Article

www.the-innovation.org

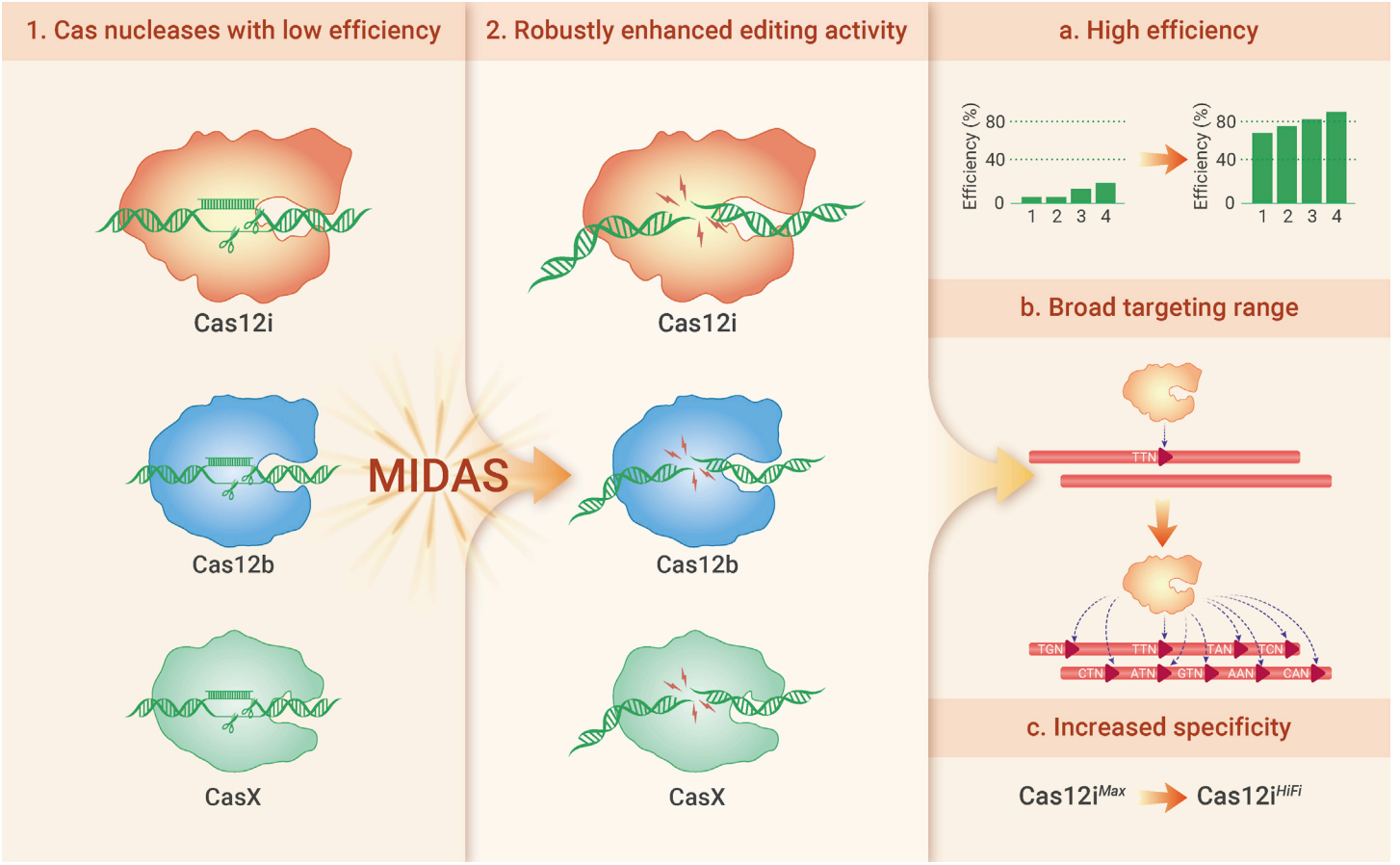
Synergistic engineering of CRISPR-Cas nucleases enables robust mammalian genome editing

Yangcan Chen,[1,2,3,5] Yanping Hu,[1,2,3,5] Xinge Wang,[1,2,3,5] Shengqiu Luo,[1,2,3] Ning Yang,[1,2,3] Yi Chen,[1,2,3] Zhikun Li,[1,3,4] Qi Zhou,[1,2,3,4,*] and Wei Li[1,2,3,4,*]

*Correspondence: zhouqi@ioz.ac.cn (Q.Z.); liwei@ioz.ac.cn (W.L.)

Received: January 12, 2022; Accepted: May 23, 2022; Published Online: May 26, 2022; https://doi.org/10.1016/j.xinn.2022.100264

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Improving Editing Activity by Synergistic Engineering (MIDAS) of Cas nucleases
- MIDAS can improve the activity of Cas12i, Cas12b, and CasX
- Engineering high-efficiency Cas12i^{Max} and high-specificity Cas12i^{HiFi}

Article

Synergistic engineering of CRISPR-Cas nucleases enables robust mammalian genome editing

Yangcan Chen,[1,2,3,5] Yanping Hu,[1,2,3,5] Xinge Wang,[1,2,3,5] Shengqiu Luo,[1,2,3] Ning Yang,[1,2,3] Yi Chen,[1,2,3] Zhikun Li,[1,3,4] Qi Zhou,[1,2,3,4,*] and Wei Li[1,2,3,4,*]

[1]State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Institute for Stem Cell and Regenerative Medicine, Chinese Academy of Sciences, Beijing 100101, China

[4]Bejing Institute for Stem Cell and Regenerative Medicine, Beijing 100101, China

[5]These authors contributed equally

*Correspondence: zhouqi@ioz.ac.cn (Q.Z.); liwei@ioz.ac.cn (W.L.)

Received: January 12, 2022; Accepted: May 23, 2022; Published Online: May 26, 2022; https://doi.org/10.1016/j.xinn.2022.100264

Citation: Chen Y., Hu Y., Wang X., et al., (2022). Synergistic engineering of CRISPR-Cas nucleases enables robust mammalian genome editing. The Innovation **3(4)**, 100264.

The naturally occurring prokaryotic CRISPR-Cas systems provide valuable resources for the development of new genome-editing tools. However, the majority of prokaryotic Cas nucleases exhibit poor editing efficiency in mammalian cells, which significantly limits their utility. Here, we have developed a method termed Improving Editing Activity by Synergistic Engineering (MIDAS). This method exerts a synergistic effect to improve mammalian genome-editing efficiency of a wide range of CRISPR-Cas systems by enhancing the interactions between Cas nuclease with the protospacer adjacent motif (PAM) and the single-stranded DNA (ssDNA) substrate in the catalytic pocket simultaneously. MIDAS robustly and significantly increased the gene-editing efficiency of Cas12i, Cas12b, and CasX in human cells. Notably, a Cas12i variant, Cas12i^{Max}, exhibited robust activity with a very broad PAM range (NTNN, NNTN, NAAN, and NCAN) and higher efficiency than the current widely used Cas nucleases. A high-fidelity version of Cas12i^{Max} (Cas12i^{HiFi}) has been further engineered to minimize off-target effects. Our work provides an expandable and efficacious method for engineering Cas nucleases for robust mammalian genome editing.

