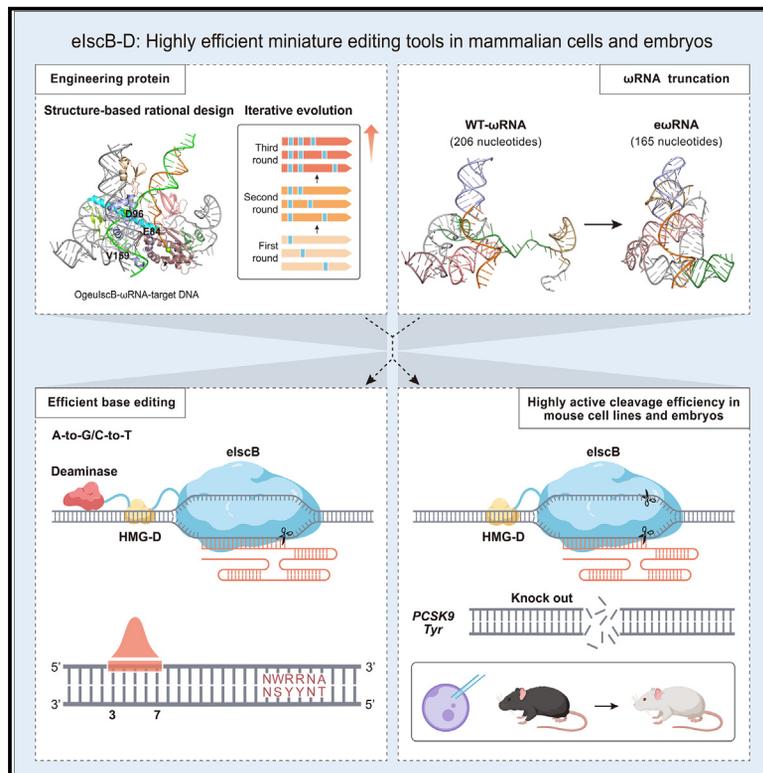


Engineering IscB to develop highly efficient miniature editing tools in mammalian cells and embryos

Graphical abstract



Authors

Niannian Xue, Dishan Hong, Dan Zhang, ..., Liren Wang, Yifan Zhu, Dali Li

Correspondence

lrwang@bio.ecnu.edu.cn (L.W.), yfzhu@bio.ecnu.edu.cn (Y.Z.), dlli@bio.ecnu.edu.cn (D.L.)

In brief

Xue et al. develop a highly efficient, miniature genome-editing system through extensive engineering of ω RNA and IscB nuclease via amino acid substitutions and HMG-D domain fusion. elscB-D can be adapted to base editors to achieve robust A-to-G or C-to-T conversion, and it is highly efficient in mouse embryos to generate disease models.

Highlights

- Engineered IscB and ω RNA significantly increase programmable DNA cleavage activity
- Fusion of DNA deaminases with elscB-D efficiently induces base conversions
- elscB-D/e ω RNA enables robust DNA editing in mouse embryos to generate disease models

Article

Engineering IscB to develop highly efficient miniature editing tools in mammalian cells and embryos

Q1 Niannian Xue,^{1,4} Dishan Hong,^{1,4} Dan Zhang,^{1,4} Qian Wang,^{1,4} Shun Zhang,¹ Lei Yang,¹ Xi Chen,² Yongmei Li,¹ Honghui Han,¹ Chunyi Hu,³ Mingyao Liu,^{1,2} Gaojie Song,¹ Yuting Guan,¹ Liren Wang,^{1,*} Yifan Zhu,^{1,*} and Dali Li^{1,5,*}

¹Shanghai Frontiers Science Center of Genome Editing and Cell Therapy, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China

²Bioray Laboratories Inc., Shanghai, China

Q8 ³Department of Biological Sciences, Department of Biochemistry, Precision Medicine Translational Research Programme (TRP), National University of Singapore, Singapore, Singapore

⁴These authors contributed equally

⁵Lead contact

*Correspondence: lrvang@bio.ecnu.edu.cn (L.W.), yfzhu@bio.ecnu.edu.cn (Y.Z.), dlli@bio.ecnu.edu.cn (D.L.)
<https://doi.org/10.1016/j.molcel.2024.07.007>

SUMMARY

The IscB proteins, as the ancestors of Cas9 endonuclease, hold great promise due to their small size and potential for diverse genome editing. However, their activity in mammalian cells is unsatisfactory. By introducing three residual substitutions in IscB, we observed an average 7.5-fold increase in activity. Through fusing a sequence-non-specific DNA-binding protein domain, the elscB-D variant achieved higher editing efficiency, with a maximum of 91.3%. Moreover, engineered ω RNA was generated with a 20% reduction in length and slightly increased efficiency. The engineered elscB-D/ ω RNA system showed an average 20.2-fold increase in activity compared with the original IscB. Furthermore, we successfully adapted elscB-D for highly efficient cytosine and adenine base editing. Notably, elscB-D is highly active in mouse cell lines and embryos, enabling the efficient generation of disease models through mRNA/ ω RNA injection. Our study

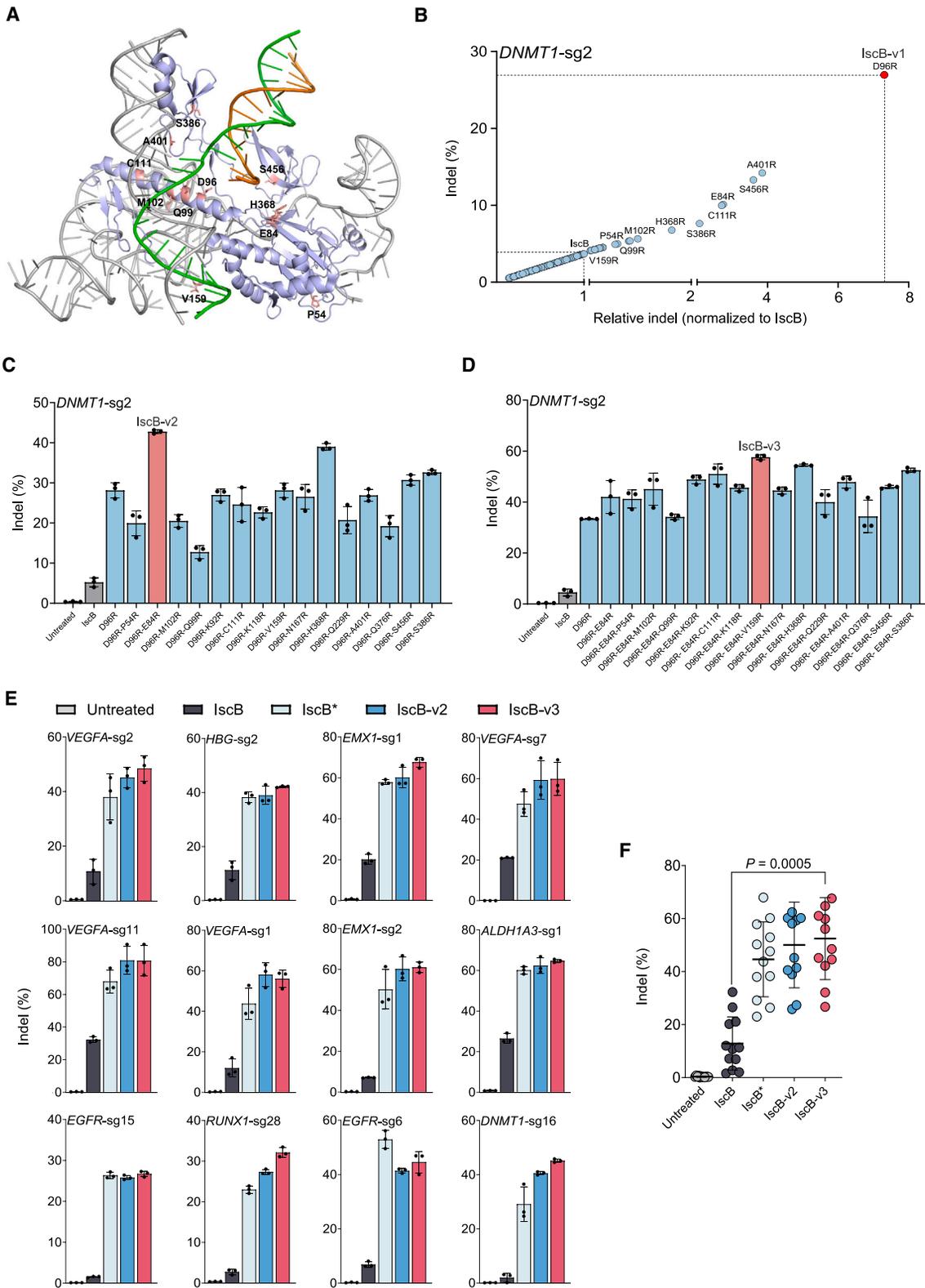
Q2 suggests that these miniature genome-editing tools have great potential for diverse applications.

Q3 Q4 Q7 INTRODUCTION

The clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins (CRISPR/Cas) system is discovered in prokaryotic organisms for adaptive immunity, which has been repurposed for gene editing in various applications, including gene therapy.^{1–5} Among the identified CRISPR systems, type II Cas9^{6–8} and type V Cas12^{9,10} systems, which contain a set of guide RNA and single Cas effector protein, are the most commonly used tools for genome editing in eukaryotic cells. Cas12, which contains a single RuvC nuclease domain and mainly functions through dimerization,¹¹ Cas9 has a more complicated nuclease architecture, a RuvC-like domain split into three conserved catalytic motifs (RuvC-I-III), and an HNH endonuclease domain inserted within it.¹² Compared with Cas12 proteins, the additional HNH domain enables Cas9 to be more efficient and versatile for genome engineering. For instance, Cas9 nickase facilitates highly effective base editing with the fusion of DNA deaminase or prime editing through the recruitment of an engineered reverse transcriptase.^{13,14} These Cas9-derived editors expand the scope of genome engineering

and significantly broaden the application scenarios.¹⁵ However, the fusion of additional domains also increased the molecular weight, which poses a challenge for efficient delivery, especially for AAV-mediated *in vivo* gene therapy.¹⁶

Q9 With the tremendous efforts, several miniCas effectors have been identified, including small Cas12 proteins (Cas12j,¹⁷ Cas12f,¹⁸ and Cas12n¹⁹ ranging from 400 to 800 amino acids [aa]), their ancestral protein TnpB,^{20,21} and the eukaryotic homolog named Fanzor.^{22,23} As mentioned above, the Cas12 system presents challenges when applied to base editors (BEs) and prime editors (PEs). The identification of the insertion sequences Cas9-like OrfB (IscB) family, consisting of ~400 residues and considered as the ancestor of Cas9 from IS200/IS605 transposons,^{24,25} inspires the field to develop highly efficient small-sized editors. The IscB protein is an RNA-guided nuclease in the obligate mobile element-guided activity (OMEGA) system.²⁴ It associates with a cognate 222-nt ω RNA to be guided to the target loci and subsequently cleaves double-stranded DNA targets through the HNH and the RuvC-like nuclease domains, which cleave the target strand (TS, paired with guide RNA) and non-target strand (NTS), respectively.^{26,27} Unfortunately, the activity of IscB in



(legend on next page)

mammalian cells is very limited, as a previous study has demonstrated that OgeulscB induces indels of less than 5% in HEK293FT cells.²⁴ A recent study has shown that the introduction of mutations in OgeulscB (named IscB*) and fusion of T5 exonuclease could dramatically increase its activity in mammalian cells.²⁸ However, the fusion of T5 exonuclease could not be adapted to BE systems. Thus, a miniature ABE was developed through fusion of two TadA-8e domains to efficiently generate A-to-G conversions but showed a wide editing window (between target position 2 and 12).²⁸

In this study, we developed an evolved version of IscB (elscB) that exhibits significantly increased activity through multiple rounds of structure-guided engineering. The editing efficiency was further boosted to as high as 91.3% in mammalian cells by fusing an HMG-D DNA-binding domain with elscB. Additionally, by fusion of TadA-8e or hAPOBEC3A (hA3A) with the evolved IscB nickase, we created compact eiABE and eiCBE capable of efficiently inducing A-to-G and C-to-T conversions guided by an engineered 165-nt ω RNA, respectively. Furthermore, we demonstrated that this engineered IscB system is highly active in mouse cells and embryos, making it suitable for generating animal models of human genetic disorders. Overall, our study presents an IscB variant and small-sized BEs that offer highly efficient genome editing, highlighting their potential for a wide range of applications.

