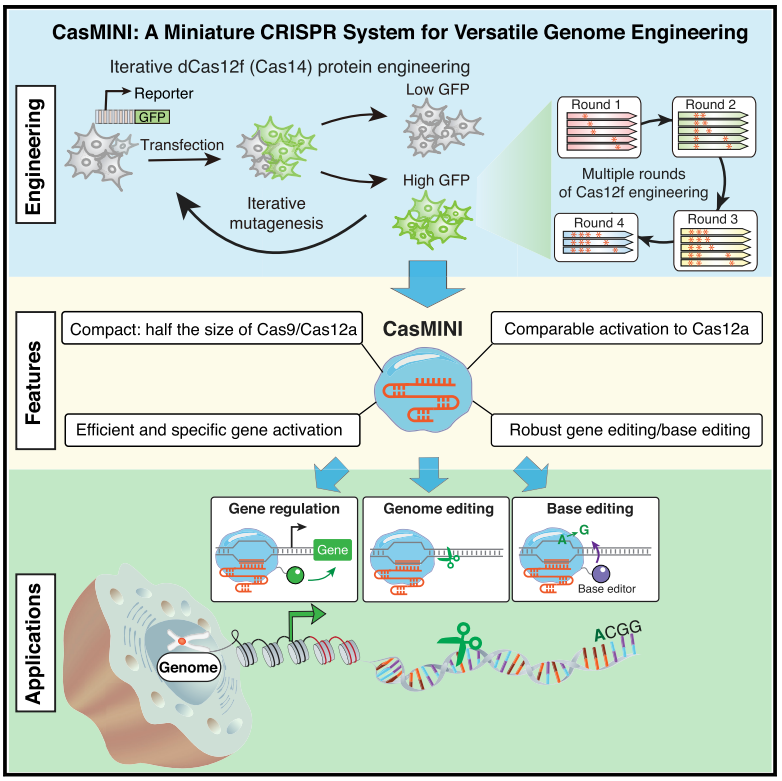
Molecular Cell

Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing

Graphical abstract



Highlights

- Protein and RNA engineering enable Cas12f to function robustly in mammalian cells
- The engineered CasMINI is compact and less than half the size of Cas9 and Cas12a
- CasMINI is efficient and specific for gene activation and is comparable with Cas12a
- CasMINI is versatile and allows robust genome editing and base editing

Authors

Xiaoshu Xu, Augustine Chemparathy, Leiping Zeng, Hannah R. Kempton, Stephen Shang, Muneaki Nakamura, Lei S. Qi

Correspondence

stanley.qi@stanford.edu

In brief

Xu et. al developed a miniature CRISPR system for genome engineering via protein and guide RNA engineering. Whereas the natural Cas12f does not function in mammalian cells, engineered Cas12f mutants, named CasMINI, show comparable activities with Cas12a for efficient gene activation. CasMINI also enables robust gene editing and base editing.

Xu et al., 2021, Molecular Cell *81*, 1–13
October 21, 2021 © 2021 Elsevier Inc.
https://doi.org/10.1016/j.molcel.2021.08.008

CellPress

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

Molecular Cell

CellPress

Technology

Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing

Xiaoshu Xu,[1] Augustine Chemparathy,[1] Leiping Zeng,[1] Hannah R. Kempton,[1] Stephen Shang,[1] Muneaki Nakamura,[1] and Lei S. Qi[1,2,3,4,*]

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[2]Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA
[3]ChEM-H, Stanford University, Stanford, CA 94305, USA
[4]Lead contact
*Correspondence: stanley.qi@stanford.edu
https://doi.org/10.1016/j.molcel.2021.08.008

SUMMARY

Compact and versatile CRISPR-Cas systems will enable genome engineering applications through high-efficiency delivery in a wide variety of contexts. Here, we create an efficient miniature Cas system (CasMINI) engineered from the type V-F Cas12f (Cas14) system by guide RNA and protein engineering, which is less than half the size of currently used CRISPR systems (Cas9 or Cas12a). We demonstrate that CasMINI can drive high levels of gene activation (up to thousands-fold increases), while the natural Cas12f system fails to function in mammalian cells. We show that the CasMINI system has comparable activities to Cas12a for gene activation, is highly specific, and allows robust base editing and gene editing. We expect that CasMINI can be broadly useful for cell engineering and gene therapy applications *ex vivo* and *in vivo*.

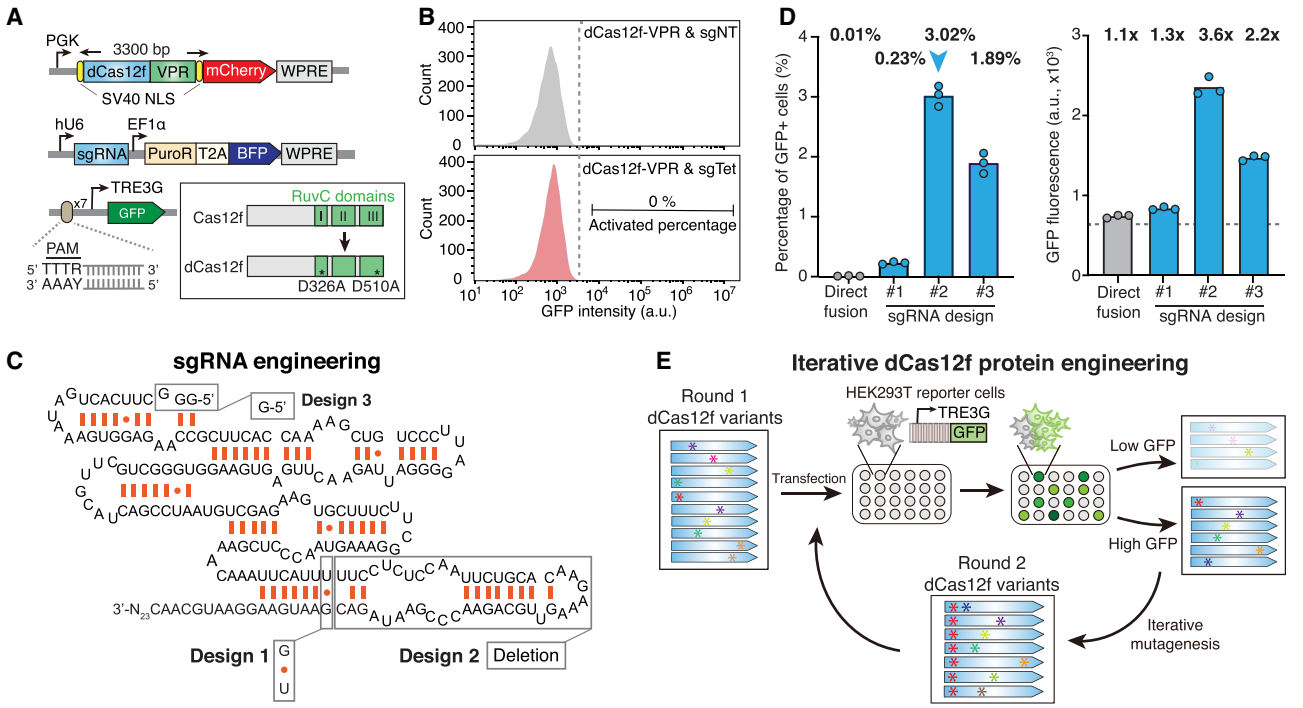
INTRODUCTION

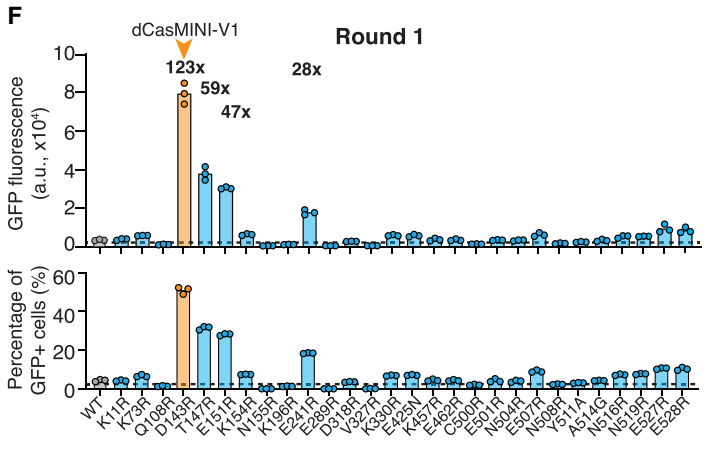
The development of CRISPR-Cas systems for human cells has revolutionized genome engineering (Cong et al., 2013; Jinek et al., 2012). Cas nucleases (e.g., *Streptococcus pyogenes* Cas9, *Lachnospiraceae bacterium* Cas12a) allow efficient and specific genome editing, while the nuclease-deactivated dCas molecules fused with transcriptional or epigenetic effectors enable targeted regulation of endogenous genes in mammalian cells via programmable binding to DNA (Kleinstiver et al., 2019; Qi et al., 2013; Swarts et al., 2017; Tak et al., 2017; Xu and Qi, 2019; Zetsche et al., 2015). These systems offer opportunities for the development of gene therapies against a variety of genetic diseases (Fellmann et al., 2017; Hilton et al., 2015; Klann et al., 2017). However, their large sizes often restrict delivery into cells, which impedes clinical applications. For example, adeno-associated virus (AAV), a vector widely applied for *in vivo* delivery, has limited packaging capacity of the payload (<4.7 kb), and many Cas fusion proteins are beyond this limit (Doudna, 2020; Wang et al., 2019a; Zhang, 2019). As a result, there is a great need to engineer highly efficient, compact Cas systems to facilitate the next generation of genome engineering applications.

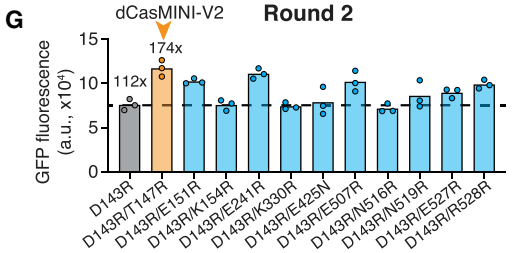
Compared with Cas9 or Cas12a (usually 1,000–1,500 amino acids), the discovery of smaller naturally occurring Cas effectors, including Cas12f (Cas14, 400–700 amino acids) and Cas12j (CasΦ, 700–800 amino acids) offers a natural reservoir of compact Cas effectors (Harrington et al., 2018; Karvelis et al., 2020; Pausch et al., 2020; Takeda et al., 2021). The class 2 type V-F system, CRISPR-Cas12f, is a family of exceptionally compact RNA-guided nucleases from uncultivated archaea. Originally identified as a single-stranded DNA (ssDNA) cutter, the wild-type Cas12f system was recently discovered to possess protospacer adjacent motif (PAM) sequences for double-stranded DNA cleavage *in vitro* (Harrington et al., 2018; Karvelis et al., 2020). However, it remains unknown whether the compact Cas12f effector can be used in mammalian cells.

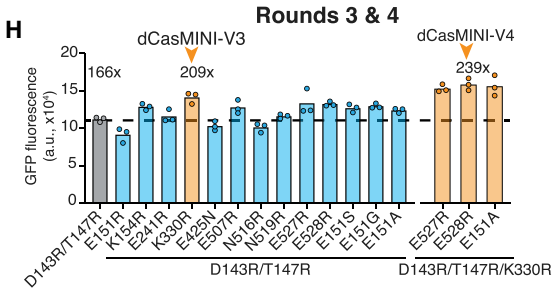
Here we apply RNA engineering and protein engineering to the type V-F Cas12f (529 amino acids) system and generate a compact, efficient, and specific system for mammalian genome engineering, which we name CasMINI. We report that the natural Cas12f protein has no detectable activity in mammalian cells. By optimizing the single guide RNA (sgRNA) design and performing multiple rounds of iterative protein engineering and screening (Qu et al., 2020; Reetz and Carballeira, 2007; Xu et al., 2016), we generate a class of Cas12f variants (i.e., CasMINI) which, when fused to a transcriptional activator, can efficiently activate reporter and endogenous gene expression. This dCasMINI-mediated gene activation has significant improvement over the wild-type dCas12f system, has comparable activation ability with the dCas12a system, and is specific in mammalian cells without detectable off targets. The dCasMINI-mediated adenine base editor (dCasMINI-ABE) allows robust conversion of A·T to G·C. Additionally, we show that the nuclease-active CasMINI enables efficient genome editing with distinct editing patterns from Cas9. Thus, CasMINI provides a useful tool for broad

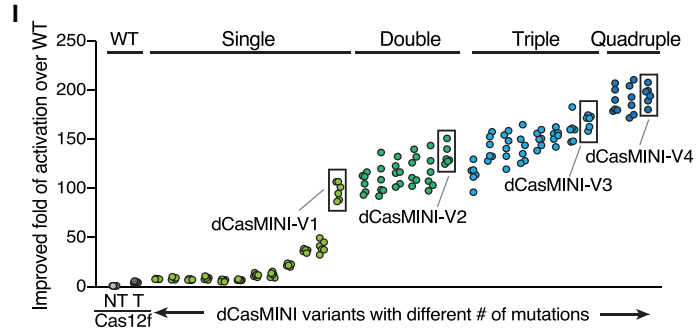
Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

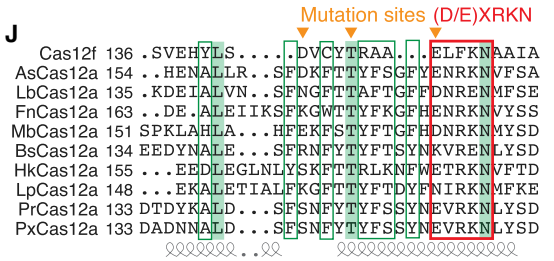












(legend on next page)

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

genome engineering applications that require compact Cas fusion proteins for delivery and cellular function.