INTRODUCTION

CRISPR-Cas systems are continually being discovered in large numbers and phylogenetic diversities[1] (Figure 1A), providing valuable resources with which to develop a variety of versatile gene-editing tools. Some of the newly discovered Cas nucleases exhibit specific features, such as smaller protein size or simpler protospacer adjacent motif (PAM) requirements,[2–4] permitting efficient *in vivo* gene delivery by adeno-associated virus (AAV) or an expanded genome-targeting scope. However, most of the newly discovered CRISPR-Cas systems show poor gene-editing activity in mammalian cells,[2,4] thereby limiting their widespread use. One solution is to characterize more orthologs within a CRISPR-Cas system and seek to find one with high gene-editing activity in mammalian cells.[5,6] However, this strategy is largely dependent on chance and can be ineffective when the number of discovered orthologs is limited.[3,4] Another important solution is to optimize the activity of existing CRIPSR-Cas nucleases. Among these, rational protein engineering, which utilizes the information inherent in a protein structure, is of great importance and power. Previous studies have so far only focused on engineering one local part of the Cas protein,[7,8] which may not achieve maximum activity optimization considering the multifunctionalities of Cas nucleases. In this study, we have developed a method, termed Improving Editing Activity by Synergistic Engineering (MIDAS) (Figure 1B), to synergistically improve the CRISPR-Cas editing activity in mammalian cells by simultaneously enhancing the interactions of Cas proteins with PAM and with single-stranded DNA (ssDNA) substrate in the catalytic pocket. MIDAS-engineered variants of Cas12i, Cas12b, and CasX showed significantly increased editing efficiencies up to 137-, 11-, and 465-fold compared with wild-type enzymes, respectively. Importantly, a Cas12i variant, Cas12i^{Max}, exhibits robust activity with a very broad PAM range (NTNN, NNTN, NAAN, and NCAN) and higher efficiency than current widely used Cas nucleases. We further engineered a high-fidelity version of Cas12i^{Max} to minimize off-target effects for applications that need high specificity.

RESULTS

Rationale for MIDAS

CRISPR-Cas nuclease-based DNA editing generally includes two processes: recognition and then hydrolysis of DNA.[9] The former ensures the binding of a Cas nuclease to the correct PAM duplexes and the formation of a cleavage-ready complex, while the latter allows cleavage of DNA substrates encased in the catalytic pocket[9] (Figure 1B). The ability to accomplish these two intrinsic steps determines the catalytic activity of a given Cas nuclease. However, relative to microbial cells, the more complex cellular environments of mammalian cells may make it harder to cleave DNA.[10,11] To that end, we developed MIDAS to concurrently enhance the interactions between Cas nucleases and the PAM duplex (EIP), and the interactions between the catalytic pocket and ssDNA substrate (EIS), aiming for facilitating both DNA recognition and hydrolysis (Figure 1B). Given that these two processes are both required and reciprocally affect each other, we anticipated that MIDAS would dramatically enhance gene-editing activity in mammalian cells via a synergistic effect (Figure 1B).

MIDAS enhances the nuclease activity of Cas12i in human cells

Cas12i is a recently discovered CRISPR-Cas system from metagenomes with a promising potential due to its small protein size, simple CRISPR RNA (crRNA), and PAM requirements as well as its pre-crRNA processing capacity.[3] We first evaluated the efficiency of the wild-type Cas12i2 nuclease to cleave DNA in human cells. The Cas12i2 protein flanked by two SV40 nuclear localization signals (NLSs) was expressed from a CAG promoter, while its cognate mature crRNA[12] with a 20 nucleotide (nt) spacer was expressed by a U6 polymerase III promoter. The two vectors were transfected into human HEK293T cells. Subsequently, 72 h post-transfection, EGFP-positive cells were collected by fluorescence-activated cell sorting (FACS), and high-throughput sequencing was used to measure the gene-editing efficiency (Figure S1A). SpCas9 targeting the same *CCR5* site was used as a reference. We found that the insertion and deletion (indel) frequency of Cas12i2 averaged from eight different sites was only 6.9%, while SpCas9 reached over 50% (Figure S1B).

We then used MIDAS to engineer CRISPR-Cas12i variants. At the EIP step, two amino acids of Cas12i2, Q163 and N164, whose side chains stack on the last base pair of the PAM duplex (Figure 2A), were first engineered to facilitate the opening of the downstream double-stranded DNA (dsDNA).[12] While mutating both amino acids into an alanine to disrupt the stacking interactions impairs the catalytic activity of Cas12i2,[12] here, we substituted them with residues carrying bigger, flat aromatic rings (e.g., phenylalanine [F], tyrosine [Y], or tryptophan [W]) to reinforce the stacking interaction with the last base pair of the PAM duplex. The editing efficiencies of the engineered Cas12i2 variants and the wild-type enzyme at two gene loci (*CCR5* and *RNF2*) were evaluated in the human HEK293T cells. Notably, variants Q163W, N164Y, and N164F showed robustly increased gene-editing efficiency at both sites (Figure 2B). Next, we substituted the amino acids surrounding the PAM duplex with positively charged arginine to increase their electrostatic attraction to the negatively charged phosphate backbone (Figure 2C). A similar strategy was used for energy compensation in previous studies during the engineering of Cas nuclease with a broadened PAM recognition profile.[8,13,14] In our study, 10 variants were tested at two human genomic sites, and four of these variants (E176R, K238R, T447R, and E563R) exhibited increased indel frequency (Figure 2D). New variants harboring combinations of these four mutations had higher gene-editing activities (Figure 2E).

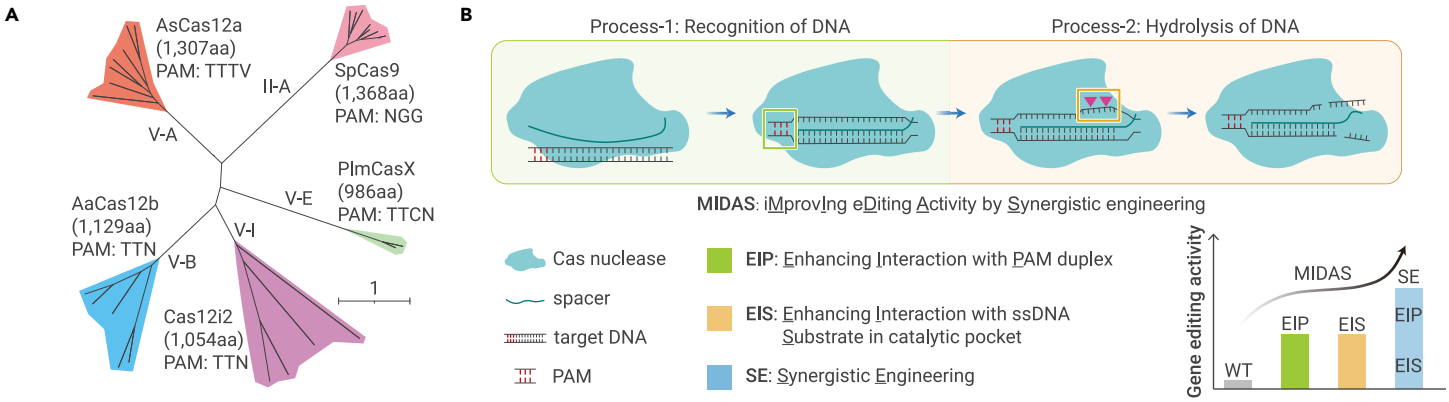


Figure 1. Developing MIDAS to enhance the mammalian genome-editing activity of Cas nucleases (A) Phylogenetic analysis of type II and V CRISPR-Cas systems. The protein size and PAM requirement of each type's representative are given. (B) Rationale and workflow of MIDAS. Specifically, in the EIP step, amino acids stacking with the last PAM base pair were mutated into W, Y, or F and/or amino acids surrounding the PAM duplex were mutated into R. In the EIS step, amino acids surrounding ssDNA substrate were mutated into R. In the last SE step, the modification from EIP and EIS were combined into one protein. R, W, Y, and F represent arginine, tryptophan, tyrosine, and phenylalanine, respectively.