RESULTS

Engineering the IscB protein to enhance its activity in mammalian cells

Previous studies have indicated that substitutions of amino acids with arginine or lysine could enhance the interaction between the Cas protein and negatively charged sgRNA/target DNA.^{29,30} To enhance the activity of the IscB protein, we hypothesized that introducing positively charged amino acids, such as arginine (R) at critical positions could improve the editing activity. Based on our recently published structure of OgeulscB in the complex with the cognate ω RNA and target DNA (Protein Data Bank [PDB] accession number: 8CSZ),²⁷ 280 residues were selected and individually substituted with arginine (R) (Figure 1A). These residues were chosen based on their accessibility and proximity to the binding interfaces with the cognate ω RNA and target DNA, as revealed by the cryoelectron microscopy (cryo-EM) structure (PDB: 8CSZ). Residues that are buried within the protein core or

not proximal to the binding interfaces were excluded from this substitution analysis to focus on regions most likely to affect binding dynamics. When testing an endogenous target site (*DNMT1*-sg2) in HEK293T cells, we observed that some candidates showed improved cleavage activity among which several variants (E84R, D96R, M102R, A401R, S456R, C111R, H368R, and S386R) significantly enhanced the editing efficiency (>1.5-fold compared with wild-type [WT] IscB) (Figures 1B and S1A). Among these variants, the D96R mutant was the most efficient, showing a 7.3-fold elevation compared with WT IscB (WT-IscB). We termed this compact protein variant IscB-v1 (Figure 1B). In the second round of optimization, the D96R variant was individually combined with 15 other mutants that demonstrated elevated activity in the first round. Among these double mutants, the D96R/E84R and D96R/H368R variants were more efficient, and the D96R/E84R mutant exhibited a 1.5- and 8.2-fold increase compared with IscB-v1 and the wt-IscB, respectively (Figure 1C). This mutant was named IscB-v2. Next, another 14 variants were generated based on IscB-v2 with one additional mutation, and most of the triple mutations exhibited increased activity compared with IscB-v2 (Figure 1D). The D96R/E84R/V159R variant exhibited the best performance, with a 12.7-fold increase in efficiency compared with the wt-IscB, and named IscB-v3 (Figures 1D and S1B).

Subsequently, 12 endogenous targets were employed to evaluate IscB-v3 activity in comparison to wt-IscB or IscB*, which was developed through the introduction of E85R, H369R, S387R, and S457R substitutions as reported.²⁸ The activity of wt-IscB was low, ranging from 1.6% to 32.3%, while the activity of IscB-v2 was comparable or slightly higher than that of IscB* (efficiency ranging from 25.8%–80.9% and 23.0%–67.9%, respectively) (Figure 1E). The activity of IscB-v3 was significantly enhanced, showing a maximum improvement of 22.4-fold at the *DNMT1*-sg16 site and an average increase of 7.5-fold compared with wt-IscB (Figures 1E and 1F). In comparison to IscB*, IscB-v3 demonstrated higher efficiency at 11 out of 12 target sites with an average 1.2-fold increase (Figure 1E).

To further improve the cleavage activity, we aimed to enhance the binding affinity of IscB to the target DNA by fusing non-sequence specific DNA-binding domains, as our previous study showed that using this strategy successfully increased the activity of Cas9.³¹ The HMG-D domain (112 aa), deriving from the high-mobility group family of chromosomal proteins, was selected due to its superior effectiveness comparing to other

Figure 1. Engineering the IscB protein to enhance its activity in mammalian cells

(A) The overall structure of the IscB- ω RNA-target DNA ternary complex (PDB: 8CSZ). IscB is in light blue, ω RNA is in gray, target strand is in green, and non-target strand is in orange. Amino acid residues crucial for interaction with DNA substrates and ω RNA have been labeled in pink within the diagram.

(B) Indel frequency and relative indel frequency (wild-type IscB is used for standardization) of 280 arginine-substitution variants evaluated in HEK293T cells at *DNMT1*-sg2. Each circle represents indel for single variant ($n = 3$). Red circle denotes D96R (IscB-v1).

(C) Incorporation of IscB-v1 with additional arginine substitution at *DNMT1*-sg2 in HEK293T cells. The best variant (D96R-E84R) is named IscB-v2 as indicated. Data are means \pm SD ($n = 3$).

(D) Frequency of indel in the third round of optimization through the synergistic incorporation of IscB-v2 with additional arginine substitution at *DNMT1*-sg2 in HEK293T cells. The best variant (D96R-E84R-V159R) is named IscB-v3 as indicated. Data are means \pm SD ($n = 3$).

(E) The cleavage efficiency of wild-type IscB, IscB*, IscB-v2, IscB-v3 at 12 endogenous genomic loci in HEK293T cells. Data are means \pm SD ($n = 3$).

(F) Comparison of average indel efficiency induced by wild-type IscB, IscB*, IscB-v2, IscB-v3 at 12 endogenous loci in (E). Each data point represents the average indel frequency at each target site calculated from three independent experiments. Error bars and p value are derived from these 12 data points. p value was determined by paired two-sided Wilcoxon rank-sum test.

See also Figure S1.

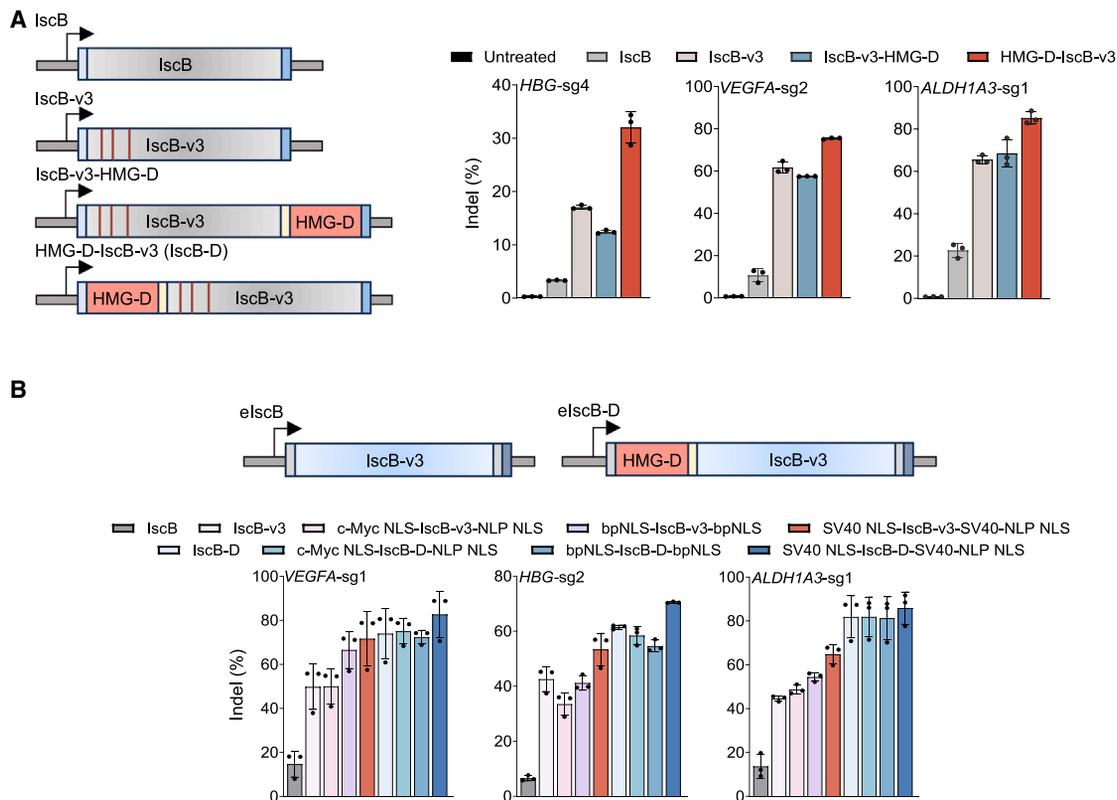


Figure 2. Optimization of IscB architecture by fusing HMG-D protein and adding extra nuclear localization signals

(A) Schematic of the HMG-D fused to N or C terminus of IscB-v3 (left). The light-blue box indicates nuclear localization signals of simian virus 40 (SV40) large T antigen (SV40 NLS) at N termini, the blue box indicates nuclear localization signal from nucleoplasm (NLP NLS), the yellow box indicates the 32-amino-acid linkers, the gray box indicates wild-type IscB or IscB-v3 protein, and the mutation sites of IscB-v3 are labeled with crimson lines. The orange box indicates HMG-D domain. The cleavage efficiency of different fusion strategies at 3 endogenous genomic loci in HEK293T cells (right). Data are means \pm SD ($n = 3$).

(B) Schematic of IscB-v3 and IscB-D with extra nuclear localization signal (upper). The gray box indicates SV40 NLS, the blue box indicates IscB-v3, the dark blue box indicates NLP NLS, the yellow box indicates 32-amino-acid linkers, and the orange box indicates HMG-D domain. Comparison of indel frequency for constructs of different nuclear localization signals at three endogenous targets in HEK293T cells (lower). c-Myc NLS, nuclear localization signal of human c-Myc; bpNLS, bipartite nuclear localization signal. Data are means \pm SD ($n = 3$).

tested domains in our prior investigation.³¹ The HMG-D domain was fused to either the N terminus or C terminus of the IscB-v3 variant. The HMG-D-IscB-v3 outperformed IscB-v3-HMG-D showing up to 85.2% indel efficiency across three tested sites, demonstrating an average improvement of 1.5-fold over IscB-v3 (Figure 2A). The HMG-D-IscB-v3 was designated as IscB-D.