RESULTS

Engineering the Cas12f sgRNA scaffold for enhanced gene activation in mammalian cells

The Cas12f used in this study was derived from uncultivated archaea, which was recently reported to possess a TTTR PAM (Karvelis et al., 2020). Using a designed sgRNA, a chimeric RNA fusion between the CRISPR-associated RNA (crRNA) and the *trans*-acting crRNA (tracrRNA), the previous work also demonstrated *in vitro* dsDNA cleavage.

We first tested whether this naturally occurring Cas12f could function in mammalian cells. We generated a nuclease-deactivated Cas12f (dCas12f) by introducing two mutations (D326A and D510A) to the conserved active sites of Cas12f in the RuvC domains (Harrington et al., 2018; Karvelis et al., 2020) (Figure S1). The resulting dCas12f protein was fused to a tripartite VP64-P65AD-Rta (VPR) transcriptional activator (Chavez et al., 2016; Figure 1A; Table S1).

Using a TRE3G-GFP HEK293T cell line, we tested whether the reported sgRNA scaffold with a spacer sequence targeting TetO repeats with a TTTA PAM in the TRE3G promoter could activate the GFP reporter (see STAR Methods; Table S2). We measured the GFP expression using dCas12f-VPR via flow cytometry and observed no GFP activation (Figure 1B), implying that the natural Cas12f system fails to function as a useful activator in the context of the human genome.

We attributed the lack of activity to the suboptimal design of the sgRNA and/or the weak binding activity of Cas12f to the genomic DNA. To test this, we first generated three designed sgRNA variants with different scaffolds based on the natural tracrRNA sequence (Figure S2A; Table S3), including a G:U swap to disrupt the poly U sequence (design 1) similar to previously reported (Chen et al., 2013), RNA hairpin truncation (design 2), and poly G removal (design 3) (Figure 1C; Figure S2A). Interestingly, all the designs improved gene activation, with design 2 outperforming the other two. Although the sgRNA by directly fusing crRNA-tracrRNA showed no activation with dCas12f-VPR, sgRNA design 2 exhibited modest activation (3% of GFP+ cells, 3.6-fold upregulation over the non-targeting sgRNA in transfected cells) (Figure 1D; Figure S2B). This optimized sgRNA design 2 scaffold was used for all subsequent experiments.

Engineering the Cas12f protein for enhanced gene activation in mammalian cells

To test the effect of dCas12f protein on gene expression, we used an iterative protein engineering strategy. First, to examine the effects of domain positioning, folding, and nuclear localization, we tested a panel of fusion variants, comparing our original dCas12f-VPR construct (fusion #1) with ten other constructs with varying positioning of the VPR domain along with linkers and nuclear localization signals (fusions #2–#11; Figure S2C). We found that one such variant (fusion #5) containing an N-terminal SV40 NLS and a C-terminal c-Myc NLS demonstrated modest enhancement of gene activation (Figures S2D and S2E). We adopted this fusion configuration for subsequent experiments.

We next tested if the gene activation activity could be further improved by protein engineering of the Cas12f itself. We used an iterative protein engineering strategy to screen the ability of dCas12f variants fused to VPR to drive activation of GFP (Figure 1E). In this assay, variants showing enhanced activation of GFP in each cycle were used as the starting point of the next cycle.

We hypothesized that by mutating amino acids in the DNA-binding pocket, we could affect DNA binding, a strategy that has been shown to enhance the activity of Cas12a protein (Kleinstiver et al., 2019; Strecker et al., 2019). To generate candidates for mutagenesis, we aligned the protein sequences of Cas12f to Cas12a proteins with available structures and predicted conserved motifs and residues in the target DNA binding pocket. On the basis of this analysis, we selected 28 candidate amino acids, creating a corresponding library of single-mutant dCas12f variants, each with one of the candidate amino acids mutated to the positively charged arginine (R), which we hypothesized might enhance the dCas12f interaction with the negative charged genomic DNA and/or sgRNA.

Figure 1. RNA engineering and iterative protein engineering to generate efficient dCasMINI variants for reporter gene activation from the type V-F CRISPR

(A) Schematic construct designs for testing dCas12f-VPR for CRISPR activation (CRISPRa). TRE3G-GFP HEK293T reporter cell line was used to measure GFP activation efficiency of dCas12f-VPR by flow cytometry 48 h post-transfection. Two mutations were introduced in the RuvC domain of Cas12f to generate nuclease-deactivated dCas12f. The sgRNA targets the seven repeats in the TRE3G promoter with a TTTR PAM.

(B) Performance of GFP activation as measured using flow cytometry. Representative histograms of targeting (sgTet) and non-targeting sgRNAs (sgNT) show the percentages of GFP-positive population and that dCas12f-VPR fails to activate GFP expression.

(C) Schematic of strategies for sgRNA engineering. design 1, G-U swap; design 2, stem-loop truncation; design 3, 5′ poly G removal.

(D) Performance of GFP activation by transfecting the TRE3G-GFP HEK293T reporter line with different sgRNA designs shown in (C). Left: bars represent percentage of GFP+ cells. Right: bars represent mean GFP fluorescence values. Dots represent three biological replicates. Dotted line, mean value of non-targeting sgRNA group. Fold changes are calculated relative to the non-targeting sgRNA. a.u., arbitrary units. Data are representative of three biological replicates.

(E) Overview of the iterative protein engineering strategy. The best dCas12f-VPR variant for GFP activation is used as the starting sequence for the next round of screening.

(F–H) Performance of engineered variants as measured by GFP expression in four rounds of screening. The best variants in each round are named "dCasMINI-V1," "dCasMINI-V2," "dCasMINI-V3," and "dCasMINI-V4" as indicated. Dots represent three biological replicates. The fold of GFP activation was calculated by normalizing to the non-targeting sgRNA. Data are representative of three biological replicates.

(I) Gradual improvement of dCasMINI-VPR-mediated GFP activation. The fold change of each group is calculated by normalizing gene activation to the wild-type dCas12f-VPR with the design 1 sgRNA scaffold. Data are representative of two independent experiments with three biological replicates for each experiment.

(J) Sequence alignment between Cas12f and representative Cas12a proteins. The enhancing residues and the conserved motif (D/E)XRKN are indicated.

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

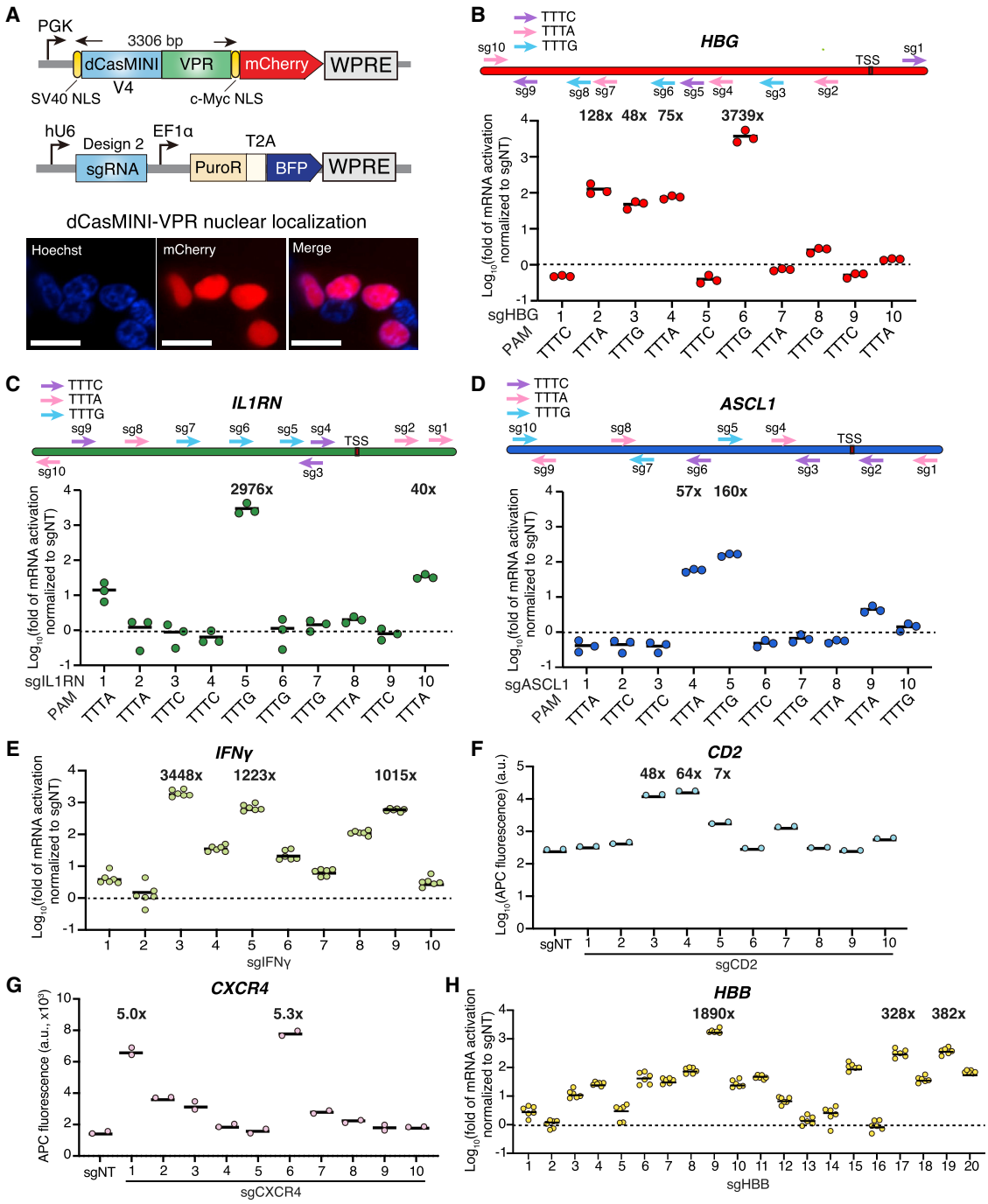


Figure 2. Endogenous gene activation using dCasMINI-VPR in mammalian cells

(A) Top: schematic of constructs used for the dCasMINI-VPR activation system. All experiments used sgRNA design 2. Bottom: confocal microscopy images show the expression and nuclear localization of dCasMINI-VPR fused to mCherry. Nuclei are stained using Hoechst 33342. Scale bars, 20 μm.

(B–D) Gene activation using dCasMINI-VPR and individual sgRNAs with different PAMs (TTTV) on three endogenous genes (*HBG*, *IL1RN*, and *ASCL1*) in HEK293T cells measured using qPCR. All data shown are fold activation of mRNA by normalizing to the mRNA expression using a non-targeting sgRNA (sgNT). Top: schematic illustrating the sgRNA distributions and PAMs for each gene. Transcriptional start site (TSS) for each gene is shown. For PAMs, arrows represent the directionality of the designed sgRNAs; purple, TTTC; pink, TTTA; blue, TTTG. PAMs for each sgRNAs are also shown below the diagram. Fold activation for top sgRNAs is shown. Dots represent individual biological replicates, and black bars represent mean values. Dotted lines represent the activation level of sgNT (normalized to 1) for each gene. Data are representative of three biological replicates.

(E–H) Characterization of dCasMINI-VPR-mediated activation in HEK293T cells using a library of sgRNAs for human endogenous genes *IFNγ*, *CD2*, *CXCR4*, and *HBB*. Ten sgRNAs were tested for *IFNγ* (E), *CD2* (F), and *CXCR4* (G), and 20 sgRNAs were tested for *HBB* (H). (E) and (H) show fold of mRNA activation measured

(legend continued on next page)

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

Although most variants showed no improvement of activation over the wild-type dCas12f-VPR, a few variants (D143R, T147R, E151R, and E241R) significantly enhanced activation (Figure 1F, red groups in Figure S3A). The most effective variant, D143R, showed 123-fold gene activation compared with a non-targeting sgRNA, which was more than 34-fold improvement over the wild-type dCas12f-VPR (Figure 1F; Figures S3B and S3C). We termed this new, compact protein variant dCasMINI-V1.

For our second round of iteration, we mutated dCasMINI-V1 by additionally incorporating other mutations that demonstrated enhanced activity in our first round, creating a library of 11 double mutants. Using the first library, we observed that D143R/T147R, D143R/E151R, D143R/E241R, and D143R/E507R showed improvement over the D143R variant (Figure 1G, blue groups in Figures S3A and S3D). The most effective variant, D143R/T147R, showed 1.55-fold improvement in activation over dCasMINI-V1, which we named dCasMINI-V2. We also performed another mutational screen using dCasMINI-V1, mutating E151 to all other amino acids. Using the second library, we observed that other than R, serine (S), glycine (G), and alanine (A) substitutions also improved activation (Figure S4A), suggesting that small-size amino acid replacement might be important for enhanced protein activity at the E151 residue.

The third round of screen contained 13 triple variants based on dCasMINI-V2 (D143R/T147R). The D143R/T147R/K330R variant (dCasMINI-V3) outperformed other variants (1.26-fold over the best double variant, D143R/T147R; Figure 1H, green groups in Figures S3A and S3E). A fourth round of screen based on dCasMINI-V3 (D143R/T147R/K330R) testing a quadruple library yielded one variant, D143R/T147R/K330R/E528R (dCasMINI-V4), showing 1.14-fold improvement over the best triple variant (Figure 1H, yellow groups in Figures S3A and S3F).

