JBC RESEARCH ARTICLE

Engineered FnCas12a with enhanced activity through directional evolution in human cells

Received for publication, October 22, 2020, and in revised form, February 2, 2021 Published, Papers in Press, February 7, 2021,
https://doi.org/10.1016/j.jbc.2021.100394

Xiexie Liu[1,‡], Xiaoyu Liu[1,‡], Chenchen Zhou[1], Jineng Lv[1], Xiubin He[1], Yuanyuan Liu[1], Haihua Xie[1], Bang Wang[1], Xiujuan Lv[1], Lianchao Tang[1], Mingchun Li[1], Changbao Liu[2], Junzhao Zhao[2], Yong Liu[1], Zongming Song[1,3], and Feng Gu[1,*]

From the [1]State Key Laboratory and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang, China; [2]The Second Clinical College, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China; [3]Henan Eye Hospital, Henan Eye Institute, Henan Provincial People's Hospital and People's Hospital of Zhengzhou University, Zhengzhou, Henan, China

Edited by Ronald Wek

Clustered regularly interspaced short palindromic repeat–Cas12a has been harnessed to manipulate the human genome; however, low cleavage efficiency and stringent protospacer adjacent motif hinder the use of Cas12a-based therapy and applications. Here, we have described a directional evolving and screening system in human cells to identify novel FnCas12a variants with high activity. By using this system, we identified IV-79 (enhanced activity FnCas12a, eaFnCas12a), which possessed higher DNA cleavage activity than WT FnCas12a. Furthermore, to widen the target selection spectrum, eaFnCas12a was engineered through site-directed mutagenesis. eaFnCas12a and one engineered variant (eaFnCas12a-RR), used for correcting human RS1 mutation responsible for X-linked retinoschisis, had a 3.28- to 4.04-fold improved activity compared with WT. Collectively, eaFnCas12a and its engineered variants can be used for genome-editing applications that requires high activity.

Clustered regularly interspaced short palindromic repeat (CRISPR)–Cas systems, developed from the adaptive immune system of prokaryotes, has become a powerful genome manipulation tool in biological research and has great potential for gene therapy (1, 2). Recently, Cas12a (also known as Cpf1), members of class 2 type V CRISPR–Cas family of nucleases, was reported to cleave the mammalian genome in human cells with distinctive characteristics, compared with the most commonly used SpCas9 (*Streptococcus pyogenes* Cas9) nuclease (3).

First, Cas12a cleaves the target DNA containing a T-rich 5′ protospacer adjacent motif (PAM) (2, 3), in contrast to SpCas9, which expands the selection of CRISPR-endonuclease-editable genomic sites, compared with the only choice of CRISPR-SpCas9. Second, Cas12a generates staggered double strand breaks 17 to 18 nt (nucleotides) distal from PAM, whereas Cas9 introduces blunt-ended double strand breaks 3 nt proximal to PAM (3, 4), which could expand its genome modification capabilities. Third, the RNase activity of Cas12a in processing the CRISPR array simplifies multiplex crRNA array for Cas12a-mediated human genome editing (5, 6). Furthermore, the *cis*-DNA-activated *trans*-ssDNA cutting of Cas12a expands the application of CRISPR-Cas for *in vitro* nucleotide acid detection (7, 8). Particularly, lower off-target effects of Cas12a have been observed compared with SpCas9 (9), enabling reliable applications of Cas12a-based gene therapies.

Four Cas12a orthologs, namely AsCas12a (*Acidaminococcus* sp.BV3L6), FnCas12a (*Francisella novicida* U112), LbCas12a (*Lachnospiraceae bacterium*), and MbCas12a (*Moraxella bovoculi* 237), have been developed as genome editing tools in prokaryotic, plant, animal, and human cells (10–17). Subsequent structural studies revealed the molecular mechanisms of recognition and cleavage of DNA (18, 19). However, for Cas12a-based clinical and basic applications, several problems, including increasing its cleavage efficacy or fidelity and recognizing more flexible PAMs for target selection, exist (20).

To address flexible PAM specificities, AsCas12a mutants (RR variant [S542R/K607R] and RVR variants [S542R/K548V/N552R]) have been identified, which recognize 5′-TYCV-3′ (Y: T/C; V: A/C/G) and 5′-TATV-3′ as PAM, respectively (21). A study demonstrated that these mutations could be used for engineering LbCas12a, FnCas12a, and MbCas12a (22). Specifically, the RR variant recognizes 5′-TWTV-3′ (W: A/T, V: A/C/G) as PAM, whereas 5′-TYYV-3′ (Y: C/T, V: A/C/G) serves as the PAM for the RVR variant (22).

FnCas12a recognizes 5′-KYTV-3′ (K: T/G; Y: C/T; V: A/C/G) as PAM in human cells (20), which may be further developed as an alternative tool for genome manipulation, compared with AsCas12a and LbCas12a (5′-TTTN-3′ as PAM, N: A/C/T/G). According to our previous study, FnCas12a has low activity or almost no activity at certain target sites (20). Theoretically, engineering of the coding sequence of FnCas12a may modulate the corresponding genome editing activity. Here, we sought to engineer and identify FnCas12a variants with high activity.

‡ These authors contributed equally to this work.
* For correspondence: Feng Gu, gufenguw@gmail.com.

Results

Generation of FnCas12a variants in vitro

Based on an assay (20), we generated the plasmid (pJET-U6-crRNA) and co-transfected it with FnCas12a in 293-SC1 cells (HEK-293 cells harboring the *EGFP* expression cassette), targeting *EGFP* sites A (also named with site 5, based on the location of target sequences in *EGFP*) and B (also named with site 44) (Table S1). The results revealed that we efficiently induced insertion or deletion mutations (indels) in *EGFP* (Fig. S1, *A* and *B*). Thus, the human cell-based *EGFP* reporter system may be adopted to quantitatively assess the activity of FnCas12a variants through flow cytometry (FCM, Fig. S1*B*).

Next, we generated variant libraries of FnCas12a by introducing mutations into domains, namely WED, REC, RuvC, NUC, PI, and BH, of FnCas12a through error-prone PCR (Table S2). Four unique restriction-enzyme (HindIII, ClaI, PshAI, and EcoRI) sites were found that could digest the coding sequence into three parts, namely HindIII-ClaI, ClaI-PshAI, and PshAI-EcoRI (Fig. 1*A*). Before library generation, we optimized mutagenesis conditions. We tested different amounts (1–100 ng) of template and found that with 1 ng plasmid template, we could introduce 2 to 7 mutations in an 860 bp fragment (the sequence between the restriction endonuclease recognition site of ClaI and PshAI, Fig. S2*A*, Sequence S1). Based on this optimized condition, library I (1051 colonies), library II (1119 colonies), and library III (505 colonies) were obtained (Fig. S2*B*). Subsequently, we randomly picked 20 colonies per library to access the genetic information of libraries (Fig. S2*C*). The results revealed that the mutations were evenly distributed across the coding sequence. Thus, we constructed three libraries of FnCas12a variants (2675 plasmids).