Previous studies have indicated the crucial importance of nuclear localization signals (NLSs) fusion for genome editing in the mammalian system.³² Given that the IscB construct mentioned above included SV40 NLS at the N terminus and nucleoplasm NLS (NLP NLS) at the C terminus, we thought to further optimize the NLS. Initially, attempts were made to utilize different NLS variants, but these were unsuccessful when substituted with bipartite NLS (bpNLS) or c-Myc NLS (Figure 2B). The addition of an extra SV40 NLS at the C terminus of IscB-v3 showed a 1.3- to 1.5-fold (mean 1.4-fold) and 4.7- to 8.2-fold (mean 5.9-fold) improvement compared with IscB-v3 and wt-IscB, respectively. Likewise, the cleavage activity of the IscB-D was slightly enhanced upon fusion with one more SV40 NLS, showing a 1.1-fold increase (Figure 2B). The SV40 NLS-IscB-v3-SV40-

NLP NLS was subsequently designated as elscB, while the IscB-D with extra NLS was denoted as elscB-D (Figure 2B). The data indicate that the enhancement of nuclear localization of IscB is highly significant for its editing activity.

Optimization of ω RNA scaffold to reduce its size and increase activity

As indicated by previous study, the sgRNA sequence plays a crucial role in the editing efficiency of both Cas9³³ and miniature Cas12³⁴ systems. This is likely due to its impact on expression and stability, which may vary between eukaryotes and prokaryotes. The structural analysis reveals that the ω RNA is composed of pairing structures including P1, J1, J2, P2, P3, P4, P5, and a terminal hairpin (Figure 3A).²⁷ The P1 stem-loop displays similarity to the crRNA repeat-tracrRNA anti-repeat duplex of Cas9.²⁶ A previous study has identified alterations to the duplex structure to improve the efficiency of Cas9.³⁵ Based on this knowledge, we first modified the P1 stem loop by incorporating truncation and replacing the linker sequence with TGCA. It showed that the P1 stem loop with 15-nt deletion (P1-del15) exhibited a

Q10

Q11

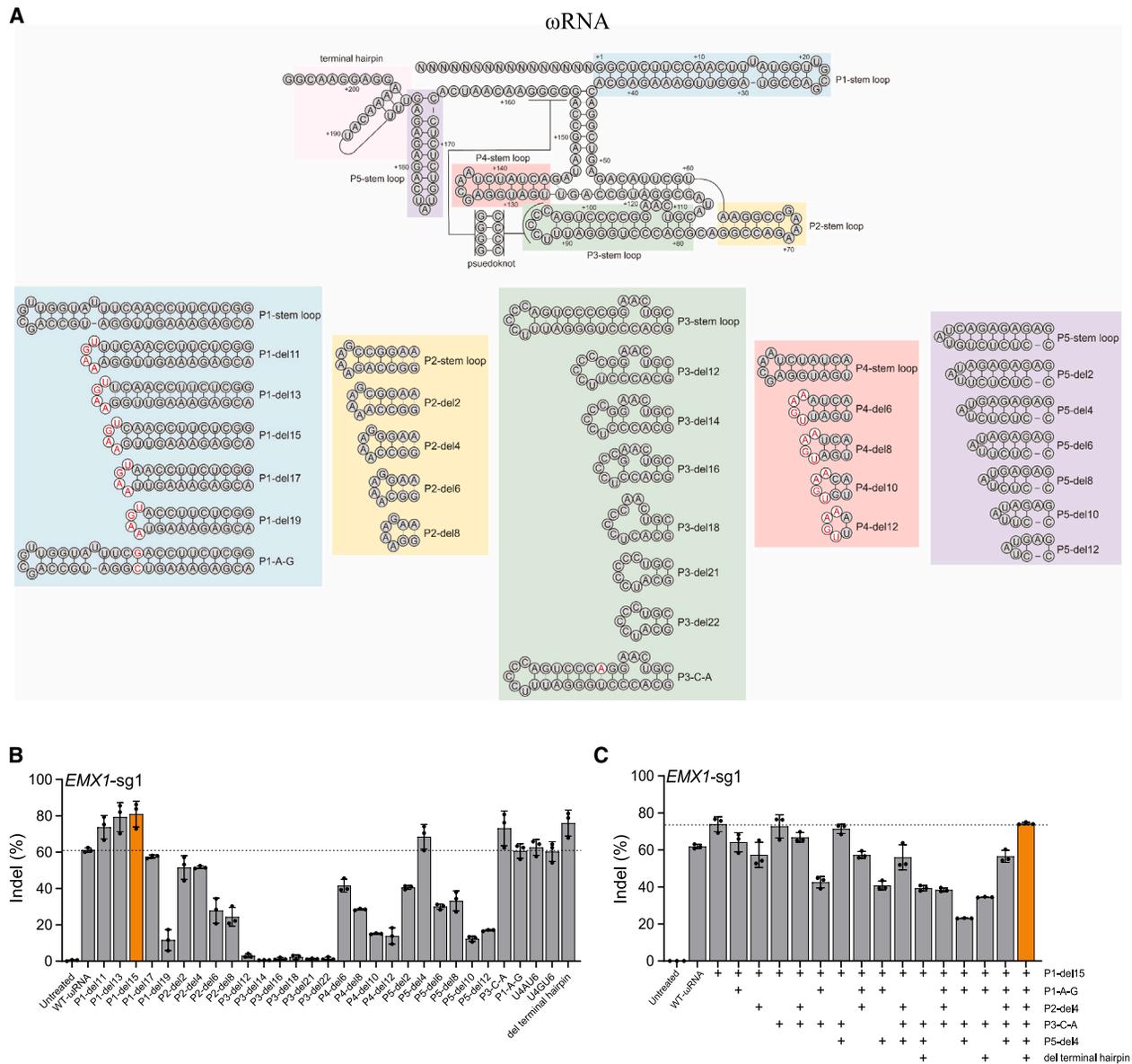


Figure 3. Optimization of ω RNA scaffold to reduce its size and increase activity

(A) Schematic diagram of different stem loop modifications and base pair substitutions. The red letter with white circle denotes base substitutions. Each stem loop is labeled in different color. The top panel is the original sequence of the ω RNA.

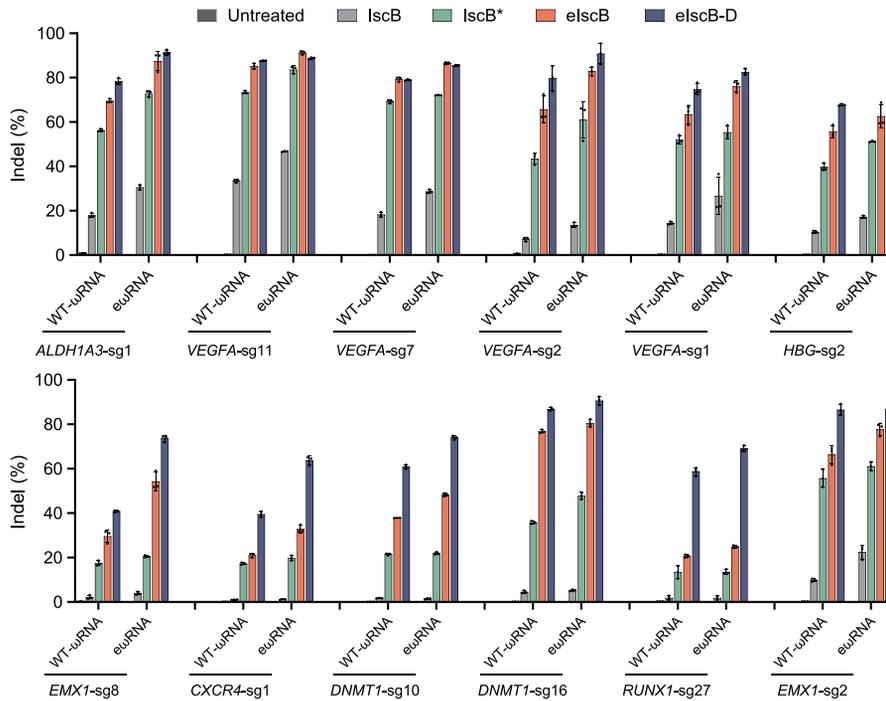
(B) The cleavage efficiency of elscB with modified ω RNA variants at *EMX1*-sg1 site in HEK293T cells. The dotted line represents the indel of elscB-WT- ω RNA. Data are means \pm SD ($n = 3$).

(C) Indel frequencies of combinations with different modified ω RNA variants based on elscB at *EMX1*-sg1 site in HEK293T cells. The dotted line represents the indel of elscB-P1-del15 variant. Data are means \pm SD ($n = 3$).

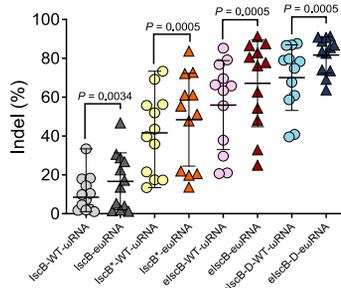
1.3-fold increase compared with WT ω RNA, while two more bases deletion (P1-del17) reduced the efficiency, consistent with previous study (Figure 3B).²⁸ Subsequently, attempts were made to shorten other stem loops of varying lengths to enhance the activity and decrease the length (Figure 3A). It indicated that P2 stem loop with 2/4-nt truncation marginally diminished the efficiency (Figures 3A and 3B). When we truncated the P3 stem loop with 12–22 nt, editing efficiency was drastically reduced,

probably due to the disruption of the pseudoknot between CCCC and the subsequent GGGG (155–158), crucial for stabilizing the overall ω RNA structure (Figures 3A and 3B).^{24,27} The deletion of several nucleotides in the P4 stem loop significantly compromised editing efficiency (Figures 3A and 3B). The P5 stem loop with the 4-nt deletion (P5-del4) showed a 1.1-fold improvement of the cleavage activity (Figures 3A and 3B). Further enhancements were involved with strategic substitution