Subsequently, we combined amino-acid alterations of variant E176R/K238R/T447R/E563R and variant N164Y into one Cas12i protein. When tested with another three crRNAs, a further additive effect was observed at two gene loci (*CD34* and *RNF2*) (Figure 2F). Taking all of this data together, we used variant N164Y/E176R/K238R/T447R/E563R as the final candidate from the EIP step optimization.

For the EIS step efficiency optimization, we mutated the non-catalytic residues in the catalytic pocket into arginine to increase their binding energy to the ssDNA substrate, aiming to substantially enhance enzymatic catalysis.[7,15] First, residues around the ssDNA substrate in the catalytic pocket of Cas12i2 were candidates for mutagenesis (Figure 3A). Following screening of 14 variants, I926R showed the highest gene-editing activity (Figure 3B). To search for more gain-of-function amino-acid mutations, we superimposed a 9 nt ssDNA substrate from Cas12i1 into the catalytic pocket of Cas12i2[16] (Figures 3C and 3D). Of note, centered on the new longer ssDNA substrate (Figure 3E), we obtained several new variants with increased activities; among them, E323R and D362R had the highest editing efficiencies (Figure 3F). Furthermore, additive effects were observed after combining E323R, D362R, and/or I926R (Figures 3G and 3H). Based on the above data, we chose variants I926R, E323R/D362R, and E323R/D362R/I926R as candidates from the EIS step optimization.

We then implemented the last SE step of MIDAS by combining the optimized mutations from the EIP and EIS steps. Robust and significant synergistic effects were observed when we combined N164Y/E176R/K238R/T447R/E563R with I926R, E323R/D362R, or E323R/D362R/I926R (Figure 4A). The additive effects were most obvious at the *EMX1* locus. Variants from the SE step exhibited 78- to 137-fold boosted gene-editing activity compared with the wild-type nuclease, with editing efficiency up to 52%, while all other variants and wild type performed poorly at this site, with efficiency less than 15% (Figure 4A). Taken together, these results suggest that our MIDAS protein-engineering strategy is very effective at turning a low-activity wild-type type V-I nuclease into a robust molecular scissor in human cells.

Expandability of MIDAS in engineering other classes of the Cas nucleases

Considering the generally similar catalytic mechanism of different single-effector Cas nucleases,[16] we next investigated the expandability of MIDAS for optimizing other classes of the Cas nucleases. First, we applied MIDAS to engineer AaCas12b, a type V-B CRISPR-Cas system (Figure 1A). Cas12b was characterized as having relatively small protein size and high specificity.[7,17] Cas12b may be a promising gene-editing tool if its editing efficiency in mammalian cells can be further improved.[17] We employed the published structure of AacCas12b[18] (Figure S2A) as a guide for implementing MIDAS to engineer AaCas12b. Following the similar workflow used for Cas12i2, we obtained variants that showed enhanced gene-editing activity in the EIP step (Figures 4B and S2B–S2F) and the EIS step (Figures 4B, S2G, and S2H). Significantly, synergistic improvement of gene editing was observed from a combination of mutants (Figure 4B), as observed earlier with the Cas12i2 nuclease (Figure 4A).

Similarly, we applied MIDAS to engineer PlmCasX from the type V-E CRISPR-Cas system[2] (Figure 1A). Here, we used the published structure of DpbCasX[2] (Figure S3A) to guide the PlmCasX engineering. Again, MIDAS significantly enhanced the genome-editing activity of PlmCasX in human cells (Figures S3B–S3F and 4C). Most impressively, whereas the editing efficiency of the wild-type PlmCasX at the three tested genomic sites was nearly 0%, the MIDAS-engineered variant, T26R/K610R/K808R, exhibited a 119- to 465-fold increase in gene-editing activity compared with the wild-type nuclease, with editing efficiency reaching up to 52% (Figure 4C). Collectively, these results showed that MIDAS could be applied to improve the gene-editing efficiency of Cas nucleases from diverse CRISPR-Cas systems.

Efficacy of MIDAS in engineering Cas nucleases

Next, we investigated in depth of the efficacy of MIDAS in engineering Cas nucleases. MIDAS-engineered Cas12i variant N164Y/E178R/K238R/T447R/E563R/E323R/D362R, which was named Cas12i^{Max}, was chosen to test (Figure 4A). First, we assessed the gene-editing activity of Cas12i^{Max} at 62 randomly chosen sites from five different genes, covering all of the possible canonical NTTN PAMs.[3] Notably, Cas12i^{Max} displayed highly robust genome-editing activity at all of the sites tested, with the average editing efficiency reaching up to 73% (Figure 5A). Importantly, the indel frequency of Cas12i^{Max} at 56 sites surpassed 60%, indicating that the high gene-editing activity of Cas12i^{Max} would be less targeted-locus dependent (Figure 5A). Next, we explored the activity of Cas12i^{Max} at targeted sites carrying non-NTTN PAMs, which are not amenable for cleavage by wild-type Cas12i2.[12] Here, 60 targeted sites bearing NTVN, NVTN, and NVVN PAMs (V = C, G, or A) were randomly chosen for testing. Impressively, Cas12i^{Max} potently edits sites bearing NTVN, NVTN, NAAN, and NCAN PAMs (V = C or A), with an average editing efficiency over 40% (Figure 5B). To confirm the robustness of Cas12i^{Max} in recognizing a broad range of PAMs, wild-type Cas12i2 and Cas12i^{Max} proteins were purified (Figure S4A), and their cleavage activities were compared at targets carrying the same protospacer and different NNNN PAMs *in vitro*. Remarkably, Cas12i^{Max} nearly fully digested dsDNA with NTTN, NTAN, NTCN, NATN, NAAN, NCTN, NCAN, and NGTN PAMs and was efficient at NTGN, NACN, and NGAN, while wild-type Cas12i2 only partially cleaved dsDNA with an NTTN PAM (Figures 5C and 5B). Encouraged by these results, we systematically characterized the PAM recognition profile of Cas12i^{Max} using a previously reported assay called PAM-DOSE[19] with some modifications. Briefly, DNA fragments containing recognizable PAMs of Cas12i^{Max} were collected by gel purification for high-throughput-sequencing analysis (Figure S4B). Of note, only subtle base preferences at -3, -2, and -1 positions of 5′-NNNN PAM were observed (Figure 5E). Further analyzing the frequency of each NNNN demonstrated that NGGN, NCGN, and NGCN were the only ones with minor proportions, indicating that Cas12i^{Max} exhibits very simple PAM requirements (Figure 5F). This result is consistent with our observation that Cas12i^{Max} enables dsDNA cleavage at all of the NNNN PAMs except at NGGN, NCGN, and NGCN (Figures 5A–5D). Collectively, these data demonstrate that the MIDAS empowers the Cas nuclease with highly potent gene-editing activity both at canonical and non-canonical PAMs.