The iterative protein engineering and screening yielded a gradually improved Cas12f variant library, with the D143R/T147R/K330R/E528R variant showing almost 200-fold improvement of reporter gene activation over the wild-type Cas12f (Figure 1I; Table S4). Interestingly, the single mutations at D143, T147, and E151 showing improved activity are near or in a (D/E)XRKN motif that is highly conserved in the Cas12a family (Figure 1J), suggesting that this domain could be important in regulating Cas12f-DNA interaction. Confirming this hypothesis, we aligned these mutations to very recently reported Cas12f structures and observed that they reside in the DNA binding pocket (Figures S4B and S4C; Takeda et al., 2021; Xiao et al., 2021). The variant D326A/D510A/D143R/T147R/K330R/E528R was chosen for further characterization for endogenous gene activation, which we refer to simply as "dCasMINI" hereafter.

dCasMINI-VPR can efficiently activate endogenous genes in mammalian cells

We next tested whether dCasMINI-VPR could activate endogenous genes. We first confirmed its nuclear localization via confocal microscope fluorescence imaging (Figure 2A). We then tested activation of endogenous genes including *HBG*, *IL1RN*, and *ASCL1* in HEK293T cells. For each gene, we designed a panel of ten sgRNAs with different PAMs (TTTA, TTTC, or TTTG) and binding orientations, targeting within 500 bp around the transcriptional start site (TSS) (Karvelis et al., 2020; Figures 2B–2D; Table S2). Testing across these sgRNAs showed that gene activation was highly dependent on the sgRNA targeting site. For all three genes, approximately 20%–40% of tested sgRNAs showed significant activation, with the best sgRNA activating its target gene by hundreds to thousands of folds of activation. TTTG and TTTA PAMs worked best, while TTTC PAM failed to show activation. Therefore, we conclude that TTTR PAMs enable highly efficient gene activation.

We further tested across a panel of endogenous genes, including *IFNγ*, *CD2*, *CXCR4*, and *HBB*, using the TTTR PAM. For each gene, we designed 10 sgRNAs (except for *HBB* with 20 sgRNAs; Figures 2E–2H). We observed a large portion of sgRNAs activated the target genes efficiently. The fold activation was more pronounced for silenced genes (*IFNγ*, *CD2*, *HBB*) in HEK293T cells, which was consistent with what has been reported for dCas9-mediated activation (Konermann et al., 2015).

dCasMINI-VPR outperforms dCas12f-VPR for endogenous gene activation and is comparable with dCas12a-VPR

We next compared side by side the performance of dCasMINI-VPR with the wild-type dCas12f-VPR for endogenous gene activation, with both effectors using the optimized sgRNA scaffold (design 2). We chose the top sgRNAs for the genes of *IFNγ*, *HBB*, *CD2*, and *CXCR4* from Figures 2E–2H and compared dCasMINI-VPR and dCas12f-VPR side by side (Figure 3A).

For all sgRNAs tested, we observed consistent and greatly enhanced activation for each gene with dCasMINI-VPR over dCas12f-VPR. For example, for *IFNγ* activation using two different sgRNAs, dCasMINI-VPR performed 45-fold or 120-fold better than dCas12f-VPR on the mRNA level measured by quantitative PCR (qPCR) or 25-fold or 7-fold better than dCas12f-VPR on the protein level measured by enzyme-linked immunosorbent assay (ELISA) (Figure 3B; see Table S2 for sgRNA sequence). When co-delivering both sgRNAs, dCasMINI-VPR showed even better activation improvement than dCas12f-VPR (300-fold by qPCR and 768-fold by ELISA). Similar improvement was observed for *HBB*, *CD2*, and *CXCR4* measured using qPCR or flow cytometry: dCasMINI-VPR showed up to 525-fold improvement over dCas12f-VPR for *HBB* activation, 64-fold improvement for *CD2* activation, and 11-fold improvement for *CXCR4* activation relative to dCas12f-VPR (Figures 3C–3E). The relative lower activation on CXCR4 was likely due to its high basal expression level in HEK293T cells.

using qPCR by normalizing to the mRNA expression using a non-targeting sgRNA (sgNT), and (F) and (G) show APC fluorescence by immunostaining the target protein measured using flow cytometry. Fold activation for top sgRNAs is shown. Dots represent individual biological replicates, and black bars represent mean values. For (F) and (G), the non-targeting sgNT data are shown. For (E) and (H), dotted lines represent the activation level of sgNT (normalized to 1). Data are representative of two biological replicates for (F) and (G) and are representative of three biological replicates with two technical replicates per biological replicates (six data points in total) for (E) and (H).

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

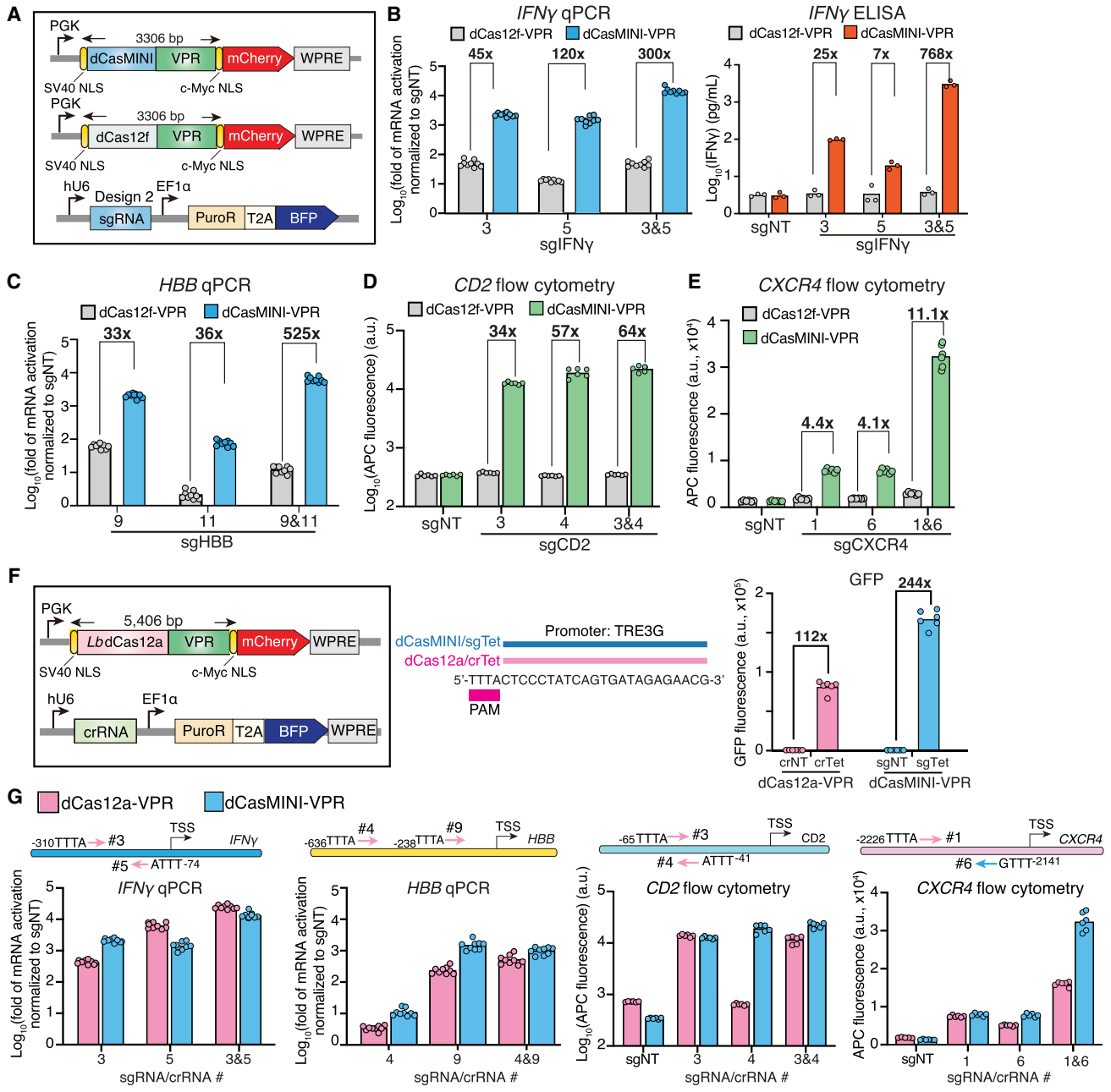


Figure 3. Comparison of CasMINI to Cas12f and Cas12a for endogenous gene activation

(A) Schematic of constructs used for dCasMINI-VPR, dCas12f-VPR, and sgRNA. All experiments used sgRNA design 2.

(B–E) Comparisons of endogenous gene activation using dCasMINI-VPR and dCas12f-VPR, including *IFNγ* (B), *HBB* (C), *CD2* (D), and *CXCR4* (E) in HEK293T cells. sgRNA numbers for each gene are indicated on the x axis (individually or in pools). For *IFNγ*, mRNA activation data measured using qPCR (left) and protein activation data measured using enzyme-linked immunosorbent assay (ELISA; right) are shown. For *HBB*, mRNA activation data measured using qPCR are shown. For *CD2* and *CXCR4*, protein activation data measured using immunostaining and flow cytometry are shown. The qPCR data show fold activation of mRNA by normalizing to the mRNA expression using a non-targeting sgRNA (sgNT). For other assays, absolute values for both targeting and sgNT are shown. a.u., arbitrary units. All qPCR data are representative of three biological replicates with three technical replicates per biological replicates (total of nine data points). All ELISA data are representative of three biological replicates. All flow cytometry data are representative of six biological replicates.

(F) Comparison of dCasMINI-VPR and dCas12a-VPR for GFP activation in TRE3G-GFP HEK293T cells measured using flow cytometry using an sgRNA (for CasMINI) or a crRNA (for Cas12a) targeting the same site. Left: constructs of dCas12a-VPR and crRNA. Middle, the target sequence of sgRNA and crRNA on the TRE3G promoter. The spacer and the PAM sequence are highlighted. Right: GFP flow cytometry data using targeting and non-targeting guides for dCas12a-VPR and dCasMINI-VPR. Data are representative of six biological replicates.

(legend continued on next page)

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

We next compared dCasMINI to *Lachnospiraceae bacterium*-derived dCas12a, a large Cas effector (1,228 amino acids) that is more than twice the size of dCasMINI (Figure 3F). We chose Cas12a because it shares an overlapping PAM (TTTV) with CasMINI (TTTR), making it possible to directly compare the performance of the two systems side by side using the guide RNAs targeting the same genomic sequence. We targeted our TRE3G-GFP reporter, along with three endogenous genes (*IFNγ*, *HBB*, and *CXCR4*) and designed Cas12a crRNAs binding to the same target sequences of the best performing sgRNAs of dCasMINI. For GFP activation, dCasMINI-VPR outperformed dCas12a-VPR by 2-fold (244-fold versus 112-fold activation; Figure 3F). For most tested endogenous genes, we observed that dCasMINI-VPR outperformed dCas12a-VPR, suggesting that the dCasMINI-VPR system is comparable with dCas12a-VPR for gene activation (Figure 3G).

dCasMINI is specific in mammalian cells and is comparable with dCas12a

To test whether dCasMINI-mediated activation is specific in the mammalian genome context, we performed whole-transcriptome RNA sequencing (RNA-seq). As it has been demonstrated that the Cas12a system is highly specific in mammalian cells (Kim et al., 2016), we performed side-by-side comparison of dCasMINI-VPR and dCas12a-VPR. To do this, we prepared HEK293T cells transfected with the respective effector and a targeting or non-targeting guide RNA. For each condition, both biological replicates showed consistent RNA-seq profiling (Figures S5A and S5B). We found that the correlation in gene expression between the targeting and the non-targeting guides was highly consistent among the biological replicates for each effector (Figures 4A and 4B). Comparison between dCasMINI-VPR and dCas12a-VPR for different guides also showed a high correlation (Pearson correlation coefficient > 0.99; Figure S5B), indicating similar specificity profiles of the two effectors. Overlaying the RNA-seq data of dCas12a-VPR and dCasMINI-VPR (two duplicates shown) demonstrated dCasMINI-VPR activated GFP with higher efficiency (Figure 4C). Comparing the SD distribution of both datasets also confirmed that the two Cas effectors had similar variation profiles for gene activation, suggesting similar specificities between the two systems (Kim et al., 2016; Figure 4D). These data together suggested the high specificity of using dCasMINI in mammalian cells.

dCasMINI-ABE allows base editing in mammalian cells

We next tested whether CasMINI could be used for other genome engineering applications. We focused on base editing, as previously developed base editors using Cas9 or Cas12 are too large to fit into the packaging capacity of AAV (<4.7 kb) (Li et al., 2018; Richter et al., 2020). On the contrary, with reduced size of CasMINI, a base editor can reasonably fit within the size limit. We first generated different designs by fusing dCasMINI to the previously reported deoxyadenosine deaminase TadA-8e (TadA*) domain or to a heterodimer TadA-TadA* (designs 1–4 in Figure 5A; Huang et al., 2021; Richter et al., 2020). We measured A·T to G·C conversion efficiency using high-throughput sequencing (HTS) using these designs at three genomic sites (Figure 5B; see STAR Methods; Tables S6 and S7). Among these protein designs, design 4 with the TadA-TadA* fusion outperformed others. We next compared the frequency of A·T to G·C conversion using dCasMINI-ABE design 4 (~3.0 kb) and dCas12a-ABE (~4.5 kb) side by side at the same genomic sites and found that the two systems exhibited similar editing efficiency across these sites (Figures S6A and S6B).