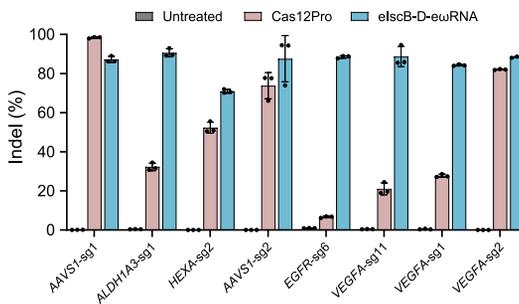
A



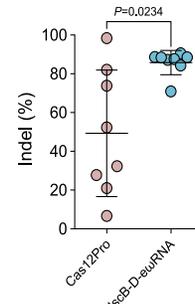
B



C



D



E

VEGFA-sg11		TAM	elscB efficiency (%)	elscB-D efficiency (%)	ALDH1A3-sg1		TAM	elscB efficiency (%)	elscB-D efficiency (%)	EMX1-sg1		TAM	elscB efficiency (%)	elscB-D efficiency (%)
ACAGGTGTGAAAACAG	<u>AAAGAA</u>		(76.9)	(83.9)	AGTGAAGAAGGAGAT	<u>AAGGTA</u>		(80.5)	(87.8)	AATGGTGGAACACAG	<u>AAGGTA</u>		(81.8)	(83.4)
ACAGCTGTGAAAACAG	<u>GTAGAA</u>		(0.9)	(2.1)	AGTGAAGAAGCAGAT	<u>TAGAGA</u>		(3.7)	(4.0)	AATGGTGGAAACACAG	<u>CAGACA</u>		(1.4)	(4.0)
ACAGCTGTGAAAACAG	<u>AAGGAA</u>		(2.0)	(2.6)	AGTGAAGAAGGAGCT	<u>CAGAGA</u>		(0.6)	(0.7)	AATGGTGA AAAACACAG	<u>GTAGTA</u>		(8.1)	(16.6)
ACAGGTGTGTAAACAG	<u>ATGGTA</u>		(1.2)	(3.4)	AGTGGGAGAAAGGAGAT	<u>AAAAATA</u>		(8.3)	(11.2)	AATGCTGGAAACACAG	<u>GAGGGA</u>		(4.0)	(11.9)
ACAGGTGTGAATACAG	<u>AAGACA</u>		(0.9)	(1.4)	AGTGAAGAAGAGAT	<u>TAGATA</u>		(2.2)	(2.4)	AATGGTGGAAAT	<u>AAAGCAAGAA</u>		(3.4)	(3.1)
ACAGGTGTAAAAACAG	<u>TAAACA</u>		(3.0)	(3.1)	AGTGAAGAAGGAT	<u>AAAAATA</u>		(0.7)	(1.2)	AGTGGTGTAAACACAG	<u>CTAGAA</u>		(1.6)	(9.1)
ACAGGTGTGAAAATAG	<u>TGTA AAA</u>		(6.0)	(6.3)	AGTGAAGAAGGAGAA	<u>GAAGAA</u>		(53.1)	(74.7)	AAAGTGGACACACAG	<u>GAGGAA</u>		(3.2)	(8.9)
AAAGGTGTGAAAACAG	<u>GTAAGA</u>		(1.5)	(2.0)	AGTGAAGAAGCAGAT	<u>GAGAAA</u>		(2.3)	(2.1)	AATGATGGAAAACAG	<u>AAAGGA</u>		(1.2)	(0.9)
ACAGGTATGAAAACAG	<u>GAAGTT</u>		(10.9)	(29.0)	AGTGAACAAGGAGAT	<u>GTGGGA</u>		(8.2)	(9.9)	AATGGAGGAAAACACAT	<u>TTTGACA</u>		(0.7)	(1.0)
ACAGGTGTGAAAACAG	<u>GAAGGA</u>		(3.1)	(7.2)	AGTGAAGAAGGAGAGAT	<u>CATGATA</u>		(3.3)	(3.4)	AATAGTGGAAAACAG	<u>GAAGGA</u>		(0.2)	(0.3)
ACAGATGGGAAAACAG	<u>GGAAGAA</u>		(1.7)	(1.6)	AGTGAAGAAGCAGGAGAA	<u>GAGGAA</u>		(1.1)	(1.6)	AATGGGGGAAAACCCAG	<u>CAAGCA</u>		(0.6)	(1.0)
ACAGGTGAAAAACAG	<u>GTAAGAA</u>		(30.0)	(37.1)	AGGGGAAGAAGGAGG	<u>TAAAGAA</u>		(0.9)	(1.9)	AATGTGTGGAAAACAG	<u>TAAAGA</u>		(0.2)	(0.1)
ACAGGTGTATTAACAG	<u>CAGGAA</u>		(1.3)	(1.4)	AGTGGCACAAGGAGAT	<u>CAAGAA</u>		(4.1)	(4.7)	CATGGTGGAAAACCTCAG	<u>AAGATA</u>		(0.2)	(0.2)
ACAGGTGGGAAAACAA	<u>AAAGAA</u>		(2.9)	(3.1)	AGTGAAGGAGGAGAAA	<u>AAAGGA</u>		(2.0)	(2.2)	AATGGTGGGAAACCCAG	<u>CAGGGA</u>		(0.3)	(0.2)
					AGTGGGAGTAGGAGAT	<u>GAGGAA</u>		(3.7)	(16.8)	AATGGTGGAAACGAG	<u>CTAAAA</u>		(0.6)	(0.4)
					AGTGGGAGAAAGGAGAA	<u>ATAACA</u>		(0.3)	(0.3)					
					AGAAGGAAAGGAGAAA	<u>AAAAACA</u>		(0.4)	(0.5)					

(legend on next page)

of unpaired bases, such as replacing C with A at position 103 to form A-U pairing (P3-C-A), leading to a 1.2-fold improvement compared with the WT ω RNA (Figures 3A and 3B). We also tested replacing A-U pairing with G-C pairing at position 11 (P1-A-G), which resulted in comparable editing efficiency to that of the WT ω RNA (Figures 3A and 3B). It has been previously documented that the incorporation of 3'-poly (uridylation) enhances the cleavage efficiency of Cas12a³⁶ and Cas12f.³⁴ However, this strategy did not work when we introduced a polyU sequence at the 3' terminus (Figure 3B). A previous study has shown that truncating the guide RNA region with poorly resolved cryo-EM density did not affect the cleavage efficiency,³⁷ as this region might indirectly interact with the protein and exhibit its flexibility. Indeed, truncation of the terminal hairpin enhanced the activities and reduced overall length of ω RNA (Figures 3A and 3B).

Subsequently, we combined variations to further enhance editing efficiency. The combination of several designs, including the P1 stem loop with deletion of 15 nucleotides, P2 stem loop with deletion of 4 nucleotides, P5 stem loop with deletion of 4 nucleotides, terminal hairpin truncation, P1-A-G, and P3-C-A, yielded a 1.2-fold increase compared with WT ω RNA (Figures 3C and S2A). Simultaneously, the length was reduced to 165 nt, which represented only 80% of the original length. We named this engineered ω RNA as $e\omega$ RNA.

The editing efficiency of ω RNA and $e\omega$ RNA was further compared within distinct IscB variants across 12 endogenous targets. In the elscB group, $e\omega$ RNA enhanced editing efficiency at all targets with an average of 1.26-fold increase compared with ω RNA. In the elscB-D group, all targets exhibited improved cleavage efficiency, resulting in 1.22-fold increase and up to 91.3% indel rates (averaging 81.7%) (Figures 4A and 4B). Notably, elscB/ $e\omega$ RNA and elscB-D/ $e\omega$ RNA displayed average increase of 13.2- and 20.2-fold (up to 60.9-fold) compared with IscB/ ω RNA, respectively (Figures 4A and 4B). The patterns of indels formed by elscB and elscB-D were similar to Cas9, mainly caused small deletions of 1–2 base pairs (Figure S2B). With the substantial increase of the activity and reduction of the length, elscB-D/ $e\omega$ RNA will be very important for various applications, especially for both *in vivo* and *in vitro* gene therapy.

To compare elscB-D with other compact highly efficient CRISPR nucleases, Cas12Pro was selected as it shared partial PAM sequence to IscB, although the guide RNA sequences were oriented opposite to PAM/TAM (Figure S2C).¹⁹ After evaluation of 8 endogenous targets with Cas12Pro preferred PAM (AAG/AAA), Cas12Pro exhibited 6.6%–98.4% (averaging

49.3%) indel rates, while elscB-D showed higher editing efficiency ranging from 70.9% to 90.6% (averaging 85.8%) (Figures 4C and 4D), demonstrating elscB-D was a highly efficient compact programmable nuclease.

To investigate the accuracy of elscB-D induced editing, 43 potential off-target sites of 3 guide ω RNAs identified by Cas-OFFinder prediction were examined (Figure 4E). High-throughput sequencing (HTS) data suggested that elscB and elscB-D showed some off-target activity (4.3% and 6.9% on average, respectively), and the targets with single mismatch adjacent to TAM could also be targeted (Figure 4E). Next, GUIDE-seq assay was performed targeting the *VEGFA*-sg1 site. elscB and elscB-D induced off-target editing at 22 and 23 sites whose reads were over 1% of total sequencing reads (Figure S2D). These data suggest that elscB variants induce obvious off-target edits, but the fusion of the HMG-D domain did not severely increase off-target sites. Further studies are required to understand the underlying mechanisms in order to engineer highly efficient and accurate IscB variants.

Generation of IscB nickase for efficient base editing

To develop a highly efficient BE, we first assessed some mutations to produce IscB nickase, which plays a crucial role in creating a nick on the TS for base editing. Our structural study demonstrated that the RuvC domain of IscB was the counterpart of Cas9 to nick the NTS, and several residues (D60, E192, H339, and D342) were critical for the catalytic activity.²⁷ The H246 residue of IscB represented the position analogous to H840 in the HNH nuclease domain of Cas9 to nick the TS.¹² To test which mutation could generate the most efficient nickase, a double-nicking strategy was employed. As shown in Figure S3A, the elscB with the H339A mutation induced the highest indel rates, suggesting it might be an efficient nickase to catalyze the TS DNA. Subsequently, a minisized ABE was engineered through fusion of the elscB^{H339A} nickase to the TadA-8e deaminase, and it induced the highest 16.2% A-to-G conversion at the *VEGFA*-sg11 target site with minimal indels (1.1%) compared with other potential nickases (Figures S3B and S3C). Catalytically dead IscB with additional H246A mutation fused with deaminase resulted in very limited base conversion (3.1%), suggesting nickase was critical for base editing. These data suggested that elscB^{H339A} nickase was suitable for base editing, and we named it elscBn.

To further enhance the efficiency of the BE, TadA-8e was combined with elscBn to either the N or C terminus using different lengths of linkers (Figure S3D). The results showed that the

Figure 4. Characterization of elscB-D- $e\omega$ RNA in mammalian cells

- (A) The cleavage efficiency of wild-type IscB, IscB*, elscB, and elscB-D with $e\omega$ RNA or wild-type ω RNA at 12 endogenous genomic loci in HEK293T cells. Data are means \pm SD ($n = 3$).
- (B) Comparison of average indel rates induced by wild-type IscB, IscB*, elscB, and elscB-D with WT- ω RNA or $e\omega$ RNA at 12 endogenous genomic loci in (A). Each data point represents the average indel frequency at each target site calculated from three independent experiments. Error bars and p value are derived from these 12 data points. p value was determined by paired two-sided Wilcoxon rank-sum test.
- (C) The cleavage efficiency induced by Cas12pro or elscB-D- $e\omega$ RNA at 8 endogenous genomic loci in HEK293T cells. Data are means \pm SD ($n = 3$).
- (D) Comparison of the average indel frequency induced by Cas12Pro or elscB-D- $e\omega$ RNA at 8 genomic loci in (C). Each data point represents the average indel frequency at each target site calculated from three independent experiments. Error bars and p value are derived from these 8 data points. p value was determined by paired two-sided Wilcoxon rank-sum test.
- (E) Analysis of IscB-dependent DNA on- and off-target at 3 targets (*VEGFA*-sg11, *ALDH1A3*-sg1, *EMX1*-sg1). The mismatched bases are labeled in red compared with the corresponding on-target sequences, and TAM is labeled in blue with underline.