Examination of high cleavage activity of FnCas12a variants

We examined three libraries of FnCas12a variants in 293-SC1 cells and screened FnCas12a variants with high cleavage activity (Fig. 1, *B* and *C*). Specifically, the plasmids of FnCas12a

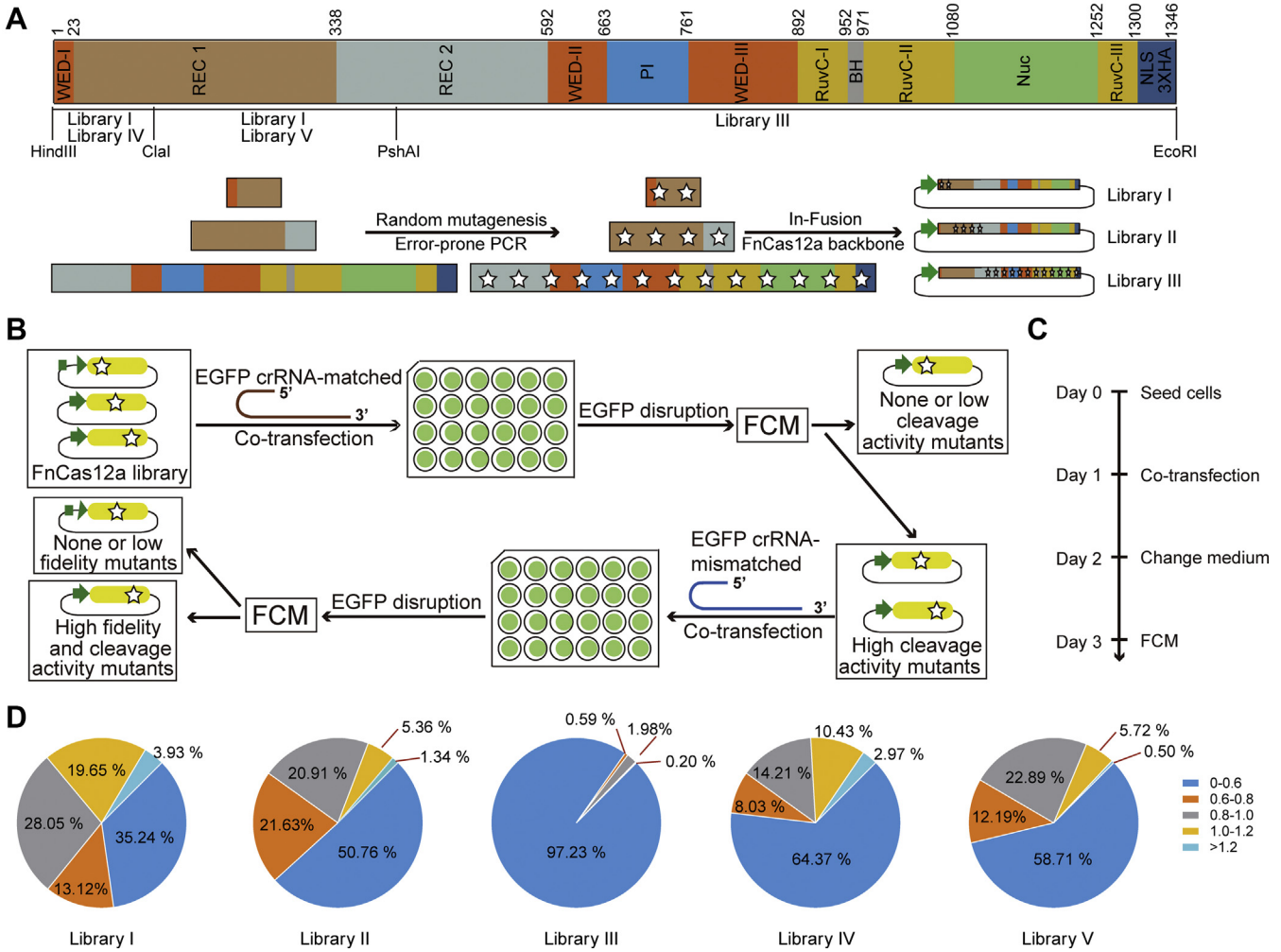


Figure 1. Generation and activities of FnCas12a variants. *A*, a schematic of FnCas12a coding sequence. HindIII-ClaI (Library I or IV), ClaI-PshAI (Library II or V), and PshAI-EcoRI (Library III). Random mutagenesis fragments are generated through error-prone PCR and inserted into the backbone. *B*, screening protocol of FnCas12a variants. Variants and matched crRNAs were co-transfected into 293-SC1 cells (HEK-293 cells expressing EGFP) and analyzed with flow cytometry (FCM). High-activity variants were tested with mismatched crRNAs and analyzed with FCM for fidelity. *C*, protocol for the EGFP disruption assay. We transfected plasmids after 24 h of seeding cells and harvested cells for FCM after 48 h of transfection. *D*, the summary of activities of FnCas12a variants. Each subgroup was divided into five parts based on relative efficiencies (ratio of the variant's EGFP disruption percentage and WT FnCas12a's), >1.2, 1.0 to 1.2, 0.8 to 1.0, 0.6 to 0.8, and 0 to 0.6. Numbers represent variants within each part and corresponding proportion. EGFP, enhanced green fluorescent protein.

variants plus crRNA targeting *EGFP* site 5 (Table S1) were transfected into the cells, and then, cleavage activity was assessed *via* FCM. We found that 430 variants (354 of library I, 75 of library II, and 1 of library III) retained 100% or higher activity (Fig. 1*D*) compared with WT FnCas12a. Then, we sought to test the fidelities of the 430 variants.

We generated single-nucleotide mismatched crRNA targeting *EGFP* sites 5 and 44 to evaluate the fidelities of the 430 variants and found that at site 5, mismatched M5-4 crRNA had comparable activity to perfect-matched crRNA, indicating high tolerance of this mismatch at the 3′ terminal of the crRNA distal to PAM for WT FnCas12a (Fig. S3, *A* and *B*, Table S3). Therefore, we selected M5-4 crRNA to screen fidelities of the 430 variants. Under the mismatched condition, we obtained 31 variants, which had no more than 80% activity, compared with WT FnCas12a (Fig. S4, *A*–*C*).

Co-transfection with perfect-matched or five single-nucleotide mismatched crRNAs (M5–1 to M5–5) revealed six variants, namely I-69, I-186, I-357, I-688, II-611, and II-717, with similar activity as WT and higher fidelity (no more than 80% activity of WT FnCas12a with single-nucleotide mismatched crRNA) than WT (Fig. S5, *A*–*E*). Further testing on site 44 (Fig. S6, *A* and *B*) revealed variants, namely I-69 and II-611, with relatively higher activities; however, variants did not exhibit a substantial increase in fidelity with different single-nucleotide mismatched crRNAs. Thus, mutants with higher activities are more likely to be identified.

Generation of library and further screening FnCas12a variants with higher activity

To boost the activity of FnCas12a, we selected these six variants as candidates (error-prone PCR templates) for second-round mutagenesis, yielding two new libraries with 1246 (library IV) and 402 variants (library V) (Fig. 1*A*). As described, we first examined their activity in an *EGFP* reporter system at site 5 (Fig. 1, *B* and *C*). Consequently, we found 39 variants from two libraries has higher activity than WT FnCas12a (Fig. 1*D*). We tested corresponding fidelities through mismatched M5-4 crRNA and identified six variants retaining no more than 80% activity of WT FnCas12a (Fig. S7, *A* and *B*). However, further testing on site 5 by co-transfecting with perfect-matched or five single-nucleotide mismatched crRNAs (M5-1 to M5-5) revealed that only variant (IV-79) had higher activity than WT. We selected this variant to perform further testing. Not surprising, we found that the IV-79 variant had increased activity at sites 5 and 44 (Fig. S7, *C*–*E*) and named it enhanced activity FnCas12a (eaFnCas12a). The other five mutants can be evaluated to identify high-fidelity variant(s).