To further evaluate the efficacy of MIDAS, we compared the gene-editing activity of Cas12i^{Max} with current widely used Cas nucleases, while the PAMs of the

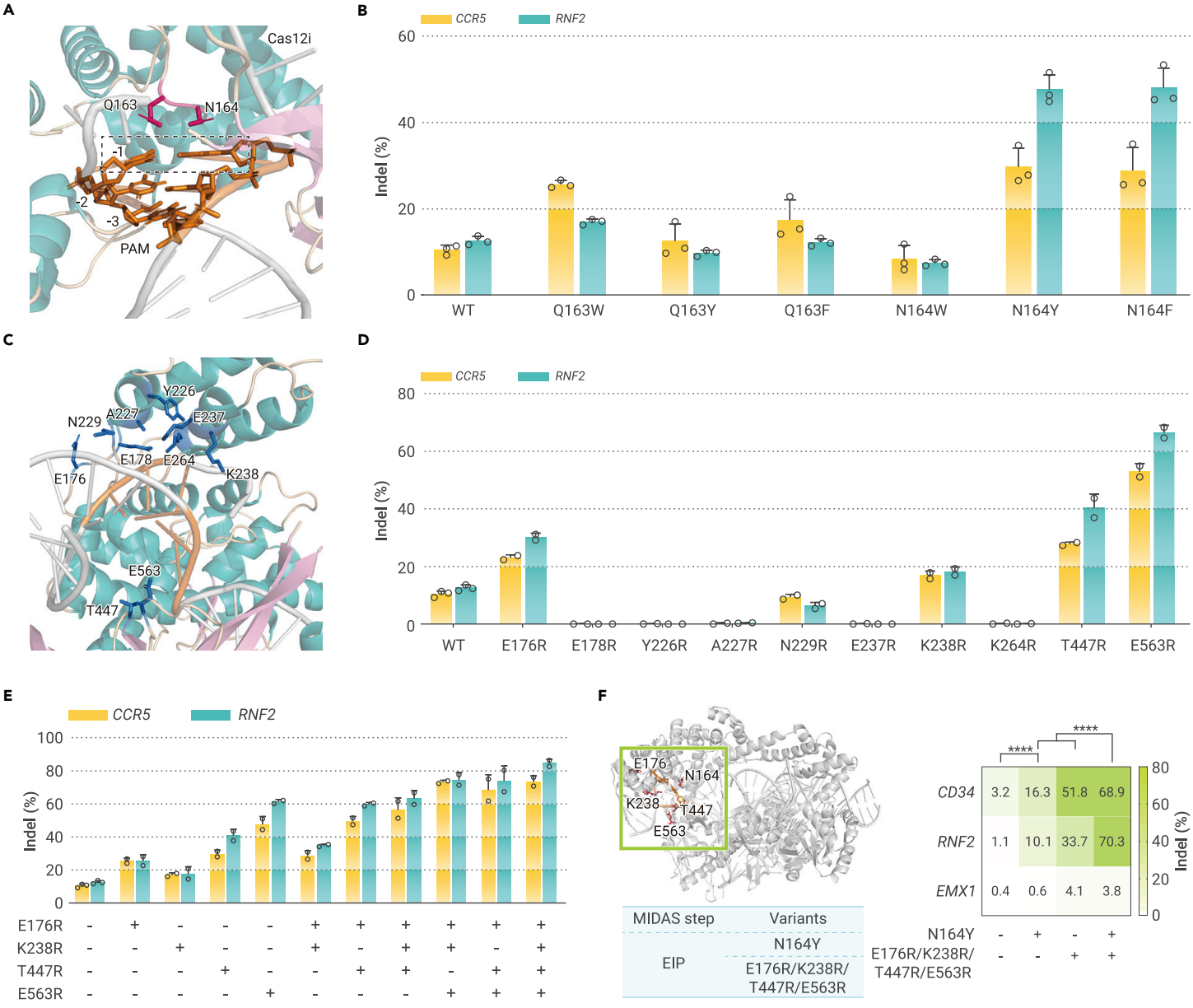


Figure 2. Implementing the EIP step of MIDAS to engineer Cas12i (A) Amino-acid stacking on the last base pair of PAM of Cas12i2 is indicated in hot pink. The last base pair of PAM was annotated as -1 (when the 20 bp targeted dsDNA was counted from +1 to +20) and highlighted by a dashed box. Orange, PAM duplex. (B) Mutating Q163 and N164 into W, Y, or F for increased gene-editing efficiency. Error bars represent standard deviation (SD) of mean, n = 3 biological replicates. (C) Amino acids chosen for arginine substitution are marine blue. Orange, PAM duplex. (D) Testing the genome-editing activity of variants at two targeted sites. Error bars represent SD of mean, n = 2 or 3 biological replicates. (E) The gene-editing activity of new Cas12i2 variants with combinations of the E176R, K238R, T447R, or E563R mutations. Error bars represent SD of mean, n = 2 or 3 biological replicates. (F) The PAM duplex of Cas12i2 is indicated in orange. Variants with robustly improved genome-editing activity are listed in the table. Corresponding mutated amino acids are indicated in red. The number in each cell of the heatmap represents the mean indel frequency averaged from n = 3 biological replicates. Two-way ANOVA test, **** $p < 0.0001$.

tested targets were most amenable to these nucleases. Impressively, at identical genomic sites bearing TTTV PAMs, Cas12i^{Max} exhibited higher average indel frequency than AsCas12a[5] (Figure 5G). Additionally, it significantly outperformed BhCas12b v4 at ATTN PAMs[7] (Figure 5G). Next, we compared Cas12i^{Max} with Cas9.[6,20–23] Using similar spacers with overlaps greater than 18 nt, Cas12i^{Max} showed higher average genome-editing efficiency than SpCas9, SaCas9, and SaCas9-KKH (Figure 5G). Taken together, these data show that MIDAS-engineered Cas12i^{Max} exhibited higher genome-editing efficiency than widely used genome editors.

The advantageous features of Cas12i^{Max}, including smaller protein size, high editing efficiency, and capability of pre-crRNA processing, make it highly suitable for multiplexed gene editing *in vivo*. To explore this potential, we tested the editing efficiency of AAV expression plasmids containing Cas12i^{Max} and crRNA targeting single genes (*Pcsk9*, *Angptl3*, and *Apoc3*) or one pre-crRNA array simultaneously targeting the three genes in mouse-derived Hepa1-6 cells (Figure 5H). Of note, Cas12i^{Max} robustly edited the three genes with a single pre-crRNA array at a comparable or higher indel level as a single crRNA (Figure 5H).

