The plasmids can be provided by Addgene pending scientific review and a completed material transfer agreement. Requests for the plasmids should be submitted to Addgene or to the corresponding authors.

Submitted 16 February 2022
Accepted 10 January 2023
Published 10 February 2023
10.1126/sciadv.abo6405

ScienceAdvances

A highly specific CRISPR-Cas12j nuclease enables allele-specific genome editing

Yao Wang, Tao Qi, Jingtong Liu, Yuan Yang, Ziwen Wang, Ying Wang, Tianyi Wang, Miaomiao Li, Mingqing Li, Daru Lu, Alex Chia Yu Chang, Li Yang, Song Gao, Yongming Wang, and Feng Lan

Sci. Adv., **9** (6), eabo6405.
DOI: 10.1126/sciadv.abo6405