To further compare the editing efficiency of eaFnCas12a with the five Cas12a family orthologs and variants (AsCas12a, enAsCas12a, enAsCas12a-HF, FnCas12a, and LbCas12a [enAsCas12a and enAsCas12a-HF represent enhanced AsCas12a variants and enhanced AsCas12a variants with high-fidelity, respectively]), we performed the enhanced green fluorescent protein (EGFP)-disruption assay at sites 5 and 44. The results illustrated that eaFnCas12a possesses higher activity than the Cas12a family orthologs and variants (Fig. S8). Specifically, eaFnCas12a possesses higher activity (approximately 34.88% at site 5 and 30.84% at site 44, respectively) than the four Cas12a family orthologs and variants (approximately 25.52%–29.38% at site 5 and 11.11%–28.74% at site 44, respectively) in human cells (Fig. S8).

To further validate our results, 50 target sites were selected with 5′-KYTV-3′ (K: G/T, Y: C/T, V: A/C/G) PAM in *EGFP* (Fig. 2, *A*–*C*, Tables S1 and S4). We observed that, compared with WT, eaFnCas12a exhibited an average 1.4-fold higher activity at all tested sites ($p < 0.01$) (Fig. 2, *B* and *C*). Notably, eaFnCas12a exhibited robust activity (∼20%) higher than WT FnCas12a (∼14%) at *EGFP* site 1 to 78 with 5′-YTV-3′ (Y: C/T, V: A/C/G) PAM (Fig. 2, *D* and *E*, Tables S1 and S4), suggesting that eaFnCas12a possessed higher activity for genome engineering. To further confirm it, PAM-DOSE (PAM Definition by Observable Sequence Excision) was harnessed to delineate the PAM in human cells (23). We set 1% of the total reads from NGS as functional PAM. The results showed that the PAM for eaFnCas12a is 5′-YTV-3′ (Y: C/T, V: A/C/G). We also observed that the PAM of eaFnCas12a is more flexible than WT at positions −4 (Fig. S9, *A*–*D*). Notably, for certain target sequence, eaFnCas12a as well as WT has its PAM for the recognition (Fig. S9, *A*–*D*), although there is no substantial difference.

The performance of eaFnCas12a at endogenous genomic sites in human cells

To further characterize DNA cleavage activity of eaFnCas12a in endogenous human genes, 18 endogenous genomic sites harboring 5′-YTV-3′ (Y: C/T, V: A/C/G) PAM were selected for the T7E1 assay (Fig. 3*A*, Table S5). Compared with WT, eaFnCas12a exhibited more favorable activity with an average 1.93-fold higher indels (Fig. 3, *A* and *B*), presenting similar performance evaluated in *EGFP* (Fig. 2, *B* and *E*). Notably, for *DNMT1*-2, *GSK3A*, *LRP6*, *RS1*-2, *TCF4*-1, and *TCF4*-2, the eaFnCas12a variant achieved at least 2-fold higher indel formation (Fig. 3*A*).

To confirm these results, we selected 16 endogenous sites from the above 18 endogenous genomic sites as targets for deep sequencing. Next-generation sequencing results revealed increased efficacy of eaFnCas12a-mediated genome editing at most endogenous genes (with an average 1.32-fold editing efficacy, Fig. S10, *A* and *B*), which is largely consistent with the results from the T7E1 assay. With the testing of another six sites, the results are similar (Fig. S10*A*). Furthermore, we predicted off-target sites using an online software (http://www.rgenome.net/cas-offinder/) at *CCR5*, *HBB*, and *GRIN2B* sites. Compared with WT, no substantial increase in off target effects of eaFnCas12a was observed (Fig. S10*C*). All results demonstrated that the fidelity of eaFnCas12a was similar to that of FnCas12a. Thus, eaFnCas12a is an engineered Cas12a nuclease with higher activity.

eaFnCas12a-RVR or -RR mutants recognizing more flexible PAMs

Recently, it was reported that RVR mutants (AsCas12a: S542R/K548V/N552R, LbCas12a: G532R/K538V/Y542R,