We next characterized the performance of this fusion for A·T to G·C base editing at a total of 12 genomic sites, including multiples sites in vicinity regions of IFNγ, HBB, and VEGFA loci. For many genomic sites, we observed detectable A·T to G·C base conversion (Figure 5C; Figure S6C). The base editing efficiency was dependent on the target site, and we further analyzed the pattern for A·T to G·C conversion. Interestingly, we observed that most efficient A·T to G·C editing occurred in a narrow window A3-A4 (3–4 bp downstream of the PAM; the "R" in the TTTR PAM is position "0") (Figure 5D), suggesting that careful sgRNA target design is needed for efficient base editing.

The nuclease-active CasMINI enables robust gene editing in mammalian cells

We next asked whether nuclease-active versions of dCasMINI variants (CasMINI) could cut and edit genomic DNA in human cells. We compared CasMINI-V2 (D143R/T147R), V3.1 (D143R/T147R/E151A), and V4 (D143R/T147R/K330R/E528R) side by side with the wild-type Cas12f (Figure 6A). We included the V2 and V3.1 variants, as we suspected that the proximity of the K330R and E528R mutations to the catalytic sites in the RuvC domains might negatively affect the DNA cleavage ability of CasMINI-V4. Using the sgRNA design 2, we tested all variants by targeting four selected sites in the VEGFA genomic locus and measured indel (insertion/deletion) formation efficiency via deep sequencing. Interestingly, we observed that CasMINI-V3.1 outperformed V2 and V4, which showed consistently higher indel formation across all tested sites (Figure 6B). We note that the use of our optimized sgRNA (design 2) also enabled modest indel formation with the wild-type Cas12f in mammalian cells, which has not been observed before.

To further characterize gene editing using CasMINI-V3.1, we quantified the indel formation efficiency at four additional genomic sites in *HBB* or *IFNγ* (Figures 2E and 2H). We observed robust gene editing using CasMINI-V3.1 at these sites, which was more efficient than the wild-type Cas12f or CasMINI-V2 (Figure 6C; Figure S7). These data suggest that the CasMINI variants enabling optimal gene editing can be different from those used for best gene activation.

(G) Comparison of dCasMINI-VPR and dCas12a-VPR for activation of endogenous genes in HEK293T cells, including *IFNγ*, *HBB*, *CD2*, and *CXCR4*, using sgRNAs and crRNAs targeting the same genomic sites. The top schematics illustrate the targeting sites of sgRNAs or crRNAs for each gene. The PAM and the genomic position of the first "T" in PAM (relative to TSS, which is "0") are shown for each site. For qPCR data, fold activation of mRNA is shown by normalizing to the mRNA expression using sgNT. For flow cytometry data, APC fluorescence values for both targeting and sgNT are shown. a.u., arbitrary units. The qPCR data are representative of three biological replicates with three technical replicates per biological replicates (nine data points in total for each group). The flow cytometry data are representative of six biological replicates.

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

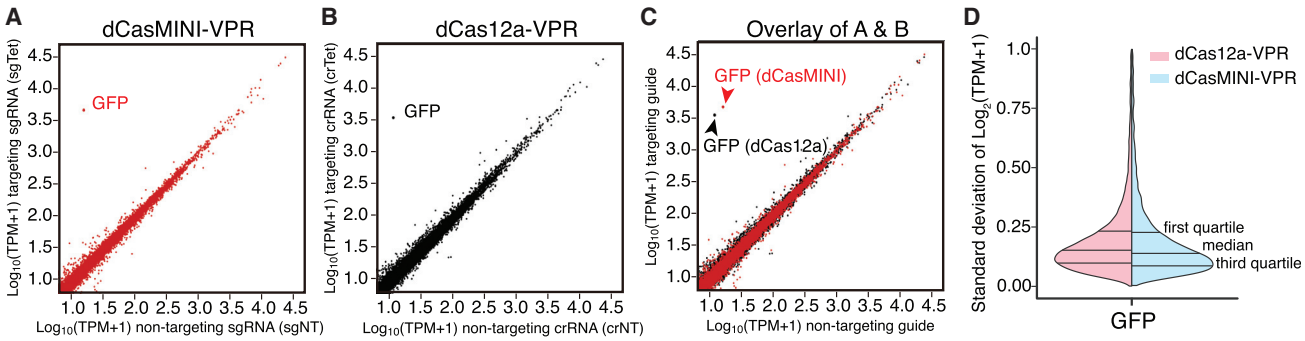


Figure 4. Characterization of off-target effects of dCasMINI-VPR in mammalian cells

(A and B) RNA-seq data of HEK293T cells transfected with an sgRNA targeting TRE3G (sgTet or crTet) versus a non-targeting sgRNA (sgNT or crTet) for dCasMINI-VPR (A) and dCas12a-VPR (B). The data points for GFP transcripts are labeled. TPM, transcripts per million mapped reads. Data represent mean values of two biological replicates.

(C) An overlay of (A) and (B). Red, dCasMINI-VPR data; black, dCas12a-VPR data. Data represent mean TPM values of two biological replicates.

(D) Distribution of SDs for log2(TPM + 1) values of all genes in RNA-seq library among targeting and non-targeting replicates for each gene for dCasMINI-VPR (blue) and dCas12a-VPR (red), respectively.

We further analyzed the indel patterns formed by the CasMINI variants by averaging the indel length at top genomic sites. Compared with the wild-type Cas12f, CasMINI-V3.1 showed larger deletions (around 20 bp), which were also larger than that reported for Cas9 (Figure 6D; Kleinstiver et al., 2019; Strecker et al., 2019). We also looked at indel formation frequency at each nucleotide position. Interestingly, we observed major indel editing at the PAM-distal region spanning outside of the sgRNA-binding sequence (Figure 6E). Previous *in vitro* assays showed that Cas12f cleavage predominantly centered around positions 20–24 bp relative to the PAM sequence (Karvelis et al., 2020). Consistently, our results using CasMINI showed that *in vivo* gene editing also peaked around positions 20–30 bp relative to PAM (Figure 6E). We thus confirm that CasMINI can be used in broad genome editing applications in addition to gene activation.

DISCUSSION

In summary, we engineered a compact and efficient CRISPR-Cas effector, termed CasMINI, derived from the naturally occurring type V-F Cas12f (Cas14) system. We showed that whereas the natural Cas12f and its sgRNA had no activity in mammalian cells, the engineered Cas12f protein variants combined with engineered sgRNAs exhibited efficient gene regulation and gene editing activity. The efficiency of the engineered dCasMINI effector is comparable with the dCas12a system, when tested for gene activation or base editing. Via RNA-seq, we observed no significant off targets. We also showed that the nuclease-active CasMINI allowed efficient indel formation. This provides a new method to engineer compact and efficient CRISPR-Cas effectors that can be useful for broad genome engineering applications, including gene regulation, gene editing, base editing, epigenome editing, and chromatin imaging (Figure 7A; Klann et al., 2017; Konermann et al., 2015; Nakamura et al., 2021; Wang et al., 2019b; Wang et al., 2018).

Our engineered CasMINI variants showed significantly improved gene activation or editing activity over the wild-type Cas12f, suggesting that enhancing interactions in the Cas12f:sgRNA:DNA complex is important for both applications. However, we note that a more efficient gene activator (e.g., V4 compared with V3.1) may not necessarily be more efficient for gene editing. Engineering the residues in the RuvC domain may further increase the editing efficiency.

For all applications, we consistently observed that the performance of CasMINI was highly dependent on the sgRNA targeting sites. Chromatin features including compactness, accessibility, epigenetic status, and nearby DNA-binding proteins may all influence the activity of CasMINI. Furthermore, for dCasMINI-ABE base editing, we observed that target sites containing A3-A4 in the sgRNA protospacer exhibited more efficient editing, suggesting that a proper adenosine window is crucial for optimal base editing. Therefore, characterizing and choosing the best sgRNAs will be an important consideration to achieve desired gene regulation or editing effects using the CasMINI system.

The CasMINI system may have different gene editing or gene regulation features compared with Cas9 or Cas12a. When we analyzed the indel pattern using CasMINI for gene editing, we observed the editing predominantly occurred distal (20–30 bp) to the PAM sequence with a large deletion size, which is distinct from Cas9-mediated indel patterns (Strecker et al., 2019). For base editing, Cas9 can be mutated into a nickase (nCas9), which enhances the base editing efficiency (Komor et al., 2016). However, like other Cas12a proteins, CasMINI cannot be easily transformed into a nickase, as it uses the RuvC domain for cutting both DNA strands. The use of dCasMINI for base editing may have a lower base editing efficiency compared with a nickase but may also avoid DNA nicking-induced indels. Very recent structural studies further suggested that Cas12f forms a dimer when binding to the target DNA (Takeda et al., 2021; Xiao et al., 2021). It is possible that a dimer of dCasMINI activator is recruited to each target site which enhances activation.

The RNA and protein engineering approach used in this work may be applicable to engineer more Cas12f/Cas14 effectors from other bacterial or archaeal species. Previous studies have used protein engineering to generate enhanced Cas12a (enAs-Cas12a) or Cas12b variants for improved editing activity (Kleinstiver et al., 2019; Strecker et al., 2019). Our work further

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

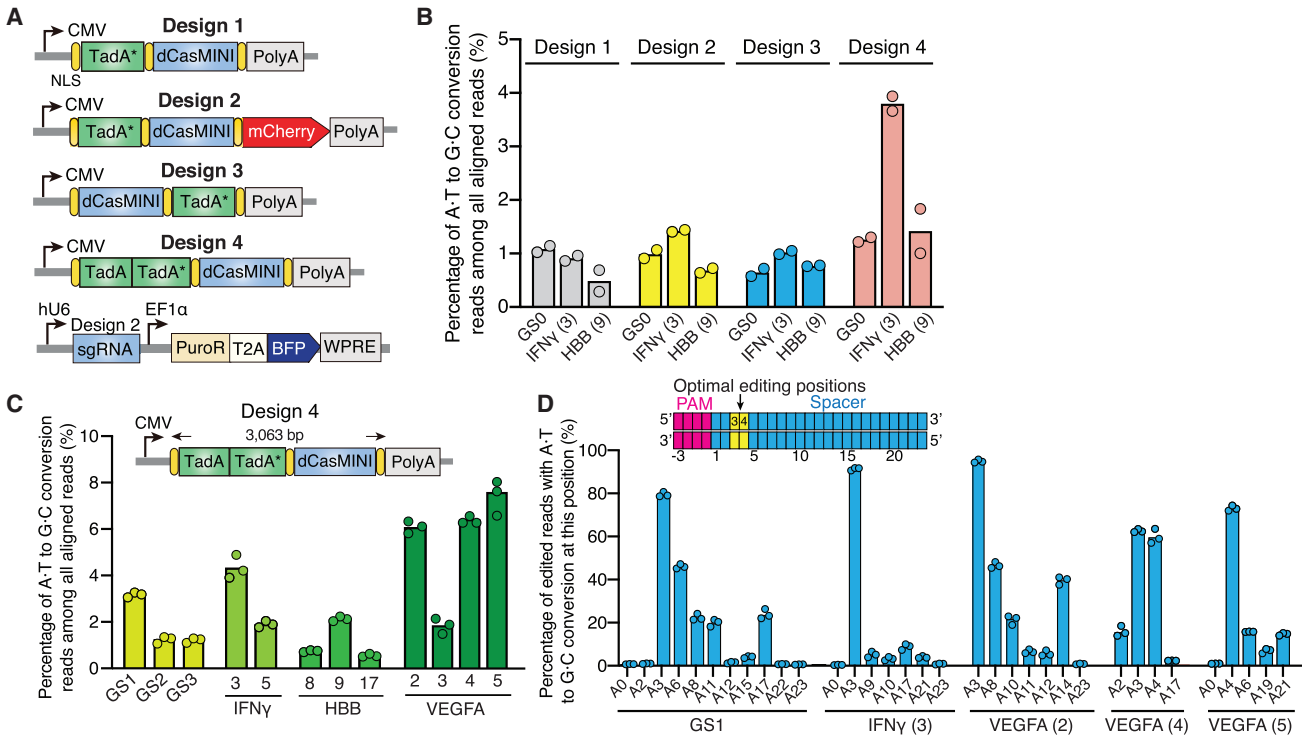


Figure 5. dCasMINI-mediated A·T to G·C base editing in mammalian cells

(A) Schematic of constructs for four designs by fusing TadA-8e (TadA*) to dCasMINI at the N terminus without (design 1) or with (design 2) mCherry, fusing TadA* to dCasMINI at the C terminus (design 3), or fusing a heterodimer TadA-TadA* to dCasMINI at the N terminus (design 4). The construct of sgRNA is shown on the bottom. All experiments used sgRNA design 2.

(B) Comparison of four dCasMINI-ABE designs for base editing efficiencies at three different genomic sites in HEK293T cells. The data shown are the percentage of reads with A·T to G·C conversion over the total aligned reads using deep sequencing. Data are representative of three biological replicates. GS0, genomic site 0. Bars represent mean values and points represent two independent biological replicates.

(C) Base editing efficiencies in HEK293T cells of more genomic sites with dCasMINI-ABE design 4, including two sites in the IFNγ locus, three sites in the HBB locus, and four sites in the VEGFA locus. The data shown are the percentage of reads with A·T to G·C conversion over the total aligned reads using deep sequencing. GS1–3, genomic sites 1–3. Bars represent mean values and points represent three independent biological replicates.

(D) A·T to G·C conversion base editing frequency in HEK293T cells by dCasMINI-ABE at adenines for five sites. The schematic of the nucleotide position is shown on the top: the "R" in TTTR PAM is position "0." The highlighted yellow boxes represent the observed most efficient A·T to G·C conversion positions (positions 3 and 4). The data shown are the number of reads with A·T to G·C conversion at a specific position over the total number of reads for A·T to G·C conversion using deep sequencing. GS1, genomic site 1. Bars represent mean values and data represent three independent biological replicates.

demonstrates that it is possible to engineer efficient Cas12f effectors starting from an initial system with no detectable activity in mammalian cells. These results likely suggest that many systems in the Cas12 family could be optimized for better efficiency via protein and guide RNA engineering.