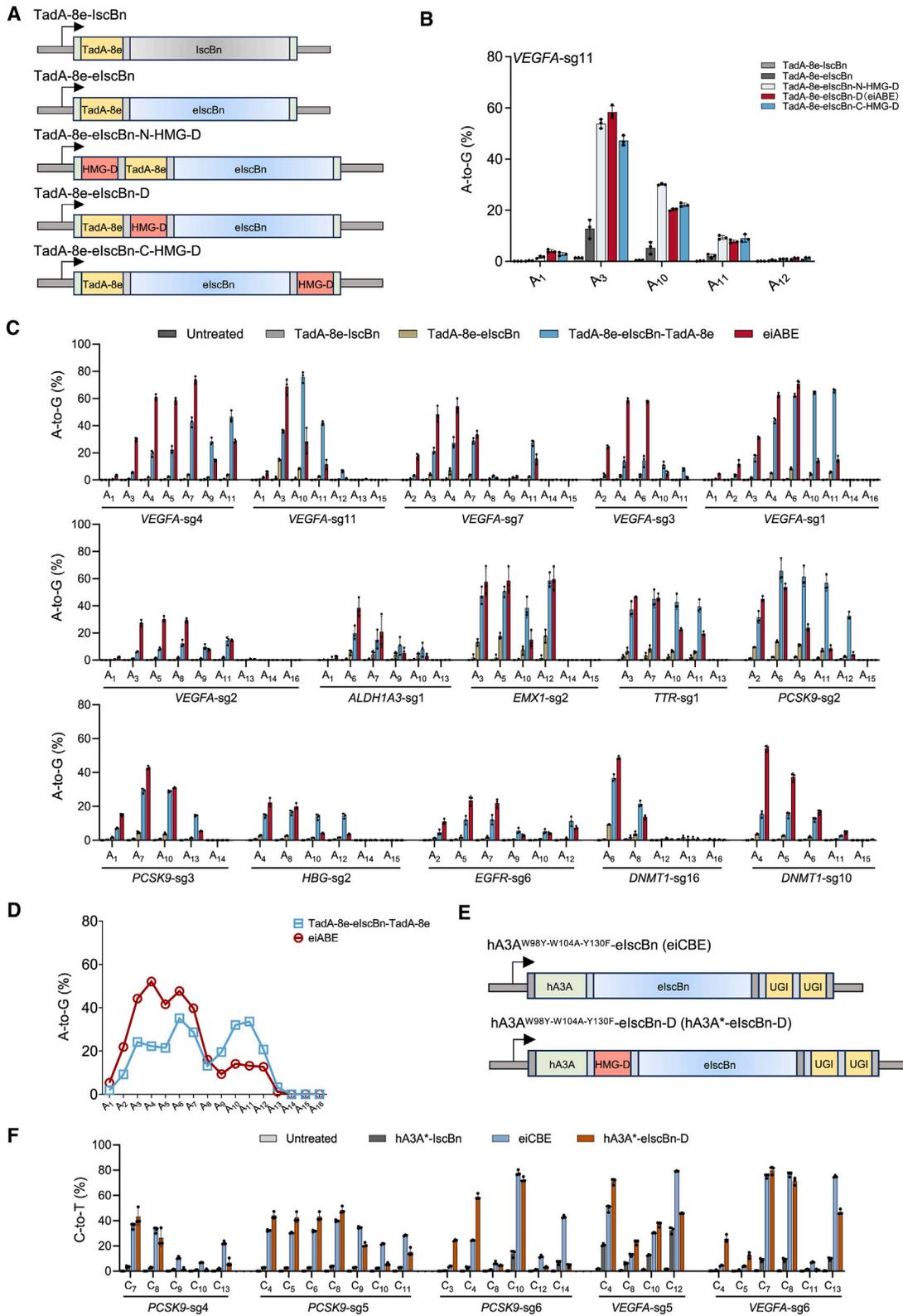


Figure 5. Generation of IscB nickase for efficient base editing

(A) Schematics of the constructs with HMG-D domain fused to TadA-8e-elscBn.

(B) The A-to-G editing efficiency of different HMG-D domain fusion strategies at *VEGFA*-sg11 site in HEK293T cells. Data are means \pm SD ($n = 3$).

(legend continued on next page)

fusion of TadA-8e at the N terminus was superior to the C terminus. Additionally, the length of the linker was found to be crucial, with the 32-aa linker resulting in an average of 10.8% A-G editing at the *VEGFA*-sg11 site (Figure S3D). As elscB-D was more efficient, we tried to fuse HMG-D domain to increase ABE activity. After evaluation of three constructions (Figure 5A), we found that fusion of HMG-D dramatically increased A-to-G activity. When the HMG-D domain was inserted between TadA-8e and elscBn, it showed the highest efficiency up to 58.3% with minimal indel frequency (Figures 5B and S3E). The highly efficient miniature ABE construct was referred as eiABE.

Next, we assessed the efficacy of eiABE at 15 endogenous targets with other ABE variants, including the construct with two TadA-8e domains flanking the elscB nickase, which demonstrated the best performance in the previous study (Figure 5C).²⁸ We found that ABEs without additional DNA-binding or deaminase domain showed very limited activity, while the efficiency of eiABE ranging from 22.2% to 73.6% (averaging 50%) with a 5-nt major editing window (A_3-A_7) and low indel rates across all 15 targets (Figures 5C, 5D, and S3F). ABE with two TadA-8e domains, which was reported in a previous study,²⁸ also exhibited substantial activity (up to 75.8%) at some targets, but statistically, it had a broader editing window (A_3-A_{12}) and lower peak or average activity compared with eiABE (Figures 5C, 5D, and S3G). After that, we compared eiABE with reported LscB* on five endogenous targets. Consistent with the data in Figures 5C and 5D, eiABE exhibited higher A-to-G editing efficiency and narrow editing window across all five targets with similar rate of indels (Figures S4A and S4B).

Moreover, LscB-based CBEs were generated through fusion of two copies of uracil glycosylate inhibitor (UGI) and hA3A* variant, which contained three mutations to reduce their off-target activity (Figure 5E).³⁸ We found that hA3A*-elscBn was highly efficient with editing frequency ranging from 35.7% to 79.2% (averaging 61.9%) and induced low indels (Figures 5F and S4C). However, in contrast to LscB-based ABE, additional fusion of HMG-D domain did not further increase the activity of C-to-T activity (Figures 5E and 5F). Therefore, we named the highly efficient hA3A*-elscBn construct as eiCBE.

Efficient generation of disease model with elscB-D through embryo microinjection

As no published data demonstrated whether LscB was efficient in animal models, we thought to generate a disease model in mice through embryonic injection of the elscB-D/ ω RNA system. We first tested the activity of LscB variants in the mouse N2a cell line and found that both elscB and elscB-D generated indels at the proprotein convertase subtilisin/kexin type 9 (*Pcsk9*) gene loci, and elscB-D induced 58.0% mutation rate, suggesting LscB also functions in mouse cells (Figure 6A). To generate a disease

model in mice, the *Tyrosinase* gene was selected as the target since its mutation caused albinism in humans and the efficiency could be estimated by fur color.³⁹ After testing three target sites within exon 1, *Tyr*-sg21 exhibited the highest efficiency in N2a cells (Figure 6B). We also analyzed the patterns of indels induced by elscB-D/ ω RNA in N2a cells and found similar results in HEK293T cells in Figure S2B (Figure S5). Thus, ω RNA was *in vitro* translated and coinjected with either LscB or elscB-D mRNA into C57BL/6J mouse one-cell embryos. 9 of 12 (75%) F0 pups injected with elscB-D contained mutations with the efficiency ranging from 34.4% to 99% (averaging 58.8%), but LscB failed to generate any mutant pups (Figures 6C–6F and S6). The albino phenotype was obviously observed when the founders grew up, and 5 of the pups showed nearly 100% mutation (Figures 6D and 6E). The data reveal that elscB-D not only demonstrates efficacy in mammalian cell lines but also possesses the ability to efficiently generate disease models through microinjection of embryos.

DISCUSSION

LscB is a promising RNA-guided endonuclease since it is about 400 aa and has two nuclease domains to generate nickase for versatile genome editing tools. In this study, we engineered highly efficient elscB-D through introducing three residual substitutions and fusing with HMG-D DNA-binding domain. elscB-D could be adapted to miniature BEs, and the engineered nuclease was very efficient to generate mouse model, suggesting its potential for a wide range of applications.

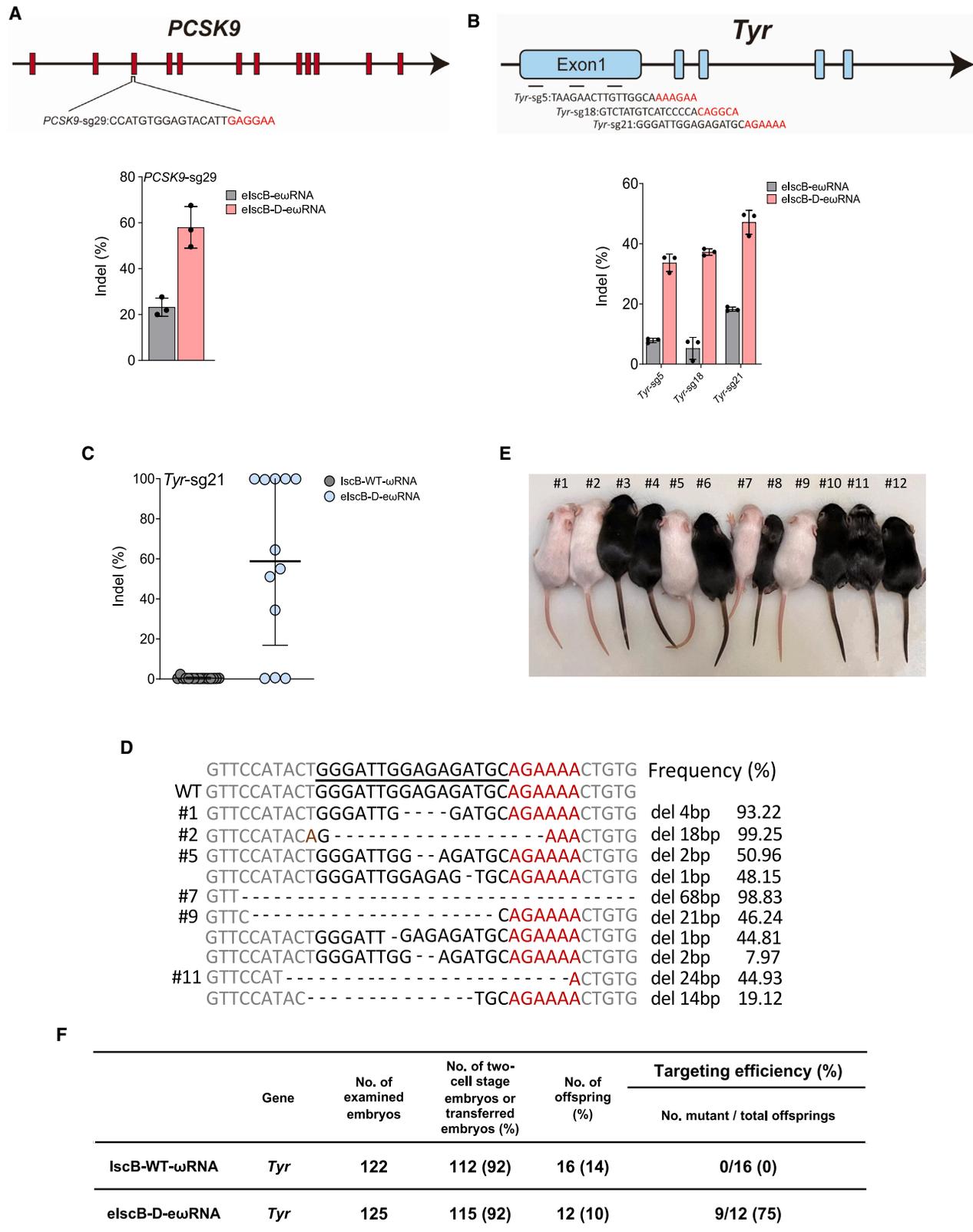
Through three rounds of engineering, we found that introducing arginine substitutions on three residues (D96, E84, and V159) of LscB significantly improved cleavage activity. According to the complex structure, the D96 locates in the bridge helix region of LscB that sits below the editing region of both NTS and TS and extends into the core of the ω RNA (Figure S1B).²⁷ Therefore, the remarkable enhancement in efficacy observed in the D96R mutant can be ascribed to its heightened interaction with the backbones of both DNA strands. This enhanced interaction may play a pivotal role in opening the double-stranded DNA more swiftly compared with the WT. The increased efficacy of E84R might because it assists the entering of TS into the enzymatic center for editing by forming hydrophilic interaction with the backbone of TS since it locates at the halfway between the editing region and enzymatic center (Figure S1B). Although single V159R mutation does not dramatically improve the activity (Figure 1B), it can further enhance the binding affinity with ω RNA in addition with the other two mutations which might change the conformation and put V159R in an optimal position to interact with nucleic acids. Additionally, the incorporation of an HMG-D DNA-binding domain further increased LscB activity,