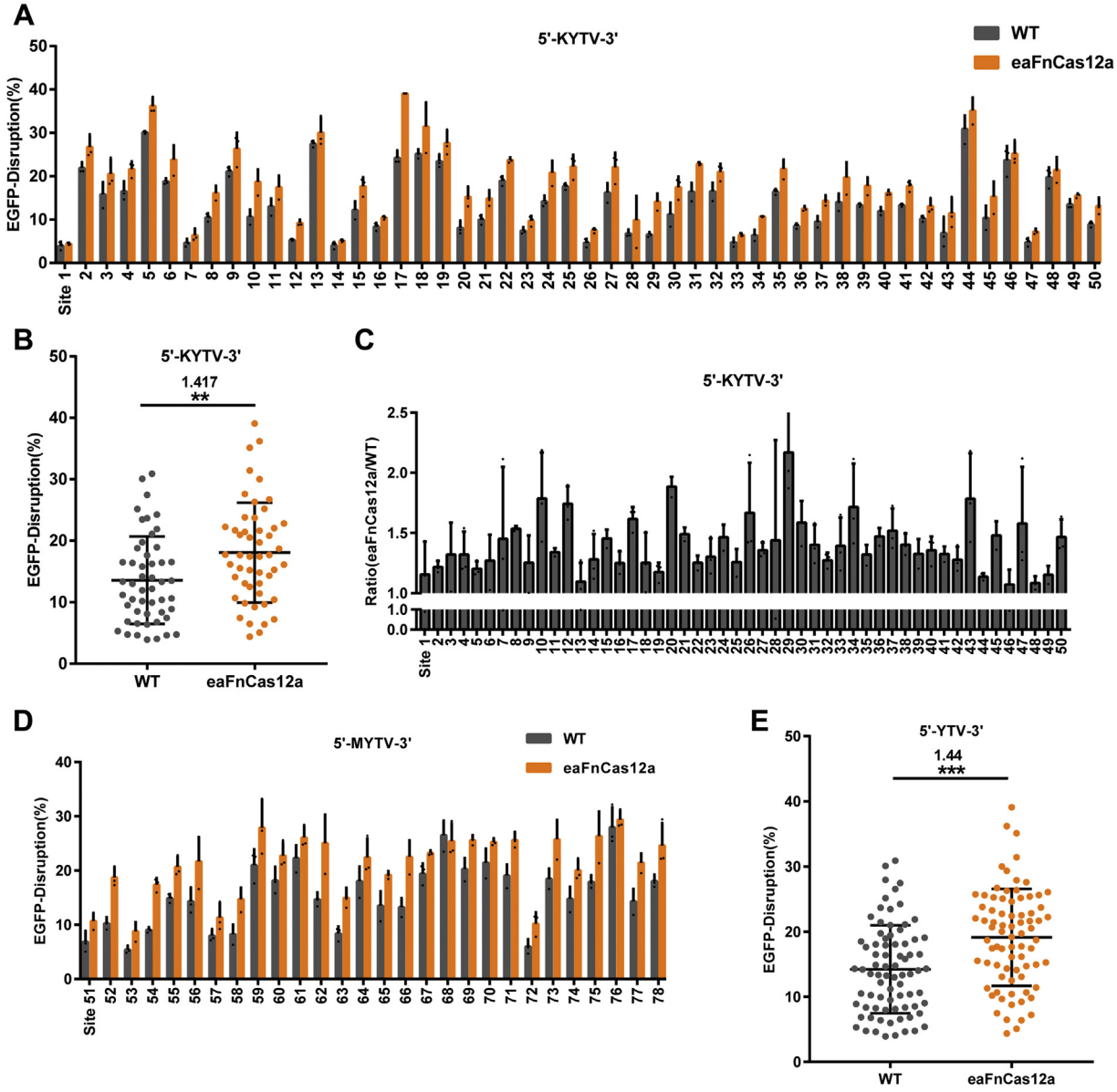


Figure 2. Examination of eaFnCas12a on 5′-KYTV-3′ and 5′- MYTV-3′ PAMs in *EGFP*. *A*, EGFP disruption of WT and eaFnCas12a on 5′-KYTV-3′ PAMs at site 1–50. *B*, Examination of eaFnCas12a at 50 *EGFP* target sites with 5′-KYTV-3′ PAMs. *C*, Relative activities of eaFnCas12a at 50 *EGFP* target sites with 5′-KYTV-3′ PAMs. *D*, EGFP disruption of WT and eaFnCas12a on 5′-MYTV-3′ PAMs at site 51–78. *E*, EGFP disruption of WT and eaFnCas12a on 5′-YTV-3′ PAMs at site 1–78. (K, Y, V, and M represent T/G, C/T, A/C/G, and A/C, respectively.) *Error bars*, SD; n = 3. $^{**}p < 0.01$. $^{***}p < 0.001$. EGFP, enhanced green fluorescent protein.

FnCas12a: N607R/K613V/N617R, MbCas12a: N576R/K582V/N586R) and RR mutants (AsCas12a: S542R/K607R, LbCas12a: G532R/K595R, FnCas12a: N607R/K671R, MbCas12a: N576R/K637R) expanded PAM preferences at targets with 5′-TATV-3′ (V: A/C/G) PAM and 5′-TYCV-3′ (Y:T/C) PAM (22). We speculated that these mutations may increase target selection of eaFnCas12a. Thus, we generated the corresponding RVR and RR mutants of eaFnCas12a (eaFnCas12a-RVR and eaFnCas12a-RR, respectively, Table S6) and compared their PAM preferences and activities on 30 target sites in *EGFP* with that in WT.

After testing with different PAMs, we observed that eaFnCas12a-RVR and eaFnCas12a-RR notably expanded PAM selection, compared with WT (Fig. S11*A*). Specifically, WT and eaFnCas12a demonstrated almost no activity with 5′-TATM-3′ (M: A/C) PAM, whereas FnCas12a-RVR exhibited considerable activity (∼24.5%) and eaFnCas12a-RVR exhibited relatively higher activity (∼27.5%) than FnCas12a-RVR (Fig. S11*A*). Next, 17 target sites with 5′-TTCR-3′ (R: A/G) PAM were selected (Tables S1 and S4), which were used in characterizing PAM preferences and activity of eaFnCas12a-RR (Fig. 4*A*, Fig. S11*B*). We observed that both eaFnCas12a and eaFnCas12a-RR exhibited significantly higher activity than WT ($p < 0.0001$) or FnCas12a-RR ($p < 0.01$), respectively (Fig. 4*A*). Moreover, we tested activity of eaFnCas12a-RR at 11 target sites with 5′-TCCV-3′ (V: A/C/G) PAM (Tables S1 and S4) and found that eaFnCas12a had higher efficacy than WT ($p < 0.01$) and eaFnCas12a-RR had the best performance (Fig. S11, *C* and *D*). Thus, we confirmed

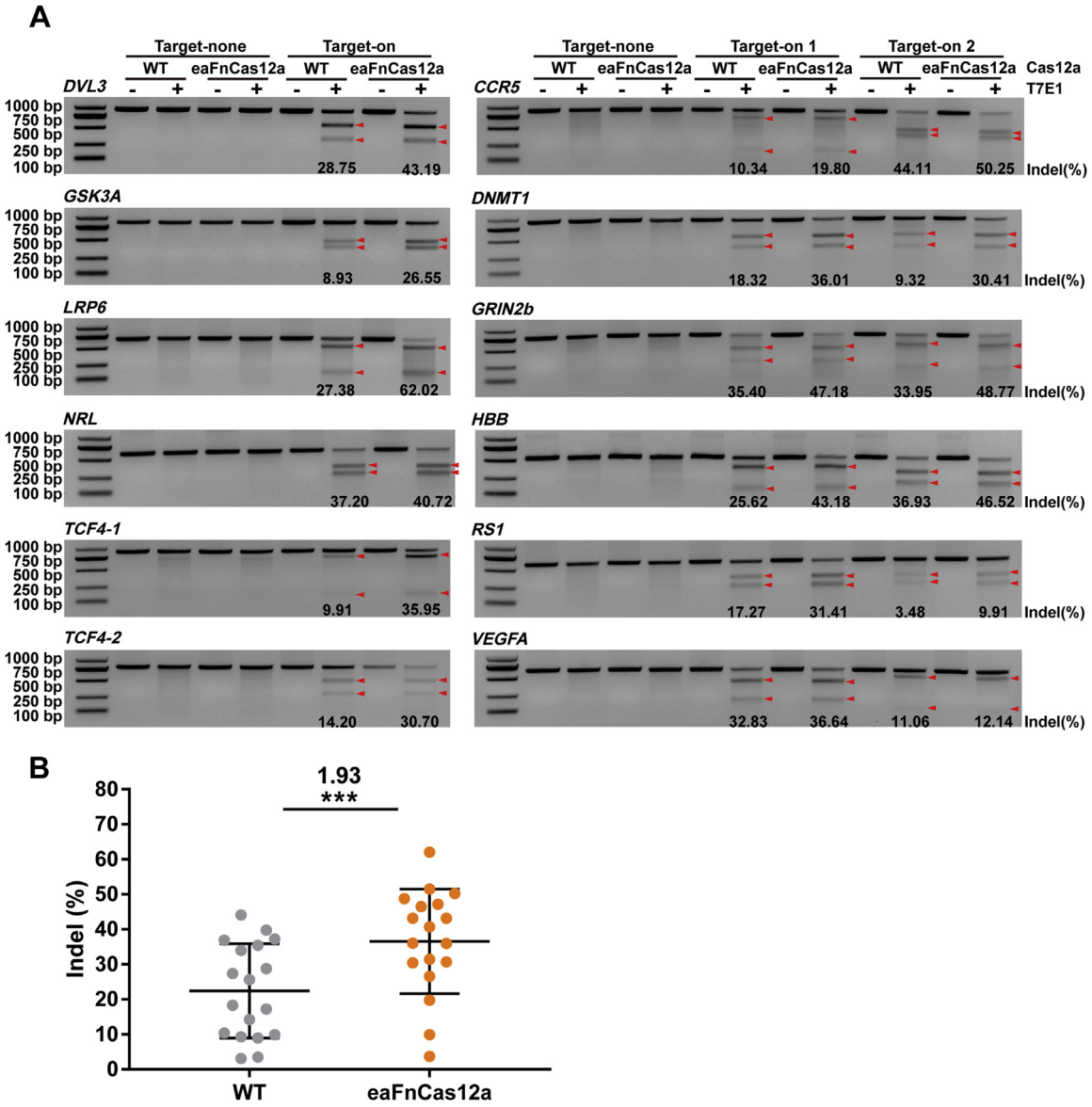


Figure 3. The activity of eaFnCas12a at endogenous sites. *A*, gel image of T7E1 assay at 18 endogenous target sites. Fragments cut by T7E1 are highlighted with *red arrows* and indel percentage is mentioned under each lane. *B*, dot plot of indels triggered with eaFnCas12a or WT at 18 endogenous target sites. Error bars, SD; ***$p < 0.001$.

that the mutation combination with eaFnCas12a-RVR or eaFnCas12a-RR may boost its target selection.