The large size of CRISPR-Cas effectors and their fusion proteins has posed a challenge for efficient cell engineering and *in vivo* delivery (Doudna, 2020; Wang et al., 2020; Zhang, 2019). The size of the engineered CasMINI molecule is 529 amino acids, which is 62% and 57% smaller than the commonly used SpCas9 (1,368 amino acids) and LbCas12a (1,228 amino acids), respectively (Figure 7B). This small size makes it suitable for a wide range of therapeutic applications. We analyzed fusions of CasMINI to widely used repressors, activators, and gene editing domains and observed that all of them were below the AAV packaging limit (<4.7 kb; Figure 7C). For example, the CasMINI fusion proteins (3.3 kb for dCasMINI-VPR or 3.0 kb for dCasMINI-ABE tested here) are well suited for AAV packaging. Furthermore, it can possibly enhance the delivery efficiency if using lipid nanoparticles (LNPs) to carry mRNA payloads for RNA therapy, as the mRNA payload size is important for effective entry into cells. We also hypothesize that its small size and non-human pathogen source make it likely less immunogenic compared with large protein payloads (Charlesworth et al., 2019; Wagner et al., 2019). Therefore, we envision that these synthetic compact Cas effectors developed in this study will be broadly useful for gene therapy and cell engineering applications.

LIMITATIONS OF THE STUDY

Future work is needed to further optimize the efficiency for base editing and gene editing by testing more protein variants and sgRNA variants. As many genome engineering applications including epigenome editing often require large protein fusions (e.g., epigenome editing usually requires fusing with 2–3 kb of epigenetic factor domains such as DNMT3A or DNMT3L), the

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

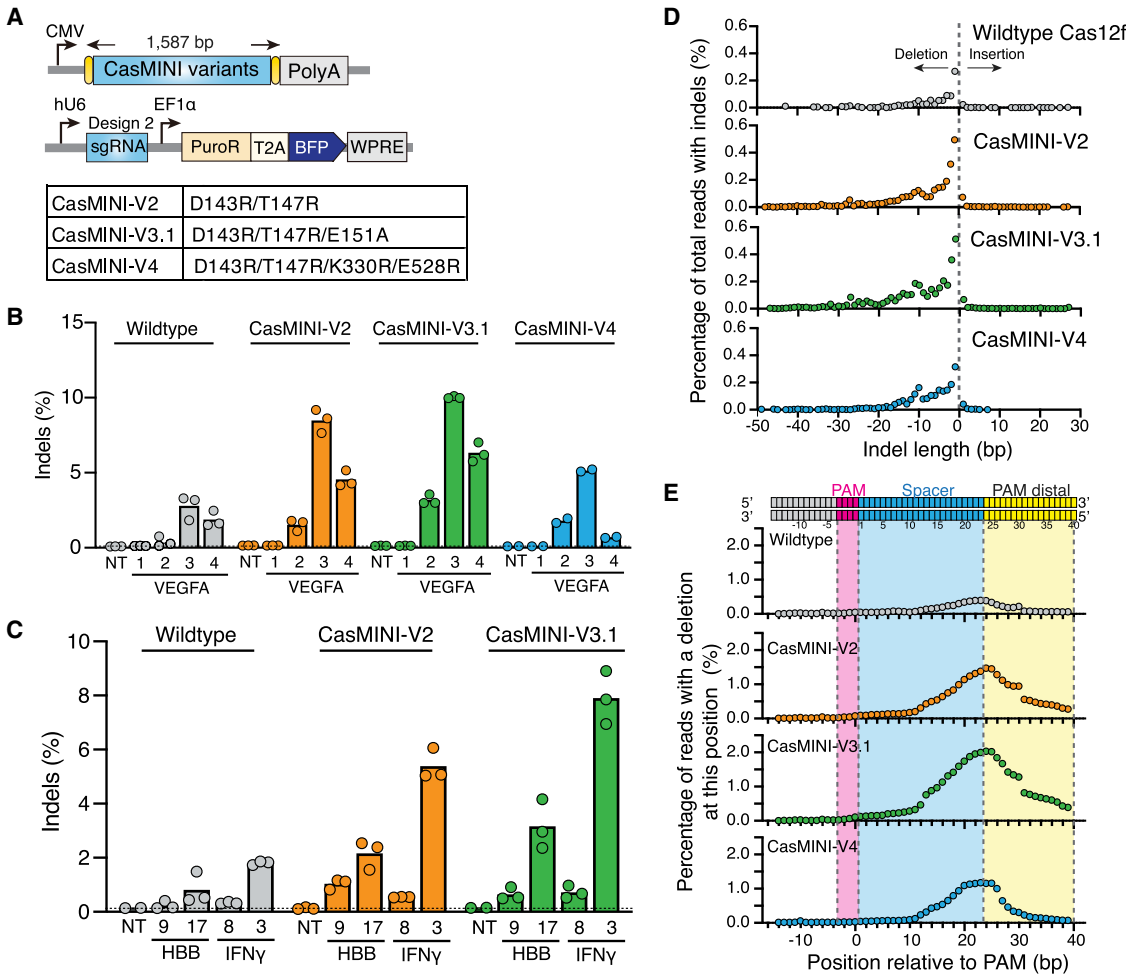


Figure 6. Genome editing using the nuclease-active CasMINI in mammalian cells

(A) Schematic of constructs encoding the nuclease-active CasMINI and its sgRNA for gene editing. All experiments used sgRNA design 2. The table shows three CasMINI variants tested.

(B) Indel activity of each CasMINI variant at four sites of the VEGFA locus measured by deep sequencing in HEK293T cells. The data using a non-targeting (NT) sgRNA are shown as a representative negative control. The dotted line shows the basal indel level detected from wild-type HEK293T cells. Bars represent mean values, and data represent three independent biological replicates.

(C) Indel activity of the wild-type Cas12f, CasMINI-V2, and CasMINI-V3.1 at two sites of the HBB and IFNγ loci in HEK293T cells. The dotted line shows the data using a sgNT as a representative negative control. Bars represent mean values and data represent three independent biological replicates.

(D) Largest indel length during genome editing over eight distinct sites (except for V4, which has four active sites). The data represent the percentage of aligned reads with an insertion or deletion of the given length.

(E) Indel activity at each nucleotide position during genome editing over eight distinct sites (except for V4, which has four active sites). The data represent the percentage of total reads with a deletion at the position. The schematic on the top show the PAM (4 bp) and the spacer (23 bp), which is aligned to each nucleotide position. The "R" in TTTR PAM is position "0."

CasMINI effector should be tested for optimal fusion design in these contexts. Furthermore, the principles for designing efficient sgRNAs at different chromatin regions for gene activation or gene editing remain unknown. Future work (e.g., via genome-wide screening) is needed to elucidate such sgRNA design principles.

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 SUBJECT DETAILS
  - Cell culture
- METHOD DETAILS
  - Plasmid cloning
  - Flow cytometry and immunostaining
  - Confocal microscopy imaging

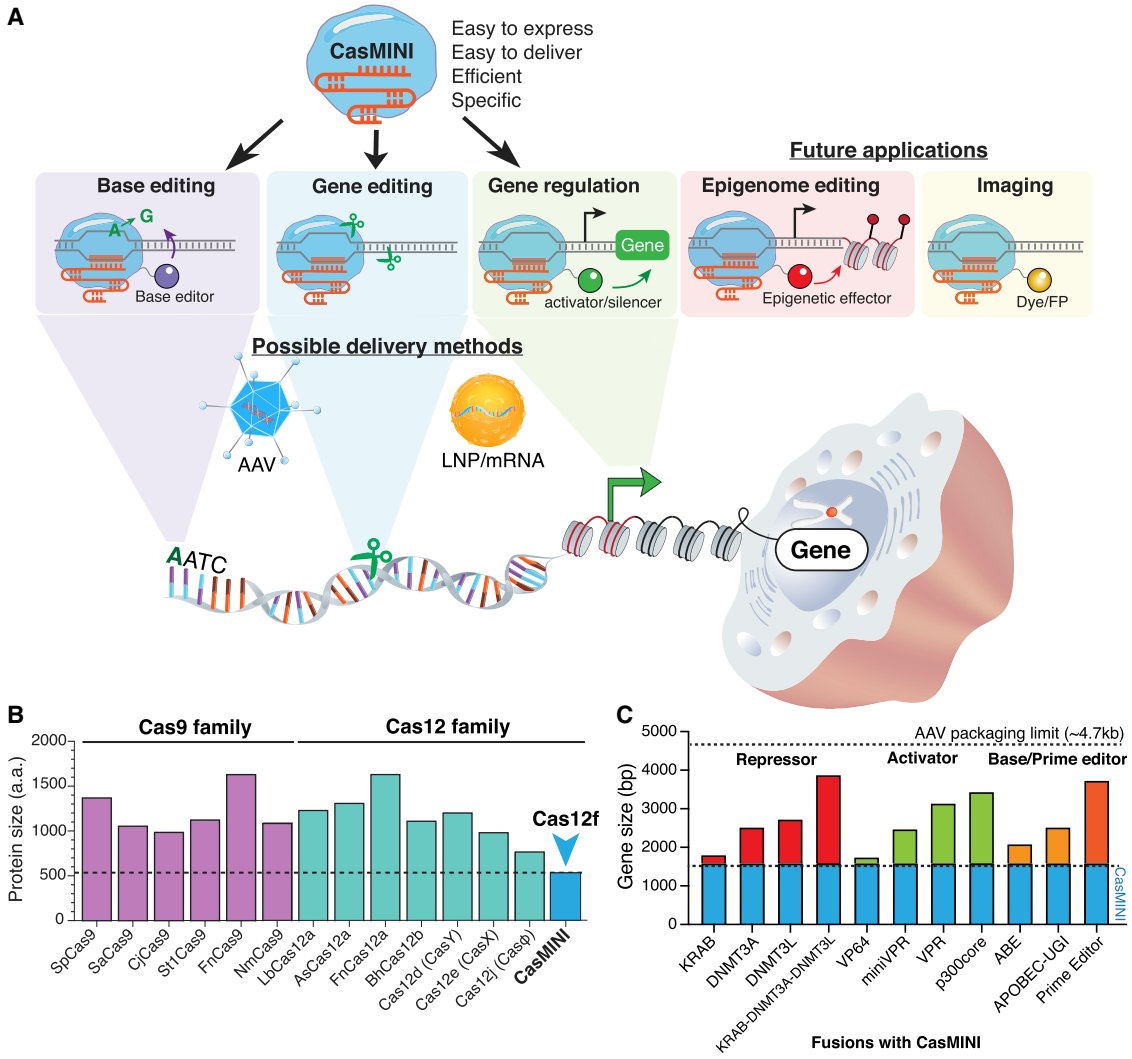


Figure 7. Overview of CasMINI for genome engineering applications

(A) In addition to the gene editing, base editing, and gene activation applications demonstrated in the study, CasMINI can be potentially used for other genome engineering applications, including epigenome editing and chromatin imaging. Its small size may facilitate high-efficiency delivery *ex vivo* or *in vivo* using AAV or LNP.

(B) Comparison of the protein size of CasMINI with other Cas effectors in the Cas9 family and the Cas12 family.

(C) Gene size of protein fusions between CasMINI and other effectors.

- ELISA
- Quantitative RT-PCR
- Transcriptome profiling of dCasMINI-VPR and dCas12a-VPR by RNA sequencing
- High-throughput sequencing (HTS) (deep sequencing) sample preparation
- Deep sequencing analysis

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.08.008.

ACKNOWLEDGMENTS

We thank the Stanley Qi lab members for technical support and comments on the manuscript. PLBH531_MBP-Cas14a1 was a gift from Jennifer Doudna (Addgene plasmids #112500). We thank the Cell Sciences Imaging Facility at Stanford University for confocal microscope usage. L.S.Q. acknowledges support from the Pew Scholar Foundation and the Alfred P. Sloan Foundation. The work is supported by a gift fund from the Li Ka Shing Foundation (L.S.Q.).

AUTHOR CONTRIBUTIONS

X.X. and L.S.Q. conceived of the idea. X.X. and L.S.Q. designed experiments. X.X., L.Z., and H.R.K. performed experiments. A.C., X.X., and L.S.Q. analyzed base editing and gene editing data. S.S., X.X., and L.S.Q. analyzed RNA-seq data for off-target analysis. X.X. and L.S.Q. analyzed other experimental

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

data. A.C. provided general computational analysis support. M.N. aided deep sequencing experiments. X.X. and L.S.Q. wrote the manuscript. All authors read and commented on the manuscript.

DECLARATION OF INTERESTS

L.S.Q. is a founder and shareholder of Epicrispr Biotechnologies and Refuge Biotechnologies. L.S.Q. is a scientific advisory board member of Epicrispr Biotechnologies and Refuge Biotechnologies. The authors have filed provisional patents via Stanford University related to the work (U.S. Provisional Patent Application Nos. 62/934,465 and 63/191,611).

Received: April 12, 2021
Revised: June 29, 2021
Accepted: August 5, 2021
Published: September 3, 2021

REFERENCES

Charlesworth, C.T., Deshpande, P.S., Dever, D.P., Camarena, J., Lemgart, V.T., Cromer, M.K., Vakulskas, C.A., Collingwood, M.A., Zhang, L., Bode, N.M., et al. (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. Nat. Med. *25*, 249–254.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat. Methods *13*, 563–567.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. *37*, 224–226.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Doudna, J.A. (2020). The promise and challenge of therapeutic genome editing. Nature *578*, 229–236.