An additional amino-acid mutation improves the specificity of Cas12i^{Max}

Considering the high gene-editing activity and broad PAM recognition profile of Cas12i^{Max}, it is particularly important to evaluate the specificity of this MIDAS-engineered system. Initially, we determined the degree of Cas12i^{Max} to tolerate mismatch(es) between spacer and target DNA. At *RNF2* endogenous locus, some degree of tolerance was observed at crRNAs with single or double mismatches (Figure S5A). These results indicated that Cas12i^{Max} might have off-target effects when targeting genes. Indeed, GUIDE-seq analysis[24] showed that Cas12i^{Max} has relatively more off-target sites compared with other widely applied Cas nucleases (Figure 6C). Analysis of the off-target profiles demonstrated that these off-target effects are attributed to simple PAM requirements and some degree of mismatch tolerance of Cas12i^{Max} (Figure S6). In order to increase the specificity of Cas12i^{Max}, we mutated amino acids with positive charge (arginine or lysine), which surrounds the RNA-DNA heteroduplex formed by spacer RNA and DNA target,[12] into alanine to attenuate the non-specific electrostatic interaction,[25,26] aiming to reduce the mismatch tolerance of Cas12i^{Max}. After screening via crRNA targeting mCherry or an endogenous gene, a new variant of Cas12i^{Max}, with only

Figure 3. Implementing the EIS step of MIDAS to engineer Cas12i (A) Amino acids chosen for arginine substitution around the 5 nt ssDNA substrate are indicated in marine blue. Orange, PAM duplex; green, catalytic amino acids (D599, E833, D1017). (B) Testing the genome-editing activity of variants at two targeted sites. Error bars represent SD of mean, n = 2 or 3 biological replicates. (C) Protein BLAST result for Cas12i2 and Cas12i1. (D) Superimposition of the 3D structures of Cas12i2 and Cas12i1. Orange, Cas12i2; teal, Cas12i1. (E) Amino acids chosen for arginine substitution around the 9 nt ssDNA substrate aligned from Cas12i1 are indicated in red. Orange, PAM duplex; green, catalytic amino acids. (F) Testing the genome-editing activity of variants at two targeted sites. Error bars represent SD of mean, n = 2 or 3 biological replicates. (G) The gene-editing activity of new Cas12i2 variants with a combination of the E323R and D362R mutations at two targeted sites. Error bars represent SD of mean, n = 2 or 3 biological replicates. (H) The 5 nt ssDNA substrate of Cas12i2 is indicated in orange, and the 9 nt ssDNA substrate from Cas12i1 superimposed on Cas12i2 is indicated in red. Variants with robustly improved genome-editing activity are listed in the table. Corresponding mutated amino acids are highlighted in marine blue. The number in each cell of the heatmap represents the mean indel frequency from n = 3 biological replicates. Two-way ANOVA test, ****$p < 0.0001$.

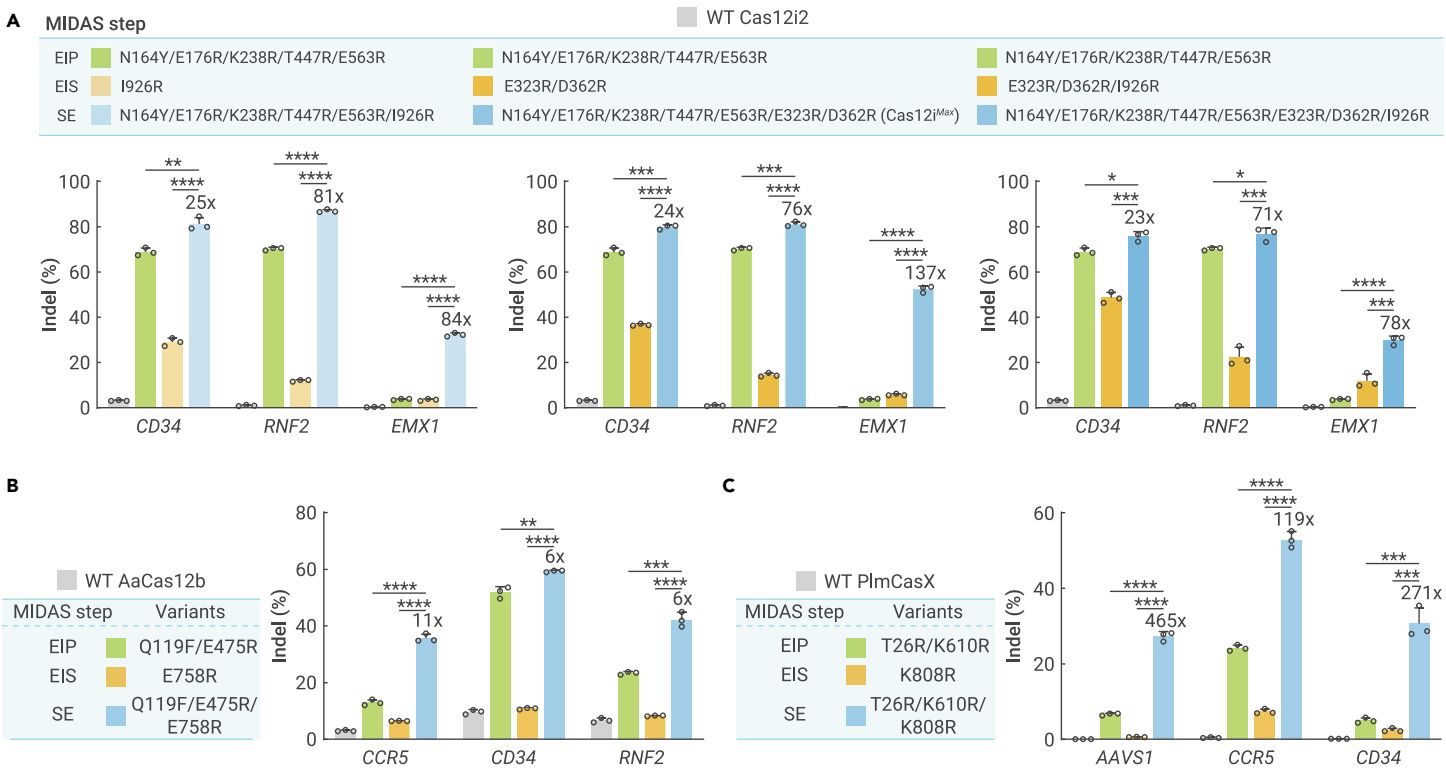


Figure 4. MIDAS enables various Cas nucleases to robustly edit the human genome (A) Activity-enhanced amino-acid mutations from EIP and EIS were combined into one protein to create the final Cas12i2 variants, which are listed in the table. Error bars represent SD of mean, n = 3 biological replicates. Fold changes of editing activity of variants from SE step compared with wild-type Cas12i2 at each site are exhibited. Two-tailed, unpaired Student's t test, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. (B) Engineering Cas12b using MIDAS. Activity-enhanced variants from each step are listed in the table. Error bars represent SD of mean, n = 3 biological replicates. Fold changes of editing activity of variants from SE step compared with wild-type AaCas12b at each site are exhibited. Two-tailed, unpaired Student's t test, **p < 0.01, ***p < 0.001, and ****p < 0.0001. (C) Engineered CasX variants with enhanced activity from each step are listed in the table. Error bars represent SD of mean, n = 3 biological replicates. Fold changes of editing activity of variants from SE step compared with wild-type PlmCasX at each site are exhibited. Two-tailed, unpaired Student's t test, ***p < 0.001 and ****p < 0.0001.

an additional single mutation, K394A, exhibited largely decreased mismatch tolerance and maintained most on-target activity of Cas12i^{Max} (Figure S5B, S5C, 6A, and 6B). We designated this high-fidelity version of Cas12i^{Max} as Cas12i^{HiFi} and further evaluated its specificity using GUIDE-seq. Impressively, no off-target sites for Cas12i^{HiFi} were detected on a genome-wide scale (Figure 6C; Figure S6). Collectively, these results demonstrate that the fidelity of Cas12i^{Max} can be further engineered to generate Cas12i^{HiFi}, which offers minimized off-target effects for applications that require high specificity.