Application of variants-mediated genome editing for correcting a pathogenic mutation

X-linked juvenile retinoschisis is the most common form of macular dystrophy in young men, with no effective treatment (24). Genome editing–mediated *in situ* gene correction holds great promise for its clinical treatment. To investigate gene correction mediated by eaFnCas12a, we used the 293-RS1 cell line, which harbors the partial coding region of *RS1* carrying the causative mutation (*RS1*-p.Y65X) from a patient with X-linked juvenile retinoschisis (Fig. 4*B*, Fig. S12*A*) (25). In this system, EGFP is activated when nonhomologous end joining (3n bp indels) or homology directed repair (HDR, choosing WT donor DNA oligo as template) is triggered by Cas12a at m*RS1*-site 1 or 2 (Fig. 4*B*, Tables S1 and S4). Consistent with PAMs recognized by eaFnCas12a, m*RS1*-sites 1 and 2 contain 5′-TTV-3′ (V: A/C/G) and 5′-TTCV-3′ PAMs, respectively. On day 5 after transfection, EGFP expression was the highest (Fig. 4*C*). eaFnCas12a had the best performance with or without the donor (Fig. S12, *B–D*). We observed a 3.28- and 4.04-fold increase in activity of the variant at m*RS1*-sites 1 and 2, respectively, compared with WT. (Fig. 4*D*). Thus, we demonstrated that gene correction efficacy can be highly increased with eaFnCas12a.

Identification of key residue(s) for eaFnCas12a

We sequenced eaFnCas12a and found that it contained one amino acid substitution (position 125, from Gln [Q] to Arg [R]). We then evaluated the reason for the presence of

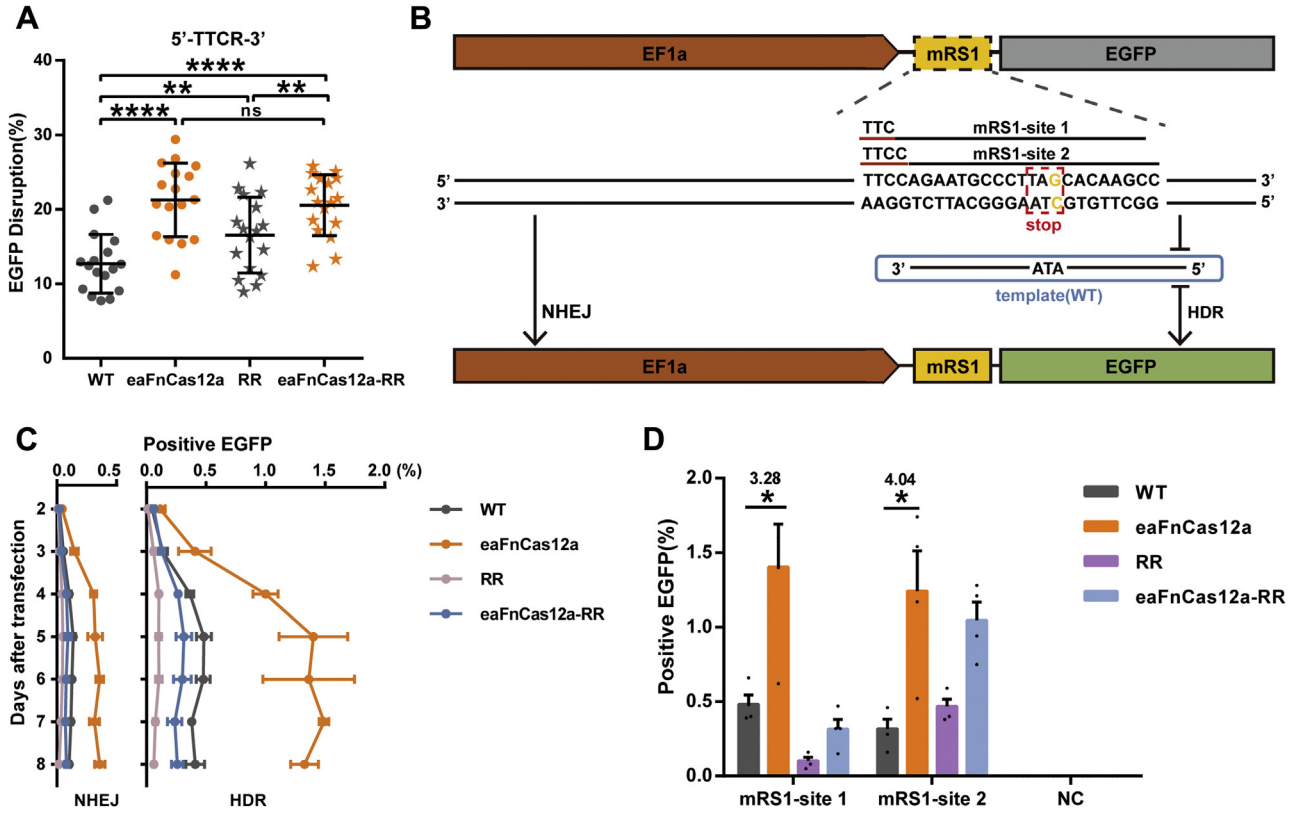


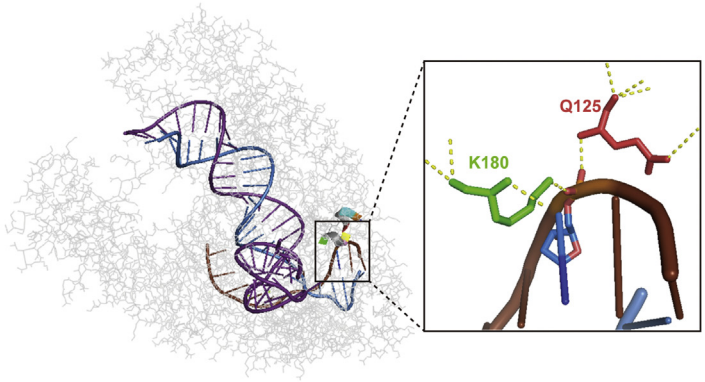
Figure 4. Correction of *RS1* mutation through eaFnCas12a and its variant. *A*, EGFP disruption triggered with WT-FnCas12a, eaFnCas12a, FnCas12a-RR (RR group), and eaFnCas12a-RR. *B*, a schematic of mRS1 (harboring TAG, stop codon) restoration by NHEJ (no template) or HDR (with ssDNA template). *C*, a correction comparison of WT-FnCas12a, eaFnCas12a, FnCas12a-RR (RR group), and eaFnCas12a-RR at mRS1-site 1. Positive EGFP is showed on different days (y-axis) after transfection. *D*, a correction comparison of WT-FnCas12a, eaFnCas12a, FnCas12a-RR (RR group), and eaFnCas12a-RR at mRS1-sites 1 and 2 at day 5 after transfection. NC represents cells without transfection. Error bars, SD; n = 4. *$p < 0.05$, **$p < 0.01$, ****$p < 0.0001$. EGFP, enhanced green fluorescent protein; HDR, homology directed repair; NHEJ, nonhomologous end joining; ns, not significant.