Fellmann, C., Gowen, B.G., Lin, P.C., Doudna, J.A., and Corn, J.E. (2017). Cornerstones of CRISPR-Cas in drug discovery and therapy. Nat. Rev. Drug Discov. *16*, 89–100.

Gao, Y., Xiong, X., Wong, S., Charles, E.J., Lim, W.A., and Qi, L.S. (2016). Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. Nat. Methods *13*, 1043–1049.

Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., and Doudna, J.A. (2018). Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science *362*, 839–842.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Huang, T.P., Newby, G.A., and Liu, D.R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. Nat. Protoc. *16*, 1089–1128.

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.

Karvelis, T., Bigelyte, G., Young, J.K., Hou, Z., Zedaveinyte, R., Budre, K., Paulraj, S., Djukanovic, V., Gasior, S., Silanskas, A., et al. (2020). PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Res. *48*, 5016–5023.

Kempton, H.R., Goudy, L.E., Love, K.S., and Qi, L.S. (2020). Multiple input sensing and signal integration using a split Cas12a system. Mol. Cell *78*, 184–191.e3.

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat. Biotechnol. *34*, 863–868.

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2017). CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. *35*, 561–568.

Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., Scarfò, I., 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.

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.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., et al. (2018). Base editing with a Cpf1-cytidine deaminase fusion. Nat. Biotechnol. *36*, 324–327.

Nakamura, M., Gao, Y., Dominguez, A.A., and Qi, L.S. (2021). CRISPR technologies for precise epigenome editing. Nat. Cell Biol. *23*, 11–22.

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. Science *369*, 333–337.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Qu, G., Li, A., Acevedo-Rocha, C.G., Sun, Z., and Reetz, M.T. (2020). The crucial role of methodology development in directed evolution of selective enzymes. Angew. Chem. Int. Ed. Engl. *59*, 13204–13231.

Reetz, M.T., and Carballeira, J.D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nat. Protoc. *2*, 891–903.

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.

Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019). Engineering of CRISPR-Cas12b for human genome editing. Nat. Commun. *10*, 212.

Swarts, D.C., van der Oost, J., and Jinek, M. (2017). Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. Mol. Cell *66*, 221–233.e4.

Tak, Y.E., Kleinstiver, B.P., Nuñez, J.K., Hsu, J.Y., Horng, J.E., Gong, J., Weissman, J.S., and Joung, J.K. (2017). Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods *14*, 1163–1166.

Takeda, S.N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Nishimasu, H., and Nureki, O. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. Mol. Cell *81*, 558–570.e3.

Wagner, D.L., Amini, L., Wendering, D.J., Burkhardt, L.M., Akyüz, L., Reinke, P., Volk, H.D., and Schmueck-Henneresse, M. (2019). High prevalence of

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

Streptococcus pyogenes Cas9-reactive T cells within the adult human population. Nat. Med. *25*, 242–248.

Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipniss, N.H., La Russa, M., and Qi, L.S. (2018). CRISPR-mediated programmable 3D genome positioning and nuclear organization. Cell *175*, 1405–1417.e14.

Wang, D., Tai, P.W.L., and Gao, G. (2019a). Adeno-associated virus vector as a platform for gene therapy delivery. Nat. Rev. Drug Discov. *18*, 358–378.

Wang, H., Nakamura, M., Abbott, T.R., Zhao, D., Luo, K., Yu, C., Nguyen, C.M., Lo, A., Daley, T.P., La Russa, M., et al. (2019b). CRISPR-mediated live imaging of genome editing and transcription. Science *365*, 1301–1305.

Wang, D., Zhang, F., and Gao, G. (2020). CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. Cell *181*, 136–150.

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.

Xu, X., and Qi, L.S. (2019). A CRISPR-dCas toolbox for genetic engineering and synthetic biology. J. Mol. Biol. *431*, 34–47.

Xu, X., Chen, J., Wang, Q., Duan, C., Li, Y., Wang, R., and Yang, S. (2016). Mutagenesis of key residues in the binding center of l-aspartate-b-semialdehyde dehydrogenase from Escherichia coli enhances utilization of the cofactor NAD(H). ChemBioChem *17*, 56–64.

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.

Zhang, F. (2019). Development of CRISPR-Cas systems for genome editing and beyond. Q. Rev. Biophys. *52*, e6.

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-human CD2	BioLegend	Cat#309224; RRID:AB_2687219
Anti-human CXCR4	BioLegend	Cat#306510; RRID:AB_314616
Bacterial and virus strains		
Stellar Competent Cells	Takara	636766
Chemicals, peptides, and recombinant proteins		
Tris-HCl, pH 7.5, 1 M solution	Thermo Fisher Scientific	Cat# 15567027
SDS, 10% (wt/vol) solution	Thermo Fisher Scientific	Cat# 15553027
Fetal Bovine Serum (FBS)	Sigma-Aldrich	Cat# F0926
Hoechst 33342, Trihydrochloride, Trihydrate	Thermo Fisher	H3570
Critical commercial assays		
Human IFN-γ ELISA MAX Deluxe	Biolegend	430104
MiSeq Reagent Kit v3 (600-cycle)	Illumina	Cat# MS-102-3003
Deposited data		
All raw FASTQ files for RNaseq	This paper	GEO: GSE180773
All raw FASTQ files for MiSeq	This paper	NCBI BioProject: PRJNA748082
Experimental models: Cell lines		
HEK293T	ATCC	Cat# CRL-3216; RRID:CVCL_0063
Oligonucleotides		
Spacer and designed full sgRNA sequences	This paper; See Tables S2 and S3	N/A
Primers for qPCR and high-throughput sequencing	IDT; See Tables S5 and S6	N/A
Recombinant DNA		
pLBH531_MBP-Cas14a1 expression	Harrington et al., 2018	Addgene Plasmid # 112500
Plasmids generated in this study	This paper; See Tables S1 and S4	N/A
Software and algorithms		
FlowJo	FlowJo.LLC	https://www.flowjo.com
ImageJ	NIH	https://imagej.nih.gov/ij/
Prism 9	GraphPad	https://www.graphpad.com/
DESeq2	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
tximportData	Bioconductor	https://bioconductor.org/packages/release/data/experiment/html/tximportData.html
STAR	Github	https://github.com/alexdobin/STAR
CRISPResso2	Github	https://github.com/pinellolab/CRISPResso2
Code for CasMINI data analysis	This paper	https://zenodo.org/record/5140169
Other		
Q5 Hot Start High-Fidelity Mastermix, 2 ×	NEB	M0494S
DMEM, high glucose, GlutaMAX Supplement	Life Technologies	10569-044
0.05% Trypsin-EDTA	Life Technologies	25300120

(*Continued on next page*)

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
In-Fusion HD Cloning Plus	Takara	638911
T4 DNA Ligase	NEB	M0202L
Proteinase K	NEB	P8107S
TransIT-LT1 Transfection Reagent	Mirus	Cat#MIR 2306

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents and resources should be directed to the Lead Contact, Lei S. Qi (stanley.qi@stanford.edu).

Materials availability

Key constructs and plasmids are available on Addgene (https://www.addgene.org/Stanley_Qi/).

Data and code availability

- Raw RNA-seq data have been deposited at GEO and are available as of the date of publication. Accession numbers are listed in the Key resources table. Raw deep sequencing data for base editing and gene editing are available at NCBI BioProject: PRJNA748082. Microscopy data reported in this paper will be shared by the lead contact upon request.
- 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 SUBJECT DETAILS

Cell culture

Wild-type HEK293T cells (ATCC) and the HEK293T TRE3G-dscGFP (destabilized copGFP) reporter line (Gao et al., 2016) were cultured in DMEM with high glucose, sodium pyruvate and GlutaMAX (Thermo Fisher), additionally supplemented with 10% FBS (Sigma). Cells were grown at 37°C and 5% CO_2 and maintained at confluency below 80%. All transfections were performed with TransIT-LT1 transfection reagent (Mirus) at a ratio of 3 μL reagent per μg of plasmid per 100 μL volume of Opti-MEM reduced serum media (Thermo Fisher). Cells were plated in a 24-well plate with 50,000 cells in 500 μL culture medium per well one day before transfection. For GFP activation assays, 500 ng of dCas constructs and 250 ng sgRNA or crRNA plasmids were transfected to HEK293T TRE3G-GFP cells in 24-well plates. For endogenous gene activation, 800 ng of dCas plasmids and 500 ng sgRNA or crRNA plasmids were transfected to HEK293T cells in 24-well plates. The transfected cells were analyzed 3 days post transfection for endogenous gene activation or 2 days for GFP activation. For base editing or gene editing assays, cells were plated at 40,000 cells per well in 48-well plates and transfected using 750 ng of Cas (dCasMINI-ABE or dCas12a-ABE for base editing or CasMINI for gene editing) plasmids and 250 ng of sgRNA or crRNA plasmids. The transfected cells were harvested 3 days post transfection for deep sequencing analysis.

METHOD DETAILS

Plasmid cloning

Plasmids were cloned by standard molecular cloning techniques. The Cas12f sequence was amplified from plasmid Addgene #112500, and its sgRNA backbone fragments were ordered via gBlocks from Integrated DNA Technologies (IDT). The TadA-8e (TadA*) and TadA fragments were ordered via gBlocks from IDT. The dCas12f was generated by introducing two mutations (D326A and D510A) to the wild-type sequence. Nuclease-dead dCas12a and the corresponding crRNA backbone were used as previously published (Kempton et al., 2020). All Cas constructs were cloned using InFusion and Stellar competent cells (Takara Bio). All sgRNA and crRNA plasmids were cloned using T4 DNA Ligase (New England Biolabs). Oligos for targeting spacers were annealed and ligated into BsmBI digested backbone vectors. The plasmids generated in the work are in Table S1. The spacer sequences of sgRNAs used in the study are shown in Table S2. The four designed sgRNAs are in Table S3. The generated CasMINI protein sequences with efficient activities for gene activation, base editing, or gene editing are shown in Table S4.

Flow cytometry and immunostaining

To analyze fluorescent protein expression, cells were dissociated using 0.05% Trypsin EDTA (Life Technologies), resuspended in PBS with 5% FBS, and analyzed by flow cytometry on CytoFLEX S flow cytometer (Beckman Coulter). For analysis of cell surface

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

protein expression, cells were dissociated using Accutase (STEMCELL) and stained with labeled antibody in PBS with 5% FBS at 4°C for 30 min. Antibodies and relevant isotypes of CD2 and CXCR4 were purchase from BioLegend (#309224, #306510, #400122, #400220). At least 10,000 cells containing constructs of interest of each sample were analyzed using FlowJo. The analyzed cells were gated for positive fluorescent protein expression based on the non-transfected control corresponding to construct expression.

Confocal microscopy imaging

Confocal microscopic imaging was performed to visualize dCasMINI-VPR expression and nuclear localization. HEK293T cells transduced by dCasMINI-VPR lentivirus were seeded in a 96-well μ-plate (Ibidi, Inc). Cells were stained with Hoechst 33342 (Thermo Fisher Scientific) to label nucleus at 37°C for 10 min. Confocal microscopy was performed with a Nikon Spinning Disk Confocal microscope with TIRF.

ELISA

Supernatants from transfected cell cultures were harvest 3 days post transfection, and stored at −80°C. The secreted protein was quantified using the ELISA MAX Deluxe kits for human IFNγ on a Synergy H1 plate reader (BioTek). Absorbance at 450 nm and 570 nm was measured and protein concentrations were determined by the standard curve fitted to a power law.

Quantitative RT-PCR

The transfected cells as described above were harvested using Accutase (STEMCELL), and total RNA was extracted using RNeasy Plus Mini Kit (QIAGEN). cDNA was prepared using iScript cDNA Synthesis kit (Bio-Rad) and stored at −80°C. qPCR reactions were prepared in 384 well plates with iTaq Universal SYBR Green Supermix (Bio-Rad) and run on a CFX384 Touch Real-Time PCR thermocycler (BioRad). Any Cq values over 35 were considered to be 35, as there were fluctuations for transcripts with weak expression level. Samples transfected with non-targeting sgRNA or crRNA plasmids were used as negative controls. The relative expression fold-changes were analyzed using the ΔΔCq method. The levels of fold activation fold over negative controls were normalized to the expression of GAPDH. All the primers used are purchased from IDT and the sequences are listed in Table S5.

Transcriptome profiling of dCasMINI-VPR and dCas12a-VPR by RNA sequencing

The TRE3G-GFP HEK293T reporter cell line was transfected with the dCasMINI-VPR or dCas12a-VPR and sgRNA or crRNA plasmids and sorted based on the expression of fluorescence proteins (mCherry and BFP) 2 days post-transfection using a Sony SH800 Cell Sorter. Total RNA was isolated using RNeasy Plus Mini Kit (QIAGEN). RNA sequencing library preparation and next-generation sequencing were conducted by Novogene Corporation (Chula Vista, CA). The libraries were sequenced on a NovoSeq 6000 platform. Paired-end 150 bp reads were acquired and aligned to the hg38 genome with added GFP using STAR. Transcript abundances were estimated using STAR and htseq using the quantmode option. The counts were imported with tximport, and then normalized and statistically compared using DESeq2. hg38 annotations were downloaded from Gencode (Harrow et al., 2012). Custom R scripts were used to perform further TPM (transcripts per million mapped reads) normalization and quality control. Downstream plots used the ggplot2. The variation of dCas12a-VPR versus dCasMINI-VPR systems was represented as violin plots by considering the distribution of standard deviations for gene expression across the four replicates (two targeting and two non-targeting replicates). Linear models and Pearson correlation coefficients were obtained using QR decomposition and regression.