DISCUSSION

In this study, we developed a powerful Cas protein-engineering method named MIDAS, which can significantly improve CRISPR-Cas mammalian genome-editing efficiency by increasing the affinity of Cas proteins with PAM and with ssDNA substrate in the catalytic pocket. Using MIDAS, we optimized Cas nucleases from diverse CRISPR branches (Figure 1A). Of note, significant and robust synergistical effects on increasing gene-editing efficiency were observed in all cases in our study when we combined gain-of-function mutations from the EIP and EIS optimization steps into a single Cas nuclease. This is particularly obvious for Cas nuclease when targeting endogenous genomic loci, which are not amenable for variants solely from EIP or EIS (Figures 4A–4C). These results suggest that when engineering enzymes with complex functionalities like Cas nucleases, the multi-step processes of enzymic catalysis should be considered holistically to achieve maximum activity optimization.

Like other rational protein-engineering approaches, MIDAS requires the three-dimensional (3D) structural information of Cas nucleases for design guidance. Of the large numbers of the CRISPR-Cas systems discovered in recent years, only a small number of Cas nucleases with the 3D structures were reported. Even so, our study demonstrates that structural information of the orthologue can also be used as guidance, even with amino-acid-sequence homology as low as 27%, such as Cas12i2 versus Cas12i1 (Figure 3C). It should be mentioned that the information of ssDNA substrates encased in the catalytic pocket might be missed in the reported 3D structures of some Cas nucleases, which could impact the design in the EIS step optimization of MIDAS engineering. Our study provided a potential solution to fill in the missing information by superimposing the ssDNA substrate from other Cas proteins. Even for CRISPR-Cas subtypes that lack any reported 3D structures, *in silico* molecular modeling can also be applied to find the conserved interface of Cas protein with PAM or/and the ssDNA substrate.[27] With the increasing availability of Cas 3D structural information[12,28] and the variety of the powerful computational tools, MIDAS engineering has great potential to be applied to the optimization of a large numbers of Cas proteins.

In addition to its great expandability, the efficacy of MIDAS in engineering Cas nuclease is also impressive. Exemplarily, one of the MIDAS-engineered nucleases, Cas12i^{Max}, exhibited potent genome-editing activity at canonical NTTN PAMs with higher indel frequency than current widely used Cas genome editors like AsCas12a, BhCas12b, SpCas9, SaCas9, and SaCas9-KKH. Importantly, the editing efficiency of Cas12i^{Max} at 56/62 sites exceeded 60%, which indicates that the potent editing activity of Cas12i^{Max} in human cells might be less targeted-locus dependent, allowing editing of more genomic loci. A further study is required to understand the relationship between the activity profiles of Cas12i^{Max} and the target sequence compositions in the mammalian cells. In addition, MIDAS also enables Cas12i to robustly edit at sites carrying non-NTTN PAMs. The results from high-throughput characterization, *in vitro* cleavage assay, and *in cellulo* validation in multiple targeted sites demonstrated that Cas12i^{Max} has a very broad PAM targeting range (NTNN, NNTN, NAAN, and NCAN). This expanded targeting flexibility is likely due to the amino-acid substitutions from the EIP step optimization, which created new nonspecific interactions between the Cas protein and PAM duplex.

Its robust gene-editing efficiency, along with the ability to process pre-crRNA, makes Cas12i^{Max} highly suitable for multiplexed gene editing *in vivo*. To test this potential, cassettes expressing Cas12i^{Max} and pre-RNA were cloned into a single AAV expression plasmid. The size of the entire construct was smaller than 4.7 kilobase pairs (kb), which can be efficiently packaged into one AAV. Using pre-crRNA to simultaneously edit multiple genes in the mammalian cells, we observed that Cas12i^{Max} achieves a comparable or higher indel as using a single crRNA separately. Moreover, we systematically analyzed the targeting scope of Cas12i^{Max} in the human genome. Remarkably, Cas12i^{Max} can potentially target 70.1% of the entire human genome (Figure S7A), which vastly outnumbers other