(C) Comparison of A-to-G conversion frequency of TadA-8e-LscBn, TadA-8e-elscBn, TadA-8e-elscBn-TadA-8e, and eiABE on ω RNA at 15 endogenous loci in HEK293T cells. Data are means \pm SD ($n = 3$).

(D) Average A-to-G editing efficiency of TadA-8e-elscBn-TadA-8e and eiABE at each position within the protospacer at 15 endogenous target sites in (C). Data are means \pm SD ($n = 3$).

(E) Schematics of designs for elscBn-CBE.

(F) Comparison of C-to-T conversion frequency of hA3A*-LscBn, eiCBE, and hA3A*-elscBn-D on ω RNA at 5 endogenous loci in HEK293T cells. Data are means \pm SD ($n = 3$).



(legend on next page)

indicating that augmenting the binding affinity of IscB with DNA is a critical direction for performance enhancement. However, enhanced DNA-binding capacity may also elevate off-target effects, as evidenced by observations of off-target effects through GUIDE-seq or HTS at predicted off-target sites. Consistent with our previous study,³¹ the fusion of the HMG-D domain did not markedly increase the off-targeting risks, as elscB-D did not dramatically increase the number of off-target sites identified via GUIDE-seq (Figure S2D). It is possible to further increase its activity with additional DNA-binding domains, but that will increase both the molecular weight and off-target potentials. Hence, it will be imperative to further optimize the evolved IscB variants to enhance their fidelity through variant molecular evolution strategies, such as continuous and non-continuous evolution, as well as machine learning-assisted directed protein evolution.^{40,41}

A previous study showed that fusing TadA-8e deaminase in each terminus of IscB could dramatically increase the editing activity,²⁸ which was confirmed in this study. However, we found that two TadA-8e domains leading to much wider editing window (from target position 3 to 12) compared with our eiABE, whose major window was from target position 3 to 7 (Figure 5D). In eiABE, we fused single TadA-8e domain at the N terminus with an HMG-D domain inserted between elscB-D and the deaminase domain. This construction is similar to our previous hyCBE, which shows that fusion of single-stranded DNA-binding domain dramatically increases CBE activity.⁴² It suggests that DNA-binding affinity is very critical for base editing as HMG-D seems performed better (the efficiency and condensed editing window) than the additional deaminase domain. Moreover, eiABE is only 837 aa in length since the HMG-D domain is about 112 aa, which is 54 aa less than the additional TadA-8e used in miABE.²⁸ In contrast to eiABE, fusion of HMG-D domain does not increase the activity of cytosine conversions of eiCBE, probably due to high deamination activity of hA3A. Compared with the 191-nt ω RNA,²⁸ the 165-nt $e\omega$ RNA further reduces the total molecular weight of the IscB-based miniature ABE and CBE system, showing the advantage of *in vivo* delivery by AAV system for gene therapy. Since engineering of nuclease domain of Nme2Cas9 could significantly increase its ABE activity,⁴³ this strategy might help to develop IscB variants for minisized ABE or CBE with higher editing efficiency.

Limitations of the study

In this study, we have engineered hyperactive IscB variants and adapted them into highly efficient miniature BEs. However, we

also discovered that IscB induced substantial, though not severe, off-target editing. It will be imperative to further evolve elscB/elscB-D to enhance their fidelity. The TAM of IscB (NWRRNA) is slightly complicated and may limit the targeting range. It would be important to identify new IscB counterparts or engineer elscB variants with more relaxed TAMs to broaden the targeting scope. Since eiABE and eiCBE are highly efficient miniature BEs, evaluating their performance in disease animal models through AAV vectors holds considerable promise for gene therapy applications.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Cell culture
- METHOD DETAILS
 - Plasmid vector construction
 - Cell transfection and genomic DNA extraction
 - *In vitro* transcription of elscB-D mRNA and ω RNA
 - Microinjection of murine zygotes
 - Next-generation sequencing and data analysis
 - Analysis of off-target sites predicted by Cas-OFFinder
 - Guide seq assay
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2024.07.007>.

ACKNOWLEDGMENTS

We are grateful to the East China Normal University Public Platform for Innovation (011). We thank Ying Zhang from the Flow Cytometry Core Facility of the School of Life Sciences in ECNU. This work was partially supported by grants from the National Key R&D Program of China (2023YFC3403400 and 2023YFE0209200 to D.L.), the National Natural Science Foundation of China (no. 32025023, no. 32230064, and no. 32311530111 to D.L.; no. 31971366 to L.W.; no. 82230002 to M.L.; and no. 82100773 to Y.G.), and grants from the Science and Technology Commission of Shanghai Municipality (21JC1402200 to D.L. and 22PJ1403300 and 23HC1400700 to Y.Z.).

Q12

Figure 6. Efficient generation of disease model with elscB-D through embryo microinjection

- (A) Schematic representation of the *PCSK9* locus. The target sequence and TAM sequence are shown in black and red, respectively (upper). Validation of ω RNA sites in N2a cells (lower). Data are means \pm SD ($n = 3$).
- (B) Schematic representation of the *Tyr* locus and different target sites. The target sequence and TAM sequence are shown in black and red, respectively (upper). Screening and validation of ω RNA sites targeting *Tyr* locus in N2a cells (lower). Data are means \pm SD ($n = 3$).
- (C) Indel rate in F0 mice microinjected with IscB mRNA ($n = 16$) and elscB-D mRNA ($n = 12$), respectively.
- (D) Genotyping of representative F0 generation pups microinjected with elscB-D mRNA (five white mice and one heterozygous mouse). Target sequence is underlined. TAM sequence is labeled in red. Mismatched bases are labeled in brown. The frequency was determined by the mutant alleles to total allele counts. Allele reads <0.1% were omitted.
- (E) Phenotypic outcomes of F0 pups after microinjecting elscB-D-mRNA and $e\omega$ RNA. The photo was taken when the mice were 14 days old.
- (F) Summary of the injection parameters used to generate albino mice.
- See also Figure S6.

AUTHOR CONTRIBUTIONS

Conceptualization, D.L.; methodology, N.X., D.H., Q.W., and D.Z.; resources, D.L.; writing – original draft, N.X. and D.L.; writing – review and editing, N.X., Y.Z., L.W., and D.L.; supervision, Y.Z., L.W., and D.L.; project administration, D.L.; funding acquisition, D.L.

DECLARATION OF INTERESTS

Patent applications related to this study have been submitted.