High-throughput sequencing (HTS) (deep sequencing) sample preparation

Cell lysate containing genomic DNA was prepared as described previously (Huang et al., 2021; Richter et al., 2020). Briefly, transfected cells were washed with 1x PBS (Thermo Fisher Scientific), followed by genomic DNA extraction by addition of 100 μL lysis buffer containing 10 mM Tris-HCl, pH 7.5, 0.05% SDS, and 0.1% (vol/vol) proteinase K (New England Biolabs). The genomic DNA lysate was incubated at 37°C for 1 hour, heat inactivated at 80°C for 30 min, and used for deep sequencing. Targeted genomic regions of interest were amplified with Q5 Hot Start High-Fidelity Mastermix, 2 × (NEB, # M0494S) using a two-round PCR strategy to add Illumina adaptors and unique barcodes for each sample. Libraries were sequenced with 1 × 200-cycle Mi-Seq runs (Illumina) as previously described (Richter et al., 2020). The primers used for generating amplicons are shown in Table S6. The amplicon sequences are shown in Table S7.

Deep sequencing analysis

For the gene editing analysis, CRISPResso2 was used to process fastq.gz files obtained from the Illumina sequencing run (Clement et al., 2019). The "–min_average_read_quality" flag was set to 30 to filter out reads with average phred33 quality scores less than 30. The quantification window center was set to −12 and quantification window size was set to 27; in cases where the window extended closer than 5 bp to either end of the amplicon, it was trimmed to 5 bp from that end. For each sample, the Alleles_frequency_table_around_sgRNA_*.txt file was used to quantify the indel percentages using the following procedure. First, all reads with 'N' in the quantification window were filtered out. For each read-amplicon alignment window, if there was a gap character in the aligned read window, it was counted as a read with a deletion; if there was a gap character in the aligned amplicon window, it was counted as a read with an insertion. For each read-amplicon alignment window, the deletion size was defined as the longest stretch of gap characters in the aligned read window and

Please cite this article in press as: Xu et al., Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing, Molecular Cell (2021), https://doi.org/10.1016/j.molcel.2021.08.008

the insertion size was defined as the longest stretch of gap characters in the aligned amplicon window. For quantifying deletions at each position, we define the position '0' as the base 'R' of the TTTR PAM in the read-amplicon alignment window.

For the base editing mutagenesis analysis, CRISPResso2 was again used to process fastq.gz files obtained from the Illumina sequencing run (Clement et al., 2019). The ''–min_average_read_quality'' flag was set to 30 to filter out reads with average phred33 quality scores less than 30. The quantification window center and quantification window size parameters were set to encompass the guide and PAM. For each sample, the Alleles_frequency_table.txt file was used to quantify the substitution percentages using the following procedure. For each position in each read-amplicon alignment, a substitution was counted if an ''A'' in the reference sequence was mutated to a ''G'' in the read sequence. If the guide was reverse complemented relative to the reference, then a substitution was counted if a ''T'' in the reference sequence was mutated to a ''C'' in the read sequence.

QUANTIFICATION AND STATISTICAL ANALYSIS

In all figure legends, n represents number of independent biological replicates, except for qPCR assays which show all technical replicates pooled from all biological replicates (3 technical replicates per biological sample). Statistical details of experiments such as values of n can be found in the figure legends. Prism 9 was used for the statistical analysis including calculation of mean values. For RNA-seq analysis, linear models and Pearson correlation coefficients were obtained using QR decomposition and regression. Custom R scripts were used to perform TPM (transcripts per million mapped reads) normalization and quality control.

Molecular Cell, Volume *81*

Supplemental information

Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing

Xiaoshu Xu, Augustine Chemparathy, Leiping Zeng, Hannah R. Kempton, Stephen Shang, Muneaki Nakamura, and Lei S. Qi

SUPPLEMENTAL FIGURES

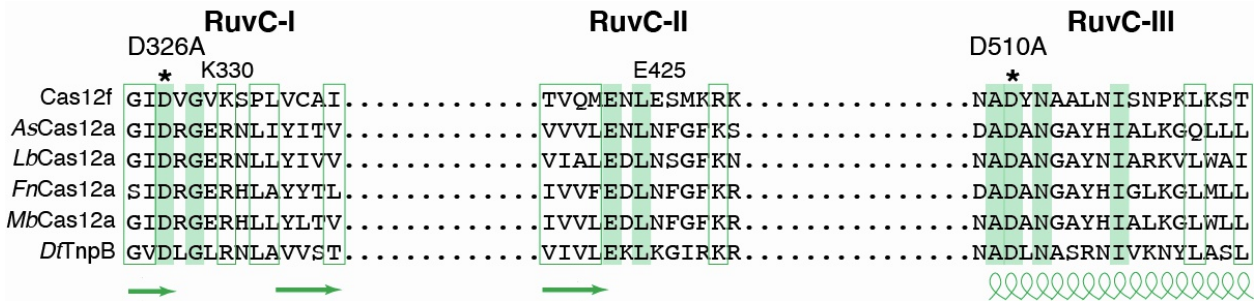


Figure S1. Alignment of Cas12f to the family of Cas12a proteins for conserved residues. Related to Figure 1A. Alignment of Cas12f to reported Cas12a proteins and DtTnpB. The conserved active residues of three RuvC domains are indicated. Highly conserved residues are shaded in green. Mutated RuvC residues to generate dCas12f are shown as asterisks.

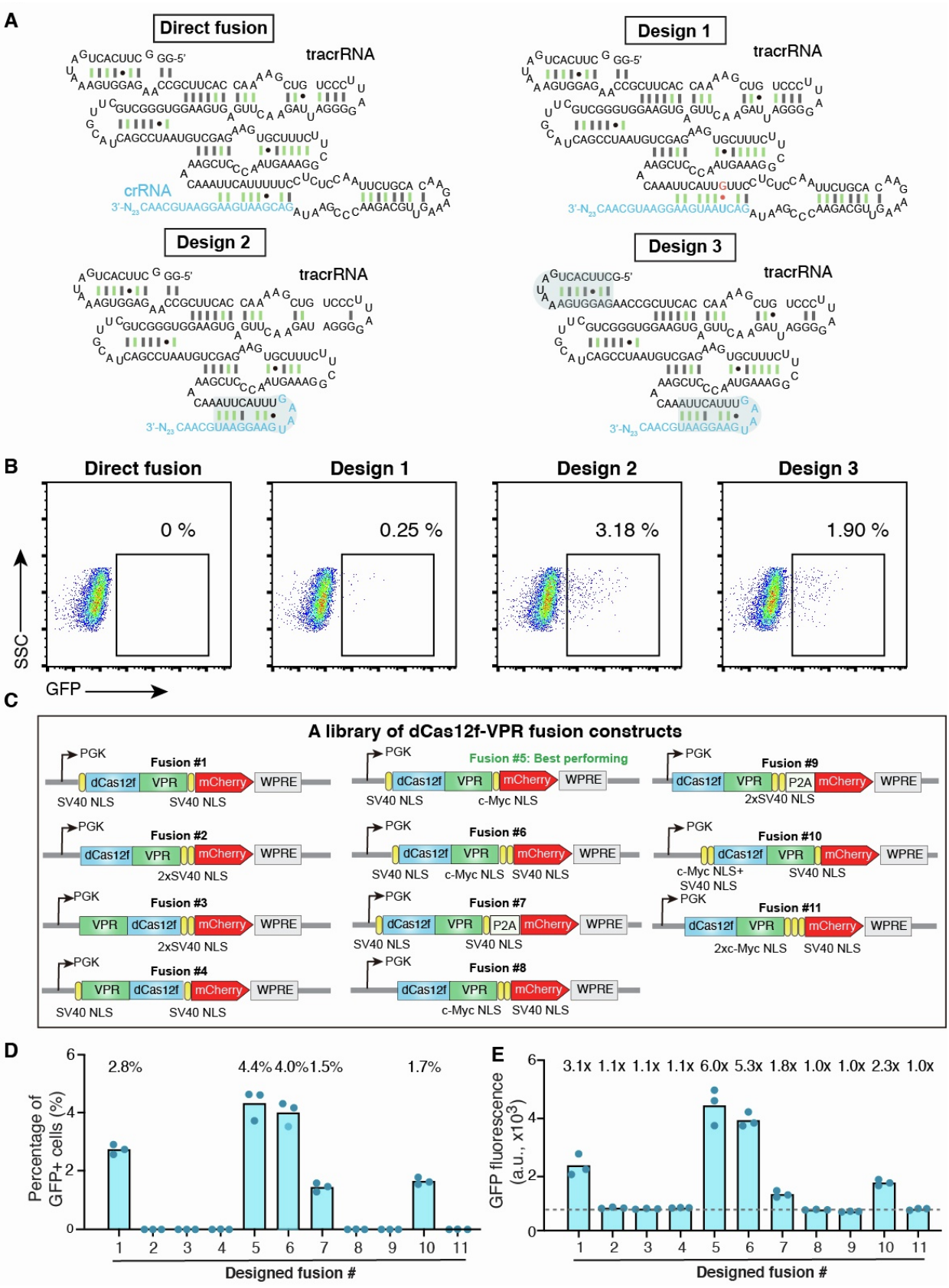

A
Direct fusion
tracrRNA
crRNA
Design 1
Design 2
Design 3
B
Direct fusion
Design 1
Design 2
Design 3
0 %
0.25 %
3.18 %
1.90 %
SSC
GFP
C
A library of dCas12f-VPR fusion constructs
PGK
dCas12f
VPR
mCherry
WPRE
P2A
Fusion #1
SV40 NLS
SV40 NLS
Fusion #2
2xSV40 NLS
Fusion #3
2xSV40 NLS
Fusion #4
SV40 NLS
SV40 NLS
Fusion #5: Best performing
SV40 NLS
c-Myc NLS
Fusion #6
SV40 NLS
c-Myc NLS
SV40 NLS
Fusion #7
SV40 NLS
SV40 NLS
Fusion #8
c-Myc NLS
SV40 NLS
Fusion #9
2xSV40 NLS
Fusion #10
c-Myc NLS+
SV40 NLS
SV40 NLS
Fusion #11
2xc-Myc NLS
SV40 NLS
D
Percentage of GFP+ cells (%)
2.8%
4.4%
4.0%
1.5%
1.7%
Designed fusion #
E
GFP fluorescence (a.u., x10^3)
3.1x
1.1x
1.1x
1.1x
6.0x
5.3x
1.8x
1.0x
1.0x
2.3x
1.0x
Designed fusion #

Figure S2. sgRNA engineering and protein fusion engineering to optimize Cas12f-mediated gene activation in mammalian cells. Related to Figure 1.

(A) Schematic of four designed sgRNAs. Design 1, G-U swap; Design 2, stem-loop truncation; Design 3, 5' poly G removal.

(B) Representative flow cytometry scatter plots showing the percentage of GFP+ cells for each sgRNA design.

(C) A library of dCas12f-VPR fusion designs (#1-#11) by fusing dCas12f to VPR at N or C terminus, with SV40 or c-MYC nuclear localization signals (NLSs), and with different linkers (P2A, glycine-serine linker).

(D) Characterization of GFP+ cells using different dCas12f-VPR fusions. The percentage of GFP+ cells for top performing designs are labeled. Bars represent the mean values, and dots represent three biological replicates.

(E) Characterization of GFP fold activation using different dCas12f-VPR fusions. The GFP fold activation by normalizing to the non-targeting sgRNA for top performing designs are labeled. Bars represent the mean values, and dots represent three biological replicates. The dotted line represents the GFP mean value of the non-targeting sgRNA.