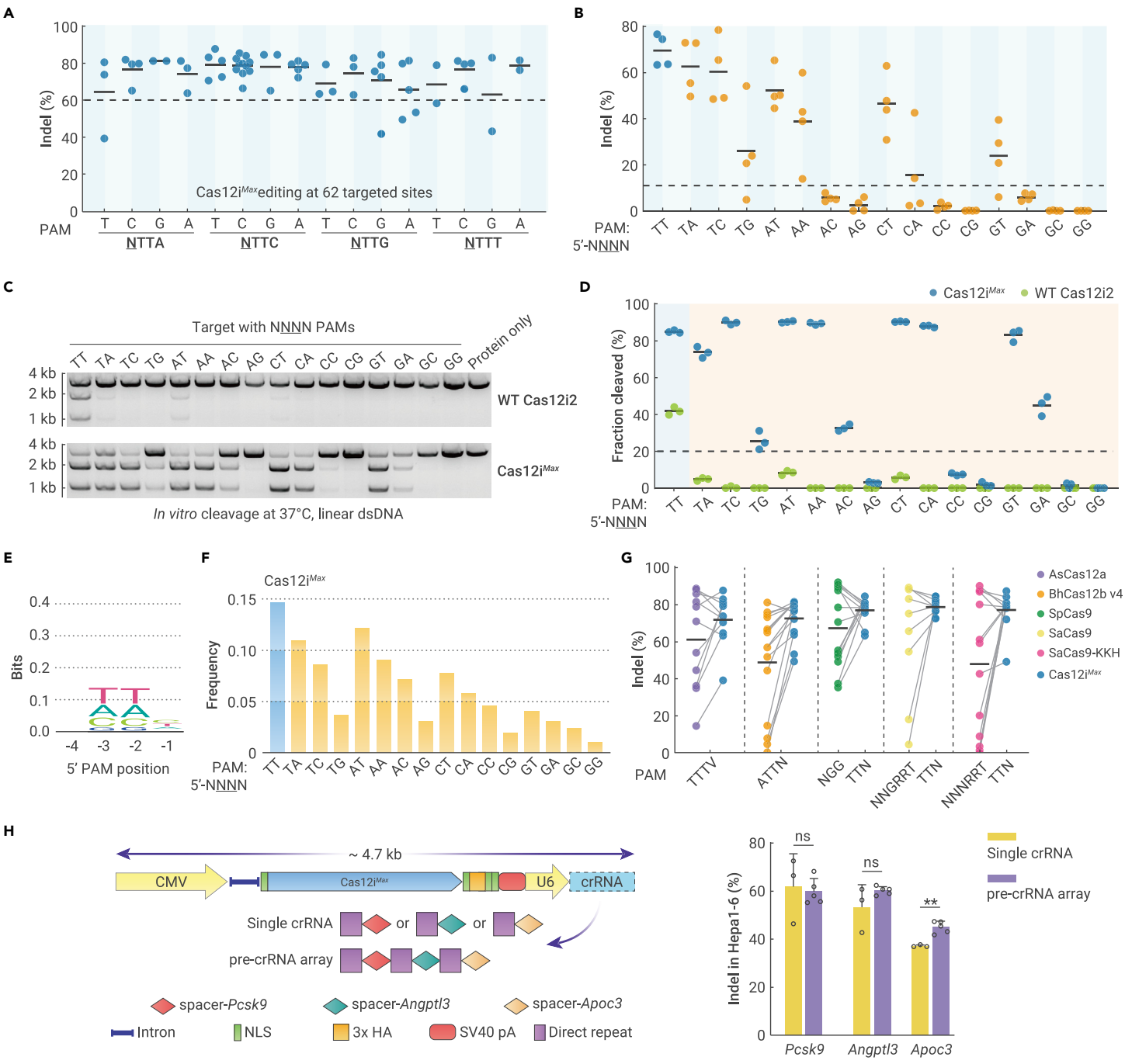


Figure 5. Efficacy of MIDAS in engineering Cas nucleases (A) Cas12i^{Max} exhibited robust indel activity at all 62 tested targeted sites bearing NTTN PAMs, with an average frequency above 70%. Each dot represents the mean indel frequency at one targeted site averaged from n = 2 biological replicates. The horizontal black lines represent the mean indel frequency of each PAM class. (B) Gene-editing activity of Cas12i^{Max} at sites carrying NNNN PAMs. Each dot represents the mean indel frequency at one targeted site averaged from n = 2 biological replicates. The horizontal black lines represent the mean indel frequency of each PAM class. (C) Purified wild-type Cas12i2 and Cas12i^{Max} proteins cleave linear dsDNA carrying the same protospacer with different NNNN PAMs *in vitro* at 37°C. The gel is a representative image from three biological replicates. WT, wild-type. (D) Quantitation of *in vitro* dsDNA cleavage efficiency. Each dot represents efficiency of one biological replicate (n = 3). The horizontal black lines represent the mean efficiency. (E) Sequence logo of Cas12i^{Max} PAM determined by PAM-DOSE assay. (F) Frequency of each recognized NNNN PAMs of Cas12i^{Max}. (G) Comparison of the gene-editing efficiency of Cas12i^{Max} with other Cas benchmarks including AsCas12a, BhCas12b v4, SpCas9, SaCas9, and SaCas9-KKH Each dot represents the mean indel frequency averaged from n = 2 biological replicates. The horizontal black lines represent the mean indel frequency of each nuclease. (F) Multiplexed genome editing mediated by Cas12i^{Max} via single crRNA and pre-crRNA array in mouse Hepa1-6 cells. The cassettes expressing Cas12i^{Max} and crRNA were incorporated into one AAV expression plasmid. Error bars represent SD of mean, n = 3 or 5 biological replicates. Two-tailed, unpaired Student's t test, ns, not significant, **p < 0.01

more commonly used Cas nucleases that could also be effectively packaged into one AAV.[5–7,22,29,30] To our knowledge, except the SpCas9-SpRY variant,[13] the PAM requirement of Cas12i^{Max} would be the simplest among all the Cas nucleases that have been reported. On the other hand, the ability to discriminate between NTNN, NNTN, NAAN, and NCAN and NRRN and NACN (R = C or G) makes Cas12i^{Max} highly suitable for allele-specific gene editing, an important strategy to treat human heterozygous dominant genetic diseases. To explore this potential, we analyzed the number of human dominant point mutations in the ClinVar database[31] to identify suitable PAMs for Cas12i^{Max} to edit[32] (Figure S7B). Notably, Cas12i^{Max} could theoretically distinguish between the mutant from the wild type for 11,668 dominant alleles (52.4% of all dominant entries), which is higher than other more commonly used Cas nucleases[5–7,13,20,22,29,30] (Figure S7C). Taken together, these results demonstrate that MIDAS-engineered Cas12i^{Max} could be a new generation of a powerful and versatile gene-editing tool.

Previous studies have demonstrated that Cas nucleases with extremely flexible PAM recognition profiles exhibit off-target effects (e.g., SpRY).[13]