only a single mutation. We obtained mutant sequences mutants from libraries and found that Q125R may be a novel mutation generated during the second-round mutagenesis (Fig. S13). Surprisingly, no tagged mutation of eaFnCas12a was derived from the first-round mutagenesis. We speculated that the tagged mutation was corrected during the generation of libraries IV and V. To further understand the improved activity of eaFnCas12a caused by Q125R, we analyzed the crystal structure of FnCas12a in complex with crRNA and target DNA (PDB: 5MGA, Fig. 5*A*). Q125 locates at the N-terminal of the REC1 domain, which participates in PAM recognition together with WED-II, WED-III, and PI domains. Briefly, the α-amines of Q125 directly contacts to the phosphate backbone of DNA, and the α-carbonyl of Q125 directly contact I128, D129, and D130 through an intrachain hydrogen bond, which does not improve activity. However, homology-modeling (SWISS-MODEL webtool) suggested that the Q125R substitution caused a loss of interaction between δ-amines of Gln125 and γ-carbonyl of Asp129 (Fig. 5*B*). *In silico* stimulation using the CUPSAT software revealed that Q125R possessed the highest stability among all predicted substitutions (2.23 kcal/mol) (Fig. 5*C*). Thus, we proposed that the Q125R substitution may increase the stability and positive charge of FnCas12a to enable extensive contacts between FnCas12a and negatively charged DNA substrates. Then, we systemically examined the effects of additional substitutions of Q125 on activity (Fig. 5*D*, Table S2). When Q125 was changed to D or E, a notable decrease in activity was observed because of the negative charge of these residues. Q125P led to local structural distortions of FnCas12a and thus compromised its catalytic activity. Q125K possessed increased activity at sites 5, 19, and 38 compared with WT FnCas12a, which may be because of the strong positive charge of K (Fig. 5*C*). Further studies will be required to ascertain the mechanisms by which eaFnCas12a achieves high activity.

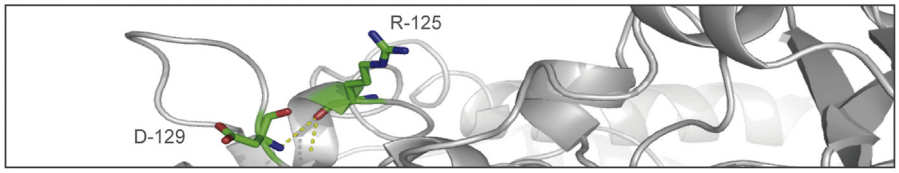
Engineering to improve eaFnCas12a and crRNA

Based on the structure of FnCas12a (26), we sought to identify mutants with higher activity. Therefore, we generated and tested additional mutants, namely N124R, I128R, D129R, A130R, T176R, T177R, K180R, N666R, T696R, K699R, and G701R (Fig. S14, *A–D*, Table S7). However, these mutants did not exhibit increased activity, except for A130R at site 44, compared with eaFnCas12a (Fig. S14*C*). Then, we generated mutants harboring Q125R in addition to these mutations; however, no boost in activity was observed (Fig. S14, *A–D*), which indicated the absence of synergic effects of Q125R with tested mutations. Thus, Q125R may play a crucial role in boosting activity.

A



B



C

Amino acid	Overall Stability	Torsion*	Predicted ΔΔG (kcal/mol)
Arg (R)	Stabilising	Favourable	2.23
Lys (K)	Stabilising	Favourable	0.21
Ala (A)	Stabilising	Favourable	0.65
Cys (C)	Stabilising	Unfavourable	5.06
His (H)	Stabilising	Unfavourable	1.85
Gly (G)	Stabilising	Unfavourable	0.33
Ser (S)	Stabilising	Unfavourable	1.21
Tyr (Y)	Stabilising	Unfavourable	2.01
Asn (N)	Stabilising	Unfavourable	0.63
Glu (E)	Stabilising	Favourable	0.28
Pro (P)	Destabilising	Unfavourable	-0.04
Asp (D)	Destabilising	Unfavourable	-1.03

D

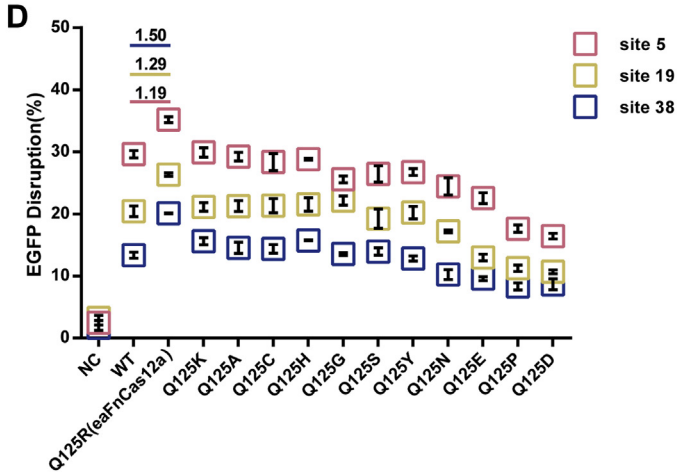


Figure 5. Structural insights of eaFnCas12a. *A*, structural context of FnCas12a at position 125. *Left panel*, overall structure of a ternary complex formed by FnCas12a (*gray*), target DNA (*blue*: target strand, *purple*: untarget strand). *Right panel*, Zoomed-in view of the *boxed region* in the *left panel*, key residues of FnCas12a (Q125 and K180 are highlighted in *red* and *green*), and target DNA. Phosphate backbone of DNA is in *chocolate*. Hydrogen bonds are indicated by *yellow dotted lines*. (PDB:5MGA). *B*, structural context of eaFnCas12a at position 125. Q125R substitution results in the loss of interaction between δ-amines of Q125 and γ-carbonyl of D129. R125 and D129 are highlighted in *green*. *C*, *in silico* predication of stability and potential energy at the Q125 substitution. *D*, variant activity with residue substitution at position 125 at EGFP sites (5, 19, and 38). EGFP, enhanced green fluorescent protein.

Engineered FnCas12a with enhanced activity

According to the previous report, Cas12a activity can be enhanced through crRNA engineering (27). We engineered the crRNA sequence and investigated whether it could boost the activity of eaFnCas12a at seven target *EGFP* sites. Our results revealed that, compared with WT, eaFnCas12a has higher editing efficacy with the engineered crRNA sequence (U4A) in crRNA. No synergistic effect of FnCas12a and engineered crRNA was found (Fig. S15).

Discussion

The Cas12a-family is a complementary nuclease family with low off-target effects and distinct PAMs compared with the Cas9-family. Although the nucleases AsCas12a and LbCas12a are commonly used for gene editing, FnCas12a can be used in plants and humans (16, 28). However, FnCas12a has relatively low cleavage activity (20, 29). Studies have used various strategies to increase the efficiency of Cas12a, including using the optimized crRNA scaffolds (30, 31), using 3′U-rich crRNA (27), and extending the 5′ end of crRNA (32). The structure-guided mutagenesis screening strategy has been adopted to expand targeting range, enhance activities, and increase fidelity of AsCas12a and LbCas12a (21, 33). Notably, the structure-guided mutagenesis approach is useful in generating special mutants, which minimizes the screening effort. However, reports demonstrating directional evolution of Cas12a with random mutagenesis to identify mutant(s) with desired traits in human cells do not exist. Random mutagenesis is beneficial, especially in cases of limited structural data and structure-activity correlations. The limitation of this approach includes requirement of larger libraries to identify variants, which may expand screening efforts and time for the target (29). In this study, we identified a novel FnCas12a variant (eaFnCas12a) with remarkably higher activity than WT FnCas12a through directional evolution.