Received: January 9, 2024

Revised: May 28, 2024

Accepted: July 10, 2024

Published: August 2, 2024

REFERENCES

- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. <https://doi.org/10.1126/science.1138140>.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. <https://doi.org/10.1126/science.1225829>.
- Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38, 824–844. <https://doi.org/10.1038/s41587-020-0561-9>.
- Li, G., Li, X., Zhuang, S., Wang, L., Zhu, Y., Chen, Y., Sun, W., Wu, Z., Zhou, Z., Chen, J., et al. (2022). Gene editing and its applications in biomedicine. *Sci. China Life Sci.* 65, 660–700. <https://doi.org/10.1007/s11427-021-2057-0>.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67–83. <https://doi.org/10.1038/s41579-019-0299-x>.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. <https://doi.org/10.1126/science.1231143>.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826. <https://doi.org/10.1126/science.1232033>.
- Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature* 520, 186–191. <https://doi.org/10.1038/nature14299>.
- Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. *Cell* 163, 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>.
- Wu, W.Y., Mohanraju, P., Liao, C., Adiego-Pérez, B., Creutzburg, S.C.A., Makarova, K.S., Keessen, K., Lindeboom, T.A., Khan, T.S., Prinsen, S., et al. (2022). The miniature CRISPR-Cas12m effector binds DNA to block transcription. *Mol. Cell* 82, 4487–4502.e7. <https://doi.org/10.1016/j.molcel.2022.11.003>.
- Xiao, R., Li, Z., Wang, S., Han, R., and Chang, L. (2021). Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res.* 49, 4120–4128. <https://doi.org/10.1093/nar/gkab179>.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949. <https://doi.org/10.1016/j.cell.2014.02.001>.
- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424. <https://doi.org/10.1038/nature17946>.
- 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., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. <https://doi.org/10.1038/s41586-019-1711-4>.
- Zhang, H., Qin, C., An, C., Zheng, X., Wen, S., Chen, W., Liu, X., Lv, Z., Yang, P., Xu, W., et al. (2021). Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer. *Mol. Cancer* 20, 126. <https://doi.org/10.1186/s12943-021-01431-6>.
- Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* 18, 358–378. <https://doi.org/10.1038/s41573-019-0012-9>.
- Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C.A., Li, Z., Cress, B.F., Knott, G.J., Jacobsen, S.E., Banfield, J.F., and Doudna, J.A. (2020). CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* 369, 333–337. <https://doi.org/10.1126/science.abb1400>.
- Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyripides, N.C., Banfield, J.F., and Doudna, J.A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* 362, 839–842. <https://doi.org/10.1126/science.aav4294>.
- Chen, W., Ma, J., Wu, Z., Wang, Z., Zhang, H., Fu, W., Pan, D., Shi, J., and Ji, Q. (2023). Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Mol. Cell* 83, 2768–2780.e6. <https://doi.org/10.1016/j.molcel.2023.06.014>.
- Karvelis, T., Druteika, G., Bigelyte, G., Budre, K., Zedaveinyte, R., Silanskas, A., Kazlauskas, D., Venclovas, Č., and Siksnys, V. (2021). Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 599, 692–696. <https://doi.org/10.1038/s41586-021-04058-1>.
- Xiang, G., Li, Y., Sun, J., Huo, Y., Cao, S., Cao, Y., Guo, Y., Yang, L., Cai, Y., Zhang, Y.E., et al. (2024). Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. *Nat. Biotechnol.* 42, 745–757. <https://doi.org/10.1038/s41587-023-01857-x>.
- Saito, M., Xu, P., Faure, G., Maguire, S., Kannan, S., Altae-Tran, H., Vo, S., Desimone, A., Macrae, R.K., and Zhang, F. (2023). Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature* 620, 660–668. <https://doi.org/10.1038/s41586-023-06356-2>.
- Jiang, K., Lim, J., Sgrizzi, S., Trinh, M., Kayabolen, A., Yutin, N., Bao, W., Kato, K., Koonin, E.V., Gootenberg, J.S., et al. (2023). Programmable RNA-guided DNA endonucleases are widespread in eukaryotes and their viruses. *Sci. Adv.* 9, eadk0171. <https://doi.org/10.1126/sciadv.adk0171>.
- Altae-Tran, H., Kannan, S., Demircioglu, F.E., Oshiro, R., Nety, S.P., McKay, L.J., Dlakić, M., Inskeep, W.P., Makarova, K.S., Macrae, R.K., et al. (2021). The widespread IS200/605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* 374, 57–65. <https://doi.org/10.1126/science.abj6856>.
- Kapitonov, V.V., Makarova, K.S., and Koonin, E.V. (2015). ISC, a novel group of bacterial and archaeal DNA transposons that encode Cas9 homologs. *J. Bacteriol.* 198, 797–807. <https://doi.org/10.1128/JB.00783-15>.
- Kato, K., Okazaki, S., Kannan, S., Altae-Tran, H., Esra Demircioglu, F., Isayama, Y., Ishikawa, J., Fukuda, M., Macrae, R.K., Nishizawa, T., et al. (2022). Structure of the IscB-ωRNA ribonucleoprotein complex, the likely ancestor of CRISPR-Cas9. *Nat. Commun.* 13, 6719. <https://doi.org/10.1038/s41467-022-34378-3>.

27. Schuler, G., Hu, C., and Ke, A. (2022). Structural basis for RNA-guided DNA cleavage by IscB- ω RNA and mechanistic comparison with Cas9. *Science* 376, 1476–1481. <https://doi.org/10.1126/science.abq7220>.
28. Han, D., Xiao, Q., Wang, Y., Zhang, H., Dong, X., Li, G., Kong, X., Wang, S., Song, J., Zhang, W., et al. (2023). Development of miniature base editors using engineered IscB nickase. *Nat. Methods* 20, 1029–1036. <https://doi.org/10.1038/s41592-023-01898-9>.
29. Xu, X., Chemparathy, A., Zeng, L., Kempton, H.R., Shang, S., Nakamura, M., and Qi, L.S. (2021). Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* 81, 4333–4345. <https://doi.org/10.1016/j.molcel.2021.08.008>.
30. McGaw, C., Garrity, A.J., Munoz, G.Z., Haswell, J.R., SenGupta, S., Keston-Smith, E., Hunnewell, P., Ornstein, A., Bose, M., Wessells, Q., et al. (2022). Engineered Cas12i2 is a versatile high-efficiency platform for therapeutic genome editing. *Nat. Commun.* 13, 2833. <https://doi.org/10.1038/s41467-022-30465-7>.
31. Yin, S., Zhang, M., Liu, Y., Sun, X., Guan, Y., Chen, X., Yang, L., Huo, Y., Yang, J., Zhang, X., et al. (2023). Engineering of efficiency-enhanced Cas9 and base editors with improved gene therapy efficacies. *Mol. Ther.* 37, 744–759. <https://doi.org/10.1016/j.ymthe.2022.11.014>.
32. Shui, S., Wang, S., and Liu, J. (2022). Systematic investigation of the effects of multiple SV40 nuclear localization signal fusion on the genome editing activity of purified SpCas9. *Bioengineering (Basel)* 9, 83. <https://doi.org/10.3390/bioengineering9020083>.
33. Dong, C., Gou, Y., and Lian, J. (2022). SgRNA engineering for improved genome editing and expanded functional assays. *Curr. Opin. Biotechnol.* 75, 102697. <https://doi.org/10.1016/j.copbio.2022.102697>.
34. Kim, D.Y., Lee, J.M., Moon, S.B., Chin, H.J., Park, S., Lim, Y., Kim, D., Koo, T., Ko, J.H., and Kim, Y.S. (2022). Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* 40, 94–102. <https://doi.org/10.1038/s41587-021-01009-z>.
35. Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., and Wu, H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* 16, 280. <https://doi.org/10.1186/s13059-015-0846-3>.
36. 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., et al. (2018). Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridylate-rich 3'-overhang. *Nat. Commun.* 9, 3651. <https://doi.org/10.1038/s41467-018-06129-w>.
37. Wu, T., Liu, C., Zou, S., Lyu, R., Yang, B., Yan, H., Zhao, M., and Tang, W. (2023). An engineered hypercompact CRISPR-Cas12f system with boosted gene-editing activity. *Nat. Chem. Biol.* 19, 1384–1393. <https://doi.org/10.1038/s41589-023-01380-9>.
38. Wang, X., Ding, C., Yu, W., Wang, Y., He, S., Yang, B., Xiong, Y.C., Wei, J., Li, J., Liang, J., et al. (2020). Cas12a base editors induce efficient and specific editing with low DNA damage response. *Cell Rep.* 31, 107723. <https://doi.org/10.1016/j.celrep.2020.107723>.
39. Mizuno, S., Dinh, T.T.H., Kato, K., Mizuno-Iijima, S., Tanimoto, Y., Daitoku, Y., Hoshino, Y., Ikawa, M., Takahashi, S., Sugiyama, F., et al. (2014). Simple generation of albino C57BL/6J mice with G291T mutation in the tyrosinase gene by the CRISPR/Cas9 system. *Mamm. Genome* 25, 327–334. <https://doi.org/10.1007/s00335-014-9524-0>.
40. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 38, 883–891. <https://doi.org/10.1038/s41587-020-0453-z>.
41. He, Y., Zhou, X., Chang, C., Chen, G., Liu, W., Li, G., Fan, X., Sun, M., Miao, C., Huang, Q., et al. (2024). Protein language models-assisted optimization of a uracil-N-glycosylase variant enables programmable T-to-G and T-to-C base editing. *Mol. Cell* 84, 1257–1270. <https://doi.org/10.1016/j.molcel.2024.01.021>.
42. Zhang, X., Chen, L., Zhu, B., Wang, L., Chen, C., Hong, M., Huang, Y., Li, H., Han, H., Cai, B., et al. (2020). Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. *Nat. Cell Biol.* 22, 740–750. <https://doi.org/10.1038/s41556-020-0518-8>.
43. Huang, T.P., Heins, Z.J., Miller, S.M., Wong, B.G., Balivada, P.A., Wang, T., Khalil, A.S., and Liu, D.R. (2023). High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. *Nat. Biotechnol.* 41, 96–107. <https://doi.org/10.1038/s41587-022-01410-2>.
44. Zhang, X., Zhu, B., Chen, L., Xie, L., Yu, W., Wang, Y., Li, L., Yin, S., Yang, L., Hu, H., et al. (2020). Dual base editor catalyzes both cytosine and adenine base conversions in human cells. *Nat. Biotechnol.* 38, 856–860. <https://doi.org/10.1038/s41587-020-0527-y>.
45. Li, D., Qiu, Z., Shao, Y., Chen, Y., Guan, Y., Liu, M., Li, Y., Gao, N., Wang, L., Lu, X., et al. (2013). Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 681–683. <https://doi.org/10.1038/nbt.2661>.
46. Xue, N., Liu, X., Zhang, D., Wu, Y., Zhong, Y., Wang, J., Fan, W., Jiang, H., Zhu, B., Ge, X., et al. (2023). Improving adenine and dual base editors through introduction of TadA-8e and Rad51DBD. *Nat. Commun.* 14, 1224. <https://doi.org/10.1038/s41467-023-36887-1>.
47. Park, J., Lim, K., Kim, J.S., and Bae, S. (2017). Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* 33, 286–288. <https://doi.org/10.1093/bioinformatics/btw561>.
48. Hwang, G.H., Park, J., Lim, K., Kim, S., Yu, J., Yu, E., Kim, S.T., Eils, R., Kim, J.S., and Bae, S. (2018). Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* 19, 542. <https://doi.org/10.1186/s12859-018-2585-4>.
49. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197. <https://doi.org/10.1038/nbt.3117>.
50. Tsai, S.Q., Topkar, V.V., Joung, J.K., and Aryee, M.J. (2016). Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.* 34, 483. <https://doi.org/10.1038/nbt.3534>.
51. Chen, L., Zhang, S., Xue, N., Hong, M., Zhang, X., Zhang, D., Yang, J., Bai, S., Huang, Y., Meng, H., et al. (2023). Engineering a precise adenine base editor with minimal bystander editing. *Nat. Chem. Biol.* 19, 101–110. <https://doi.org/10.1038/s41589-022-01163-8>.