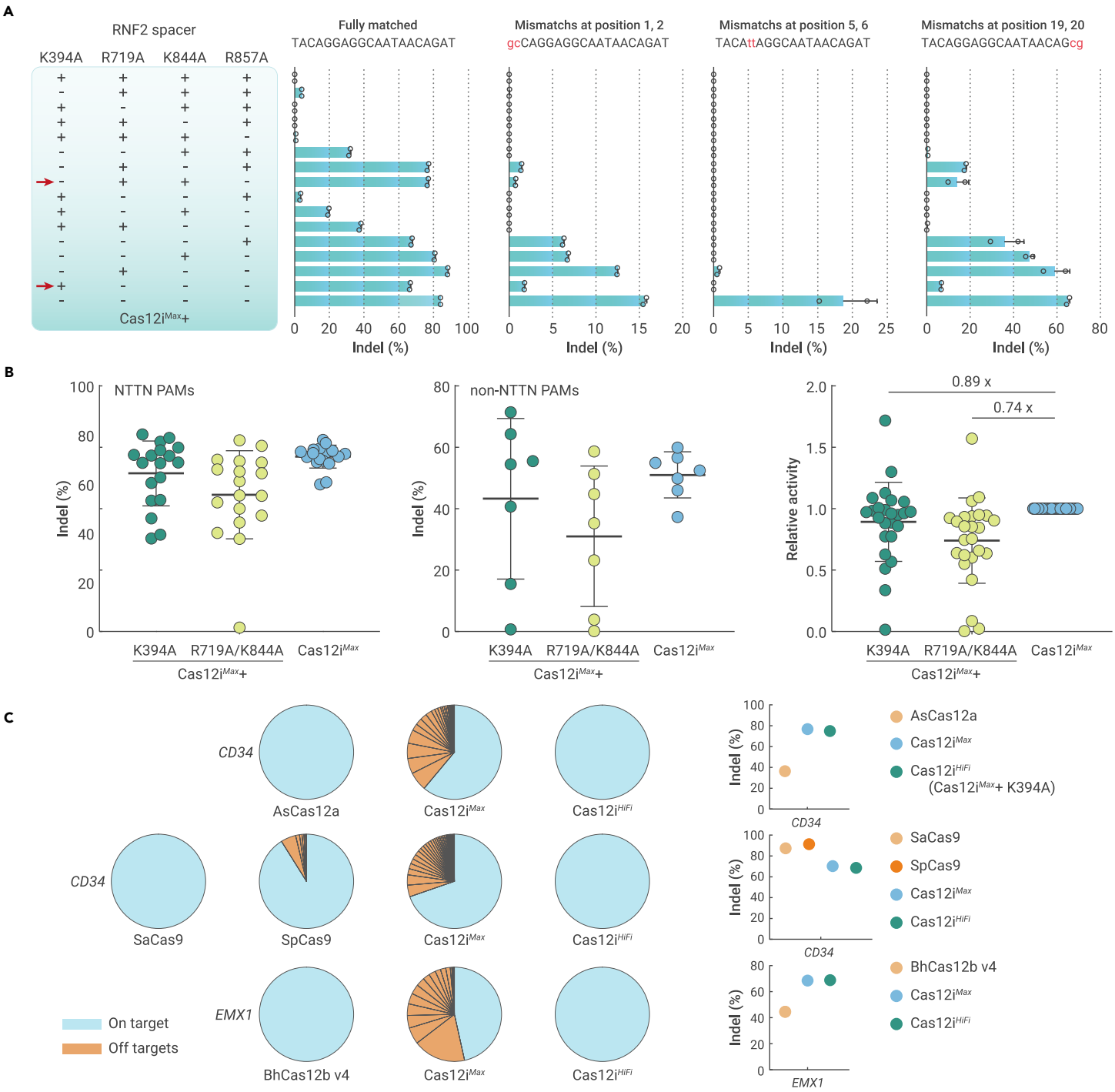


Figure 6. Improving the specificity of Cas12i^{Max} by further engineering (A) Screening variants of Cas12i^{Max} for reducing mismatch tolerance at endogenous locus. Error bars represent SD of mean, n = 2 biological replicates. (B) Comparison of on-target gene-editing efficiency of Cas12i^{Max}, Cas12i^{Max} (+ K394A), and Cas12i^{Max} (+ R719A/K844A) at sites bearing canonical NTTN PAMs and non-NTTN PAMs. Each dot represents efficiency of one experiment. Error bars represent SD of mean. The horizontal black lines represent the mean indel frequency of each nuclease. (C) Genome-wide specificity analysis of AsCas12a, BhCas12b v4, SaCas9, SpCas9, Cas12i^{Max}, and Cas12i^{HiFi} at endogenous loci via GUIDE-seq. For each nuclease, each sector represents the relative proportion of detected cleavage sites. The gene-editing efficiency of the nucleases at on-target sites chosen for GUIDE-seq was shown in graphs. Each dot represents efficiency of one experiment.

Indeed, our GUIDE-seq experiments showed that Cas12i^{Max} has relatively more potential off-target sites compared with other widely applied Cas nucleases (Figure 6C). To minimize the off-target effects of Cas12i^{Max}, we introduced one additional amino-acid mutation to create a high-fidelity version of Cas12i^{Max} (Cas12i^{HiFi}). Notably, Cas12i^{HiFi} exhibited undetected off-target effects via GUIDE-seq and maintained nearly 90% on-target activity of Cas12i^{Max} (Figures 6B and 6C). We believe that high-fidelity Cas12i^{Max} offers minimized off-target effects for applications that need high specificity, especially in the biomedical field.

MATERIALS AND METHODS

Materials and methods related to this work are available in Supplemental information.

Data availability

The sequencing data have been deposited in Genome Sequence Archive for human (GSA-Human) of Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/gsa-human/). The accession number for the sequencing data reported in this paper is HRA001700.

REFERENCES

1. Makarova, K.S., Wolf, Y.I., Iranzo, J., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. ***18***, 67–83.
2. Liu, J.-J., Orlova, N., Oakes, B.L., et al. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature ***566***, 218–223.
3. Yan, W.X., Hunnewell, P., Alfonse, L.E., et al. (2019). Functionally diverse type V CRISPR-Cas systems. Science ***363***, 88–91.
4. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., et al. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. Science ***369***, 333–337.
5. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell ***163***, 759–771.
6. Ran, F.A., Cong, L., Yan, W.X., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature ***520***, 186–191.
7. Strecker, J., Jones, S., Koopal, B., et al. (2019). Engineering of CRISPR-Cas12b for human genome editing. Nat. Commun. ***10***, 212.
8. Kleinstiver, B.P., Sousa, A.A., Walton, R.T., et al. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol. ***37***, 276–282.
9. Jiang, F., and Doudna, J.A. (2017). CRISPR-Cas9 structures and mechanisms. Annu. Rev. Biophys. ***46***, 505–529.
10. Lazzarotto, C.R., Malinin, N.L., Li, Y., et al. (2020). CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. Nat. Biotechnol. ***38***, 1317–1327.
11. Strohkendl, I., Saifuddin, F.A., Gibson, B.A., et al. (2021). Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. Sci. Adv. ***7***, eabd6030.
12. Huang, X., Sun, W., Cheng, Z., et al. (2020). Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. Nat. Commun. ***11***, 5241.
13. Walton, R.T., Christie, K.A., Whittaker, M.N., and Kleinstiver, B.P. (2020). Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science ***368***, 290–296.
14. Nishimasu, H., Shi, X., Ishiguro, S., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science ***361***, 1259–1262.
15. Page, M.I. (1977). Entropy, binding energy, and enzymic catalysis. Angew. Chem. Int. Ed. ***16***, 449–459.
16. Zhang, H., Li, Z., Xiao, R., and Chang, L. (2020). Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. Nat. Struct. Mol. Biol. ***27***, 1069–1076.
17. Teng, F., Cui, T., Feng, G., et al. (2018). Repurposing CRISPR-Cas12b for mammalian genome engineering. Cell Discov. ***4***, 63.
18. Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. Cell ***167***, 1814–1828.e12.
19. Tang, L., Yang, F., He, X., et al. (2019). Efficient cleavage resolves PAM preferences of CRISPR-Cas in human cells. Cell Regen. ***8***, 44–50.
20. Cong, L., Ran, F.A., Cox, D., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science ***339***, 819–823.
21. Mali, P., Yang, L., Esvelt, K.M., et al. (2013). RNA-guided human genome engineering via Cas9. Science ***339***, 823–826.
22. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., et al. (2015). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. ***33***, 1293–1298.
23. Chen, B., Gilbert, L.A., Cimini, B.A., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell ***155***, 1479–1491.
24. Tsai, S.Q., Zheng, Z., Nguyen, N.T., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. ***33***, 187–197.
25. Slaymaker, I.M., Gao, L., Zetsche, B., et al. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science ***351***, 84–88.
26. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., et al. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature ***529***, 490–495.
27. Xu, X., Chemparathy, A., Zeng, L., et al. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol. Cell ***81***, 4333–4345.e4.
28. Takeda, S.N., Nakagawa, R., Okazaki, S., et al. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol. Cell ***81***, 558–570.e3.
29. Kim, D.Y., Lee, J.M., Moon, S.B., et al. (2021). Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. Nat. Biotechnol. ***40***, 94–102.
30. Edraki, A., Mir, A., Ibraheim, R., et al. (2019). A compact, High-Accuracy Cas9 with a dinucleotide PAM for in vivo genome editing. Mol. Cell ***73***, 714–726.e4.
31. Landrum, M.J., Lee, J.M., Riley, G.R., et al. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. ***42***, D980–D985.
32. György, B., Nist-Lund, C., Pan, B., et al. (2019). Allele-specific gene editing prevents deafness in a model of dominant progressive hearing loss. Nat. Med. ***25***, 1123–1130.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program (2019YFA0110800 and 2020YFA0707900 to W.L. and 2018YFA0108400 and 2019YFA0903800 to Q.Z.); the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030403 to W.L.); the National Natural Science Foundation of China (31621004 to Q.Z. and W.L.); and the CAS Project for Young Scientists in Basic Research (YSBR-012 to W.L.). We thank Hua Qing and Xia Yang for their help with FACS. We also acknowledge Jin-Xin Yang and Jing Xu for their help with cell transfection.

AUTHOR CONTRIBUTIONS

W.L., Q.Z., and Y.-C.C. conceived the project and designed the experiments; Y.-C.C., Y.-P.H., X.-G.W., S.-Q.L., Y.N., Y.C., and Z.-K. L. performed the experiments; W.L., Q.Z., Y.-C.C., Y.-P.H., and Y.N. analyzed the data; W.L., Q.Z., and Y.-C.C. wrote the manuscript with assistance from the other authors.

DECLARATION OF INTERESTS

Patents related to this work have been filed.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xinn.2022.100264.