To increase the compatibility of eaFnCas12a, we combined RR or RVR mutants to eaFnCas12a (eaFnCas12a-RR or eaFnCas12a-RVR). Notably, eaFnCas12a-RR performed better than the FnCas12a-RR variant both on 5′-TTCR-3′ (R: A/G) and 5′-TCCV-3′ (V: A/C/G) PAMs; eaFnCas12a performed the best of all on 5′-TTCR-3′ PAMs (Fig. 4*A*). On 5′-TATM-3′ (M: A/C) PAMs, eaFnCas12a-RVR also outperformed the RVR variant (Fig. S11*A*), suggesting that the Q125R mutation is synergic with RR or RVR mutants. To further assess eaFnCas12a, we targeted *RS1*, which contains a pathogenic nonsense mutation. Similar results were obtained—eaFnCas12a performed the best, and eaFnCas12a-RR outperformed the FnCas12a-RR variant (Fig. 4*C*). Thus, our results demonstrated the compatibility of Q125R with RR or RVR mutants.

We evaluated whether Q125R could be used for AsCas12a, LbCas12a, or MbCas12a engineering. We aligned amino acid sequences from these four nucleases and found that Q125 is not highly conserved among them (Fig. S16). Notably, residues (K, A), which were the exact residues in AsCas12a or LbCas12a and MbCas12a, were the top substitutions in stability analysis (Fig. 5). We speculated that to increase its activity, a substitution with R(Arg) may be beneficial for engineering AsCas12a or MbCas12a and LbCas12a.

Recently, FnCas12a-based base editing (BE) was realized at T-rich PAM sequence with high efficacy, overcoming the limitation of Cas9-based BE at G-rich PAM sequences (33). Prime editing (PE) is an exciting step toward more versatile genome editing; whereas limited target selection may remain a bottle neck for SpCas9-based PE application (34, 35). Thus far, we do not know the performance of the dead eaFnCas12a for BE and PE. Our results may be used for developing next-generation genome editing tools. Taken together, we identified eaFnCas12a, a high-activity variant, through directional evolution in human cells. eaFnCas12a and its engineered variants could be used for genome editing applications that require high activity with more flexible PAM recognition.

Experimental procedures

Plasmids, cells, and cell culture

Plasmids for the expression of FnCas12a were obtained from Addgene (Addgene plasmids #69976). The crRNA expression cassette was generated by insertion of DNA annealing products into vector pJET1.2 (CloneJET PCR Cloning Kit, Thermo Fisher Scientific). The *EGFP* and m*RS1* gene target sequences used are summarized in Table S1. The oligonucleotide sequences used for introducing FnCas12a mutations are summarized in Tables S2 and S7. The oligonucleotide sequences used for crRNA construction are summarized in Tables S3 and S4. Plasmid DNA and genomic DNA were isolated by standard techniques. DNA sequencing confirmed the identity of the desired specific sequence in the constructs.

HEK-293 cells were obtained from ATCC (CAT#CRL-1573) and grown at 37 °C in 5% CO_2 in Dulbecco's modified Eagle's medium (Life Technologies), 10% heat-inactivated fetal bovine serum, 100 U/ml PS (Penicillin/Streptomycin). HEK-293 cells expressing GFP (named 293-SC1) were described previously (20, 36). HEK-293 cells expressing partial RS1 coding region (named 293-RS1) were generated by transduction with lentivirus at serial dilution and selection with puromycin (1 mg/ml) until all cells in control dishes had detached (6–8 days). Drug-resistant single colonies were isolated. To maintain RS1 expression, the medium for 293-RS1 culture included puromycin. The detailed sequence information has been described previously (37).

Generation of FnCas12a variants libraries

The FnCas12a libraries were generated by random mutagenesis by PCR. Specifically, plasmid (pY004) harboring FnCas12a coding sequence was digested with HindIII/ClaI, ClaI/PshAI or PshAI/EcoRI, respectively. The HindIII/ClaI, ClaI/PshAI, or PshAI/EcoRI fragments were mutated by random mutagenesis kit (Agilent, Cat#200550) and then purified, in-fusion with the pY004 backbone without the corresponding fragment. The individual colonies from LB plate were then manually picked. The plasmids from individual colonies were isolated. Several single and double mutants (with N125R) were designed and assembled according to the structure of FnCas12a (26).

Images and flow cytometry analysis

The FCM protocol was described previously. Briefly, on day 0, 0.9 × 10^5 293-SC1 cells/well were seeded in 24-well plates and transfected with 375 ng Cas12a and 125 ng crRNA expression plasmids by the TurboFect Transfection Reagent (Thermo Scientific) on day 1. Fresh medium was added to the transfected 293-SC1 cells on day 2. Cells were imaged or harvested for FCM or genomic DNA isolation on day 3.

Nuclease assay (T7E1) and next-generation sequencing

For T7 Endonuclease I assay, 250 ng purified PCR products (using primer sets listed in Table S5) were mixed with 1 μl 10× NEB buffer 2 and ultrapure water to a final volume of 9.5 μl and were subjected to re-annealing process to enable heteroduplex formation. After re-annealing, products were treated with T7 Endonuclease I at 37 °C for 2 h, and quantification was based on relative band intensities. Indel percentage was determined by the formula, 100 × (1− sqrt (b + c)/(a + b+ c)), where a is the integrated intensity of the undigested PCR product; b and c are the integrated intensities of the cleavage product.

As to the next-generation sequencing, amplicons were generated using two rounds of PCR (using primer sets listed in Tables S5, S8 and S9). Equal amounts of the PCR amplicons were subjected to paired-end read sequencing using the Hiseq 1500 platform (Illumina) at Novogene. Adaptors and low-quality reads were removed from the resulting 150-bp paired-end reads using Trimmomatic (version 0.36). Reads were then mapped to the template using Bowtie2 (version 2.3.3). All NGS-data obtained with this method in this study. The oligonucleotide sequences used for T7E1 and NGS are summarized in Tables S7–S9.

RS1 mutation correction via HDR

For HDR assay, briefly, 1.8 × 10^5 293-RS1 cells/well were seeded in 12-well plates on day 1 and transfected with 375 ng FnCas12a, 125 ng crRNA expression plasmids, and 2 pmol WT donor DNA oligotemplate (Table S8) by transfection reagent (TurboFect, Thermo Scientific) on day 2. Fresh medium was added on day 3. Cells were harvested for FCM on day 4 to 8, respectively.

Statistics

All data were expressed as mean ± SEM. Differences were determined by 2-tailed Student's *t* test or Mann-Whitney test between two groups. The criterion for statistical significance was $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

Data availability

The data that support the findings of this study are available in the Supplementary Materials. The deep sequencing data are available at the NCBI Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/) under BioProject PRJNA613361 (SRA: SRR11398536-SRR11398541; sample accession number, SAMN13944353); BioProject PRJNA689427 (SRA: SRR13358344; sample accession number, SAMN13944353). Preprocessed data are available upon request.

Supporting information—This article contains supporting information.

Acknowledgments—We thank Fayu Yang, Fanfan Li and Huihui Sun for technical assistance and comments from Prof. Yanli Wang (Institute of Biophysics, Chinese Academy of Sciences).

Author contributions—F.G. conceived the idea, XX. L., XY. L., C. Z., J. L., X. H., Y. L., H. X., and X. L. performed the experiments, and B. W., L. T., M. L., C. L., J. Z., Y. L., and F. G. performed data analyses, and F. G. wrote the manuscript. All authors have read and approved the final manuscript.

Funding and additional information—This work was supported by grants from National Key R&D Program of China (2018YFA0107304), National Natural Science Foundation of China (81201181 and 81670882), Science and Technology Project of Zhejiang Province (2017C37176), Natural Science Foundation of Zhejiang Provincial (LQ18C060003), Zhejiang Provincial &Ministry of Health research fund for medical sciences (WKJ-ZJ-1828), Wenzhou City Key Innovation Team of Reproductive Genetics Grant, Zhejiang, China (No. C20170007), Lin He's Academician Workstation of New Medicine and Clinical Translation (17331209 and 18331105) and Project of State Key Laboratory of Ophthalmology, Optometry and Visual Science, Wenzhou Medical University (J02-20190201).