Q5 Q6 STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
DH5 α Competent Cells	Sangon Biotech	Cat# B528413-0010
Chemicals, peptides, and recombinant proteins		
Ampicillin	Sangon Biotech	Cat# A610028
DNA Marker	Thermo	Cat# A100193
DMEM	Thermo-Gibco™	Cat# 10567014
Fetal Bovine Serum	Thermo-Gibco™	Cat# 10099141
Fast Digest Enzyme	Thermo	N/A
Gel-Red	HyCyte	Cat# 20210712004
Kanamycin sulfate	Sangon Biotech	Cat# A600286
KOD-Plus-Neo polymerase	TOYOBO	Cat# KOD-401
Penicillin and streptomycin	Thermo-Gibco™	Cat# 15140122
Polyethylenimine	78BIO	Cat# 49553-93-7
PrimeSTAR® Max DNA Polymerase	TaKaRa	Cat# R045A
QuickExtract™ DNA Extraction Solution	Lucigen	Cat# QE0905T
Trelief™ solution A&B	Tsingke	Cat# TSP001
T4 DNA Ligase	Beyotime	Cat# D7008
6x Loading Buffer	Tsingke	Cat# TSJ010
20x PBS Buffer	Sangon Biotech	Cat# B548117-0500
Critical commercial assays		
One step Mouse genotyping kit	Vazyme	Cat# PD101
Clone Express II Multis One Step Cloning Kit	Vazyme	Cat# C113-02
HiPure DNA Clean Up Maxi Kit	Magen	Cat# D2111-03
mMESSAGE mMACHINE T7 Ultra Kit	Invitrogen	Cat# AM1345
MEGAscript™ Kit	Invitrogen	Cat# AM1354
MEGAclear™ Kit	Invitrogen	Cat# AM1908
TIANPrep Mini Plasmid kit	TIANGEN	Cat# DP103-03
Deposited data		
All raw FASTQ files for NGS	This paper	SRA: PRJNA1061449 PRJNA1113560 PRJNA1113943; PRJNA1116803
Structure of the IscB- ω RNA-target DNA ternary complex	This study	PDB: 8CSZ
Experimental models: Cell lines		
Human HEK293T cells	ATCC	Cat# CRL-3216
Neuro-2a cells	ATCC	Cat# CCL-131
Experimental models: Organisms/strains		
Mouse: C57BL/6J wild type	Li et al. ⁴⁵	N/A
Oligonucleotides		
Target sites, see Table S1	This paper	N/A
Primers for PCR and high-throughput sequencing, see Tables S2 and S3	This paper	N/A
Recombinant DNA		
pcDNA-3.1(+) eGFP	Addgene	#78583
pUC19-U6-EF1 α -GFP	Chen et al. ⁵¹	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ABE8e	Addgene	#138489
OgeulscB	Altae-Tran et al. ²⁴	N/A
ω RNA scaffold	Altae-Tran et al. ²⁴	N/A
HMG-D domain	Yin et al. ³¹	N/A
Cas12Pro	Addgene	#203811
Plasmids generated in this study, see Tables S5, S6, S7, S8, and S9	This paper	N/A

Software and algorithms

NCBI	N/A	https://www.ncbi.nlm.nih.gov/
Primer-Blast	N/A	https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome
SnapGene	N/A	SnapGene7.0.2
Prism	GraphPad	GraphPad Prism10.1.2
Cas-Analyzer	Park et al. ⁴⁷	http://www.rgenome.net/cas-analyzer/#!
BE-Analyzer	Hwang et al. ⁴⁸	http://www.rgenome.net/be-analyzer/#!
GUIDE-seq	Github	https://github.com/aryeelab/guideseq
Code for IscB data analysis	This paper	https://zenodo.org/records/10462705 https://zenodo.org/records/10462794

RESOURCE AVAILABILITY

Lead contact

Further Information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dali Li (dlli@bio.ecnu.edu.cn).

Materials availability

Key constructs and plasmids are available on Addgene.

Data and code availability

- The raw high-throughput sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive database under accession codes PRJNA1061449, PRJNA1113943, PRJNA1116803. GUIDE-seq data have been deposited in the NCBI Sequence Read Archive database under accession codes PRJNA1113560. The structure of the IscB- ω RNA-target DNA ternary complex has been deposited under PDB: 8CSZ. These data are publicly available as the data of publication.
- All original code has been deposited at Zenodo and is publicly available. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell culture

Human HEK293T cells (ATCC) and Neuro-2a cells (ATCC) were cultured in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% (v/v) fetal bovine serum (Gibco) and 1% (v/v) penicillin and streptomycin. All cell types were cultured at 37°C with 5% CO₂ and passaged every 2 or 3 days.

METHOD DETAILS

Plasmid vector construction

The plasmid DNA sequences and PCR primers used in this study are listed in [Tables S1](#) and [S2](#). Human codon-optimized OgeulscB and ω RNA scaffold were synthesized by GenScript. Co. Ltd and then cloned into pcDNA-3.1(+) eGFP and pUC19-U6-EF1 α -GFP, respectively. Plasmids of Cas12Pro were the gift from Quanjiang Ji (Addgene plasmid no. 203811). For human genome editing, the AcCas12n and sgRNA-v6 were cloned into two plasmids, respectively. The CMV-driven AcCas12n cassette was constructed into the pcDNA-3.1(+) eGFP plasmid, and the sgRNA-v6 driven by U6 promoter was cloned in pUC19-U6-EF1 α -GFP plasmid. For IscB plasmid constructions, PrimeSTAR® MAX DNA Polymerase (TaKaRa) or KOD-Plus-Neo DNA Polymerase (TOYOBO) was

used for PCR and the Clone Express Multis One Step Cloning Kit (Vazyme) was used for fragment assembly. The construction of ω RNA expression plasmids was consistent with the sgRNA of Cas9.⁴⁴ Briefly, a pair of oligonucleotides were annealed from 95°C down to room temperature slowly and then ligated into Bpil-linearized pUC19-U6- ω RNA-scaffold-EF1 α -GFP. The ω RNA target oligonucleotides were synthesized by BioSune Biotechnology (Shanghai) Co. Ltd and the sequences were listed in [Table S1](#).

Cell transfection and genomic DNA extraction

To measure cleavage and editing efficiency in mammalian cells, HEK293T cells were seeded into 24-well plates (Corning) until they reached nearly 80% density prior to transfection. Next, plasmids (750 ng IscBs and 250 ng ω RNA) were transfected into cells simultaneously using polyethylenimine (PEI, Polysciences) with the ratio of 3 μ l PEI for every 1 μ g plasmids. After 72 h, transfected cells were digested with 0.25% trypsin (Gibco) for fluorescence-activated cell sorting (FACS), and the top 20% GFP positive cells were collected for genome extraction ([Figure S7](#)). About 20,000 cells were collected, and the genomic DNA was extracted by QuickExtract DNA Extraction Solution (Lucigen) following the manufacturer's recommended protocol. The extraction solution was incubated at 65°C for 6 min and then at 98°C for 2 min. To assess the efficiency of gene editing, genome loci of interest were amplified using KOD-Plus-Neo DNA polymerase (TOYOBO). To determine the genotype of modified mice, genomic DNA for PCR was extracted from the toes of newborn mice using the One Step Mouse Genotyping Kit (Vazyme) according to the manufacturer's instructions. The extraction solution was incubated at 55°C for 30 min and then at 95°C for 5 min, and using the One Step Mouse Genotyping Kit (Vazyme) to amplify the genomic of interest as above. All PCR primers used in this experiment are listed in [Tables S2](#) and [S3](#).

In vitro transcription of elscB-D mRNA and ω RNA

For the transcription of elscB-D mRNA, mMMESSAGE mMACHINE T7 Ultra Kit (Invitrogen) was utilized following the manufacturer's instructions. The ω RNA template was transcribed with MEGAshortscript Kit (Invitrogen) following the manufacturer's instructions. Both ω RNA and mRNA were purified with MEGAclean™ Kit (Invitrogen) and then eluted in nuclease-free water with storing at -80°C.

Microinjection of murine zygotes

The manipulation of rodent embryos was previously described.⁴⁵ Murine zygotes were obtained by superovulating female mice and mating them with males of the same genetic background. Before microinjection, the murine zygotes were cultured in KSOM embryo culture medium. Solutions containing elscB-D mRNA (100 ng/ μ l) and ω RNA (200 ng/ μ l) were injected into the cytoplasm of one-cell stage embryos after being diluted in nuclease-free water. Eppendorf TransferMan NK2 micromanipulator was used to administer the injection. Zygotes injected into the complex were promptly transferred into pseudo-pregnant female mice. All mice were cultured in a pathogen-free environment with a 12 h light/dark cycle and constant access to food and water at all times. All animal experiments were conducted following the guidelines established by the Association for Assessment and Accreditation of Laboratory Animal Care in Shanghai and were approved by the East China Normal University Center for Animal Research.

Next-generation sequencing and data analysis

The genomic DNA was amplified from 100-150 ng using high-throughput sequencing primer pairs specified in [Tables S2](#) and [S3](#) to target the regions of interest in the genome. All PCR primers containing an adapter sequence at the 5'-end (forward 5'-GGAGTGAG-TACGGTGTGC-3'; backward 5'-GAGTTGGATGCTGGATGG-3') were prepared for high-throughput sequencing (HTS) libraries. Next, the amplified product underwent a second-round of PCR using primers with different barcodes. The resulting HTS libraries were pooled and purified by electrophoresis using the HiPure Gel Pure DNA Mini Kit (Magen) with a 1.5% agarose gel. The elution was done with 50 μ l of nuclease-free water, and then sequenced with 150-bp paired-end reads on an Illumina HiSeq platform. For the batch analysis of FASTQ files, the reference sequences were set to full-length and analyzed as previously described.⁴⁶ Any reads with mismatches to both sides of the given reference sequence and a frequency below 1 was discarded. The alleles containing insertions or deletions (indels) were quantified by counting the number of reads that were shorter (deletions) or longer (insertions) than the reference sequence. This was done by dividing by the total number of reads in the sample, using a custom script based on the Cas-analyzer pipeline.⁴⁷ Conversions (A-to-G/C-to-T editing) were quantified by counting the number of reads where the expected A/C position was called as a G/T and dividing by the total number of reads in the sample. This was done using a custom script based on the BE-Analyzer pipeline.⁴⁸ It should be noted separately that there is a ploy A near the *VEGFA*-sg7 target, and during HTS, there is a tendency for the number of A to change, leading to the natural generation of indels. We filtered this for analysis, focusing only on the presence of indel products within the target and its 5-bp range.

Analysis of off-target sites predicted by Cas-OFFinder

To evaluate the safety and specificity of elscB and elscB-D, the Cas-OFFinder (<http://www.rgenome.net/cas-offinder/>) was used to predict potential off-target sites. Firstly, the target sequence and TAM should be determined on the Cas-OFFinder website. We entered 22nt sequence of interest into the box including 16nt-target sequence and 6nt-TAM ("NWRRNA") since no choice for the TAM of IscB protein. The mismatch was then restricted up to two nucleotides and the PAM sequence was set to "NNN". Among the sequences output from the website, we persisted the sequence of TAM as "NWRRNA" and then selected the mismatched sequence. The primers for PCR were provided in [Table S3](#).

Guide seq assay

GUIDE-seq assay were performed as previously described.⁴⁹ Briefly, target plasmids and double-stranded oligodeoxynucleotide (dsODN) were electroporated into 1 million HEK293T cells, which were then cultured for 72 hours before genomic DNA isolation. The purified genomic DNA was subjected to fragmentation, end-repair, A-tailing, adapter ligation, and dsODN-specific amplification. The libraries were sequenced on an MGI-seq 2000 sequencer (BGI company) using a pair-end mode. Data were analyzed and visualized using open-source guide-seq software with an NWRRNA TAM, and allowing up to six mismatches before downstream data processing.⁵⁰ Output sequence data were analyzed using the GUIDE pipeline (<https://github.com/aryeelab/guideseq>). All off-target sites identified by GUIDE-seq was provided in [Table S4](#).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were repeated at least three times on different days (normally $n = 3$). These replicates were subsequently used to calculate means and standard deviations (SD). Prism 10.1.2 (GraphPad) was used for statistical analysis. *P* values were calculated using GraphPad Prism 10.1.2 to assess the significance of the differences between two groups with a paired two-sided Wilcoxon rank-sum test. *P* < 0.05 was considered significant.