Conflict of interests—The authors declare that they have no conflict of interest with the contents of this article.

Abbreviations—The abbreviations used are: BE, base editing; CRISPR, clustered regularly interspaced short palindromic repeat; EGFP, enhanced green fluorescent protein; FCM, flow cytometry; HDR, homology directed repair; PAM, protospacer adjacent motif; PE, prime editing.

References

1. Chen, X., and Gonçalves, M. A. F. V. (2018) DNA, RNA, and protein tools for editing the genetic information in human cells. *iScience* **6**, 247–263
2. Komor, A. C., Badran, A. H., and Liu, D. R. (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36
3. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771
4. Jinek, M., Chylinski, K., Fonfara, I., Hauer, H., Doudna, J., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821
5. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521
6. Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., Wu, W. Y., Scott, D. A., Severinov, K., van der Oost, J., and Zhang, F. (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34
7. Li, S. Y., Cheng, Q. X., Liu, J. K., Nie, X. Q., Zhao, G. P., and Wang, J. (2018) CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491–493

8. Chen, J., Ma, E., Harrington, L., Costa, M., Tian, X., Palefsky, J., and Doudna, J. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439
9. Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S. H., and Kim, J. S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868
10. Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., Sun, B., Chen, B., Xu, X., Li, Y., Wang, R., and Yang, S. (2017) CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum. *Nat. Commun.* **8**, 15179
11. Tang, X., Lowder, L. G., Zhang, T., Malzahn, A. A., Zheng, X., Voytas, D. F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., Kirkland, E. R., Zhang, Y., and Qi, Y. (2017) A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants* **3**, 17018
12. Kim, H., Kim, S. T., Ryu, J., Kang, B. C., Kim, J. S., and Kim, S. G. (2017) CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* **8**, 14406
13. Kim, Y., Cheong, S. A., Lee, J. G., Lee, S. W., Lee, M. S., Baek, I. J., and Sung, Y. H. (2016) Generation of knockout mice by Cpf1-mediated gene targeting. *Nat. Biotechnol.* **34**, 808–810
14. Hur, J. K., Kim, K., Been, K. W., Baek, G., Ye, S., Hur, J. W., Ryu, S. M., Lee, Y. S., and Kim, J. S. (2016) Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat. Biotechnol.* **34**, 807–808
15. Zhang, Y., Long, C., Li, H., McAnally, J. R., Baskin, K. K., Shelton, J. M., Bassel-Duby, R., and Olson, E. N. (2017) CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. *Sci. Adv.* **3**, e1602814
16. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N., and van der Oost, J. (2018) Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651
17. Zhu, H. C., Li, C., and Gao, C. X. (2020) Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* **21**, 661–677
18. Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., Ma, C., Wang, S., Wu, D., Ma, Y., Fan, S., *et al.* (2016) The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526
19. Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. V., Ishitani, R., Zhang, F., and Nureki, O. (2016) Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962
20. Tu, M., Lin, L., Cheng, Y., He, X., Sun, H., Xie, H., Fu, J., Liu, C., Li, J., Chen, D., Xi, H., Xue, D., Liu, Q., Zhao, J., Gao, C., *et al.* (2017) A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic Acids Res.* **45**, 11295–11304
21. Gao, L., Cox, D. B. T., Yan, W. X., Manteiga, J. C., Schneider, M. W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017) Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792
22. Toth, E., Czene, B. C., Kulcsar, P. I., Krausz, S. L., Talas, A., Nyeste, A., Varga, E., Huszar, K., Weinhardt, N., Ligeti, Z., Borsy, A. E., Fodor, E., and Welker, E. (2018) 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
23. Tang, L., Yang, F., He, X., Xie, H., Liu, X., Fu, J., Xi, H., Lu, X., Liu, C., Song, Z., Qu, J., Zhao, J., and Gu, F. (2019) Efficient cleavage resolves PAM preferences of CRISPR-Cas in human cells. *Cell Regen.* **8**, 44–50
24. Ambrosio, L., Hansen, R. M., Kimia, R., and Fulton, A. B. (2019) Retinal function in X-linked juvenile retinoschisis. *Invest. Ophthalmol. Vis. Sci.* **60**, 4872–4881
25. Chen, D., Xu, T., Tu, M., Xu, J., Zhou, C., Cheng, L., Yang, R., Yang, T., Zheng, W., He, X., Deng, R., Ge, X., Li, J., Song, Z., Zhao, J., *et al.* (2017) Recapitulating X-linked juvenile retinoschisis in mouse model by knock-in patient-specific novel mutation. *Front. Mol. Neurosci.* **10**, 453
26. Stella, S., Alcon, P., and Montoya, G. (2017) Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563
27. Bin Moon, S., Lee, J. M., Kang, J. G., Lee, N. E., Ha, D. I., Kim, D. Y., Kim, S. H., Yoo, K., Kim, D., Ko, J. H., and Kim, Y. S. (2018) Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3'-overhang. *Nat. Commun.* **9**, 3651
28. Schindele, P., Wolter, F., and Puchta, H. (2018) Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13. *FEBS Lett.* **592**, 1954–1967
29. Xie, H., Ge, X., Yang, F., Wang, B., Li, S., Duan, J., Lv, X., Cheng, C., Song, Z., Liu, C., Zhao, J., Zhang, Y., Wu, J., Gao, C., Zhang, J., *et al.* (2020) High-fidelity SaCas9 identified by directional screening in human cells. *PLoS Biol.* **18**, e3000747
30. Teng, F., Li, J., Cui, T., Xu, K., Guo, L., Gao, Q., Feng, G., Chen, C., Han, D., Zhou, Q., and Li, W. (2019) Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. *Genome Biol.* **20**, 15
31. Lin, L., He, X., Zhao, T., Gu, L., Liu, Y., Liu, X., Liu, H., Yang, F., Tu, M., Tang, L., Ge, X., Liu, C., Zhao, J., Song, Z., Qu, J., *et al.* (2018) Engineering the direct repeat sequence of crRNA for optimization of FnCpf1-mediated genome editing in human cells. *Mol. Ther.* **26**, 2650–2657
32. Park, H. M., Liu, H., Wu, J., Chong, A., Mackley, V., Fellmann, C., Rao, A., Jiang, F., Chu, H., Murthy, N., and Lee, K. (2018) Extension of the crRNA enhances Cpf1 gene editing *in vitro* and *in vivo*. *Nat. Commun.* **9**, 3313
33. Kleinstiver, B. P., Sousa, A. A., Walton, R. T., Tak, Y. E., Hsu, J. Y., Clement, K., Welch, M. M., Horng, J. E., Malagon-Lopez, J., Scarfo, I., Maus, M. V., Pinello, L., Aryee, M. J., and Joung, J. K. (2019) Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282
34. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., and Liu, D. R. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157
35. Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., Wang, Y., Anzalone, A. V., Raguram, A., Doman, J. L., Liu, D. R., and Gao, C. (2020) Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585
36. Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., Jin, Z. B., Qu, J., and Gu, F. (2014) Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci. Rep.* **4**, 5405
37. Yang, F. Y., Ge, X. L., He, X. B., Liu, X. X., Zhou, C. C., Sun, H. H., Zhang, J. S., Zhao, J. Z., Song, Z. M., Qu, J., Liu, C. B., and Gu, F. (2018) Functional non-homologous end joining patterns triggered by CRISPR/Cas9 in human cells. *J. Genet. Genomics* **45**, 329–332