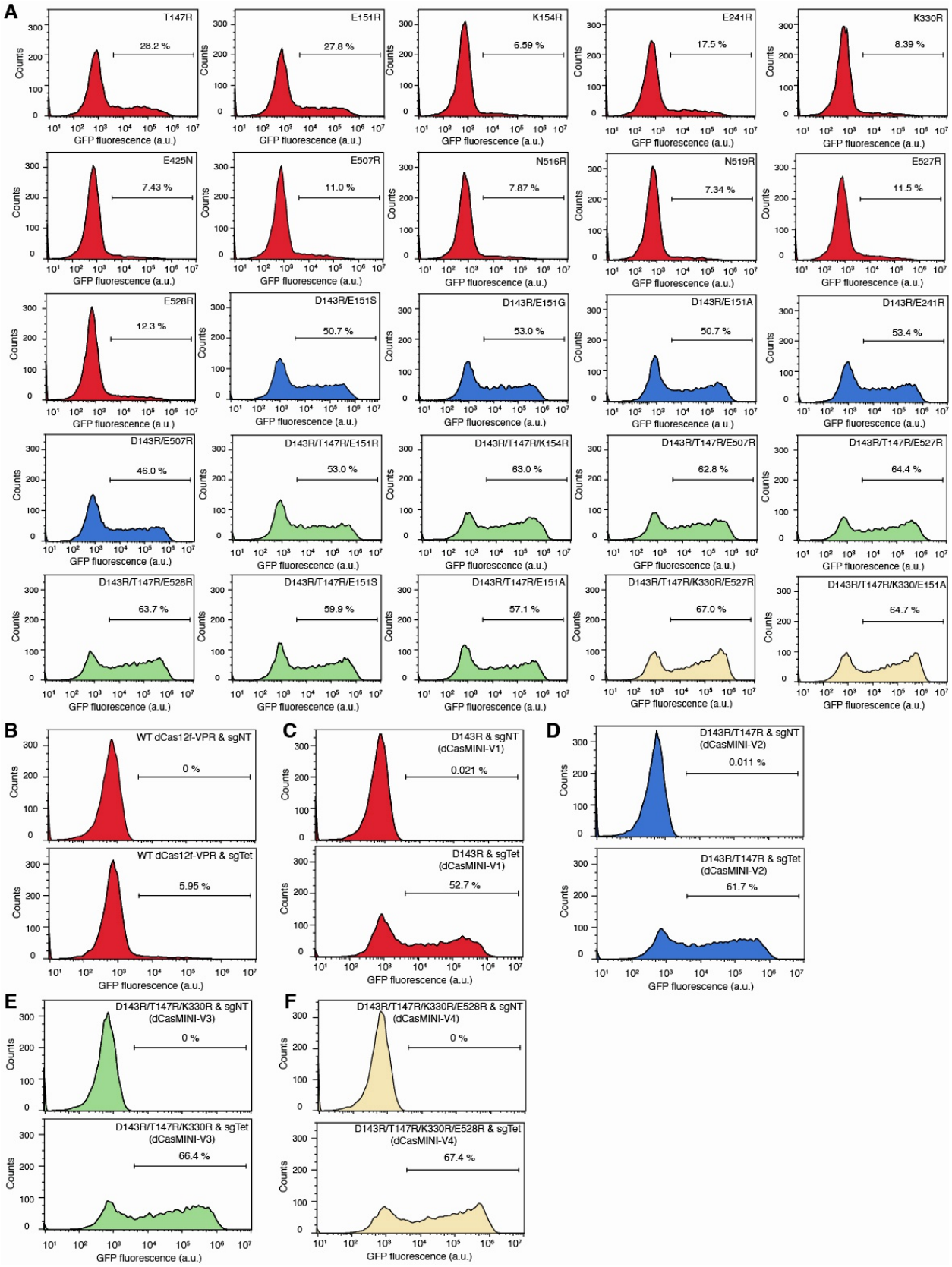

A
T147R
28.2 %
E151R
27.8 %
K154R
6.59 %
E241R
17.5 %
K330R
8.39 %
E425N
7.43 %
E507R
11.0 %
N516R
7.87 %
N519R
7.34 %
E527R
11.5 %
E528R
12.3 %
D143R/E151S
50.7 %
D143R/E151G
53.0 %
D143R/E151A
50.7 %
D143R/E241R
53.4 %
D143R/E507R
46.0 %
D143R/T147R/E151R
53.0 %
D143R/T147R/K154R
63.0 %
D143R/T147R/E507R
62.8 %
D143R/T147R/E527R
64.4 %
D143R/T147R/E528R
63.7 %
D143R/T147R/E151S
59.9 %
D143R/T147R/E151A
57.1 %
D143R/T147R/K330R/E527R
67.0 %
D143R/T147R/K330/E151A
64.7 %
B
WT dCas12f-VPR & sgNT
0 %
WT dCas12f-VPR & sgTet
5.95 %
C
D143R & sgNT
(dCasMINI-V1)
0.021 %
D143R & sgTet
(dCasMINI-V1)
52.7 %
D
D143R/T147R & sgNT
(dCasMINI-V2)
0.011 %
D143R/T147R & sgTet
(dCasMINI-V2)
61.7 %
E
D143R/T147R/K330R & sgNT
(dCasMINI-V3)
0 %
D143R/T147R/K330R & sgTet
(dCasMINI-V3)
66.4 %
F
D143R/T147R/K330R/E528R & sgNT
(dCasMINI-V4)
0 %
D143R/T147R/K330R/E528R & sgTet
(dCasMINI-V4)
67.4 %
Counts
GFP fluorescence (a.u.)

Figure S3. Raw data of flow cytometry showing GFP activation of functional variants for each round of screening and the best performing dCasMINI variants (V1-V4). Related to Figure 1.

(A) Representative flow cytometry histograms showing percentage of GFP+ cells for variants with single (red), double (blue), triple (green) and quadruple (tan) mutations. Values show the percentage of GFP+ cells.

(B-F) Representative flow cytometry histograms showing percentage of GFP+ cells for wildtype dCas12f-VPR (**B**), dCasMINI-V1-VPR (**C**), dCasMINI-V2-VPR (**D**), dCasMINI-V3-VPR (**E**), dCasMINI-V4-VPR (**F**). The non-targeting sgRNA (top) and targeting sgRNA data are shown for each group. Values show the percentage of GFP+ cells.

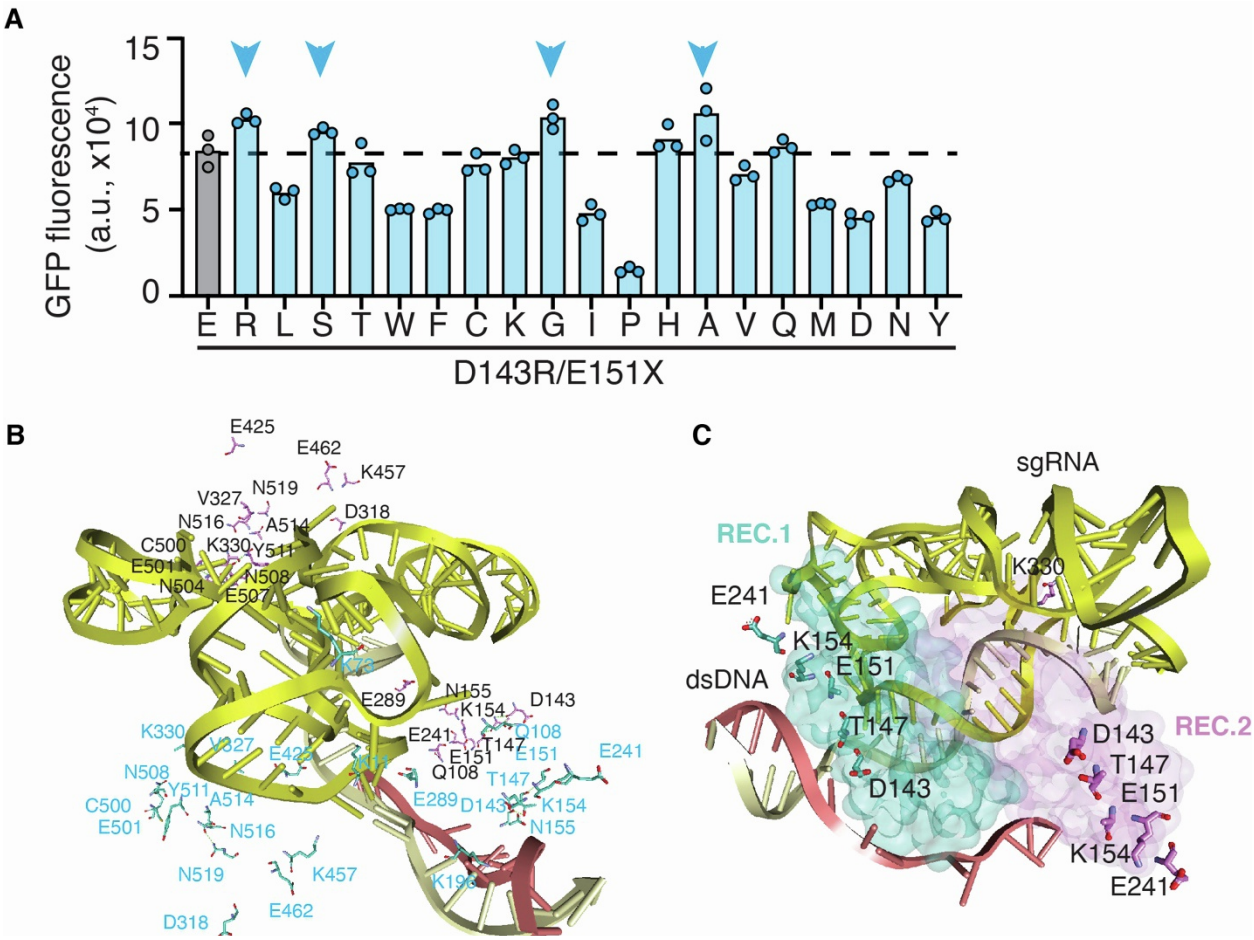


Figure S4. Performance of variants as measured by GFP expression in the second round of screening, and a general overview of mutated residues on the Cas12f:sgRNA:dsDNA (double-stranded DNA) complex structure. Related to Figure 1.

(A) All variants contained D143R and E151 is mutated to every other 19 possible amino acid. Bars represent mean values, and data represent three biological replicates. Constructs with enhanced activation relative to D143R mutant are indicated with blue arrows.

(B) Positions of residues chosen for mutagenesis for iterative dCas12f protein engineering. These residues of two dimers of the Cas12f are shown, one in blue and another in black. The dsDNA substrate is shown in red & grey and the sgRNA is shown in yellow. PDB: 7C7L.

(C) Selected residues for iterative dCas12f protein engineering overlayed to the two dimers. The dsDNA substrate is shown in red & grey and the sgRNA is shown in yellow. The binding centers of the dimer are shown as surfaces in light green and purple. PDB: 7C7L.

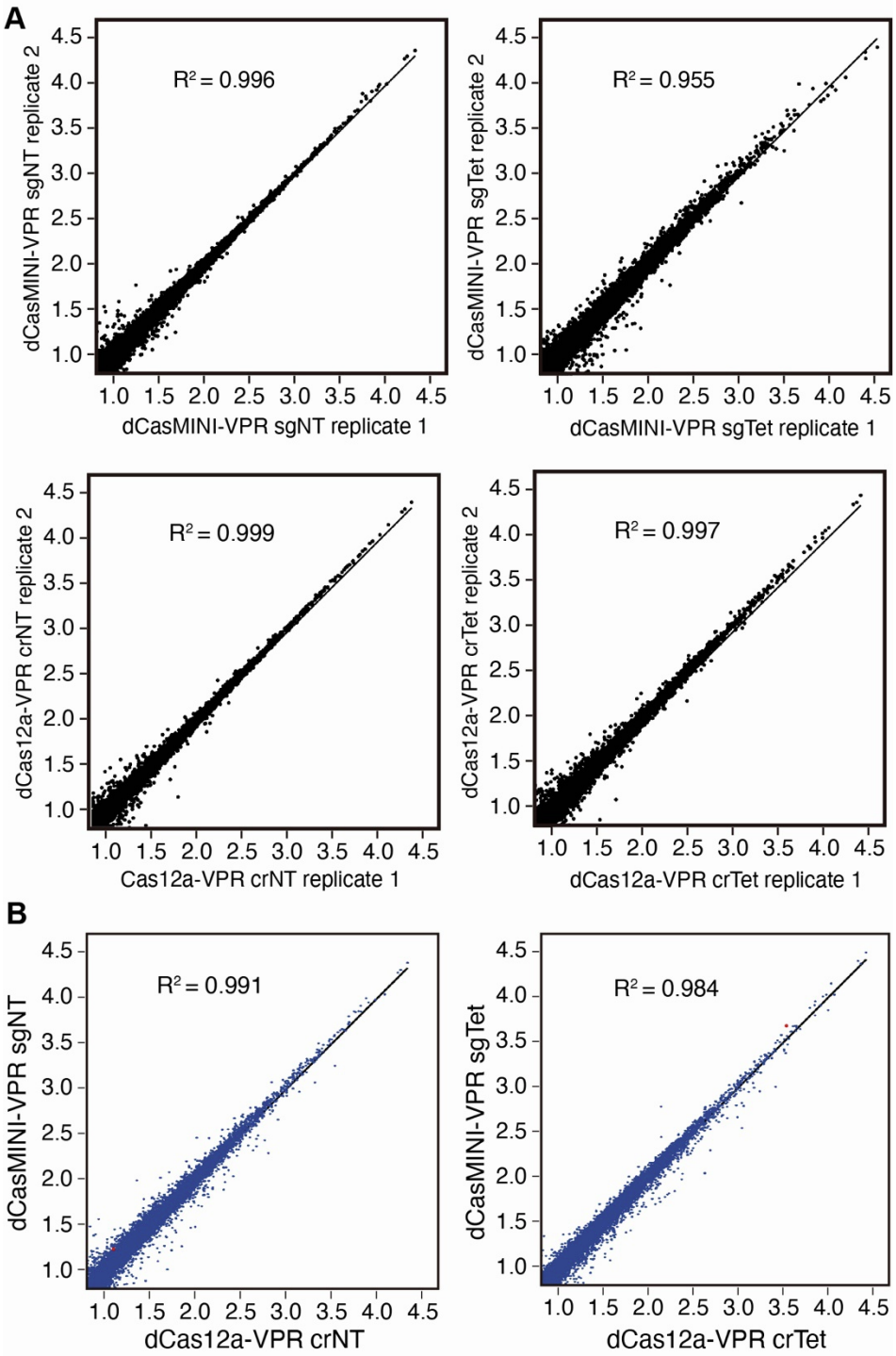


Figure S5. RNA-seq profiling of dCasMINI-VPR and dCas12a-VPR in HEK293T cells. Related to Figure 4.

(A) Scatterplots comparing $\log_{10}$(TPM+1) values for biological replicates for each condition. Top row: left, dCasMINI-VPR + non-targeting sgRNA (sgNT); right, dCasMINI-VPR + targeting sgRNA (sgTet). Second row: left, dCas12a-VPR + non-targeting crRNA (crNT); right, dCas12-VPR + targeting crRNA (crTet). The calculated Pearson correlation coefficient for each condition is shown.

(B) Scatterplots comparing $\log_{10}$(TPM+1) values of dCasMINI-VPR vs. dCas12a-VPR for the non-targeting guide (left) and the targeting guide (right). The calculated Pearson correlation coefficient for each condition is shown.

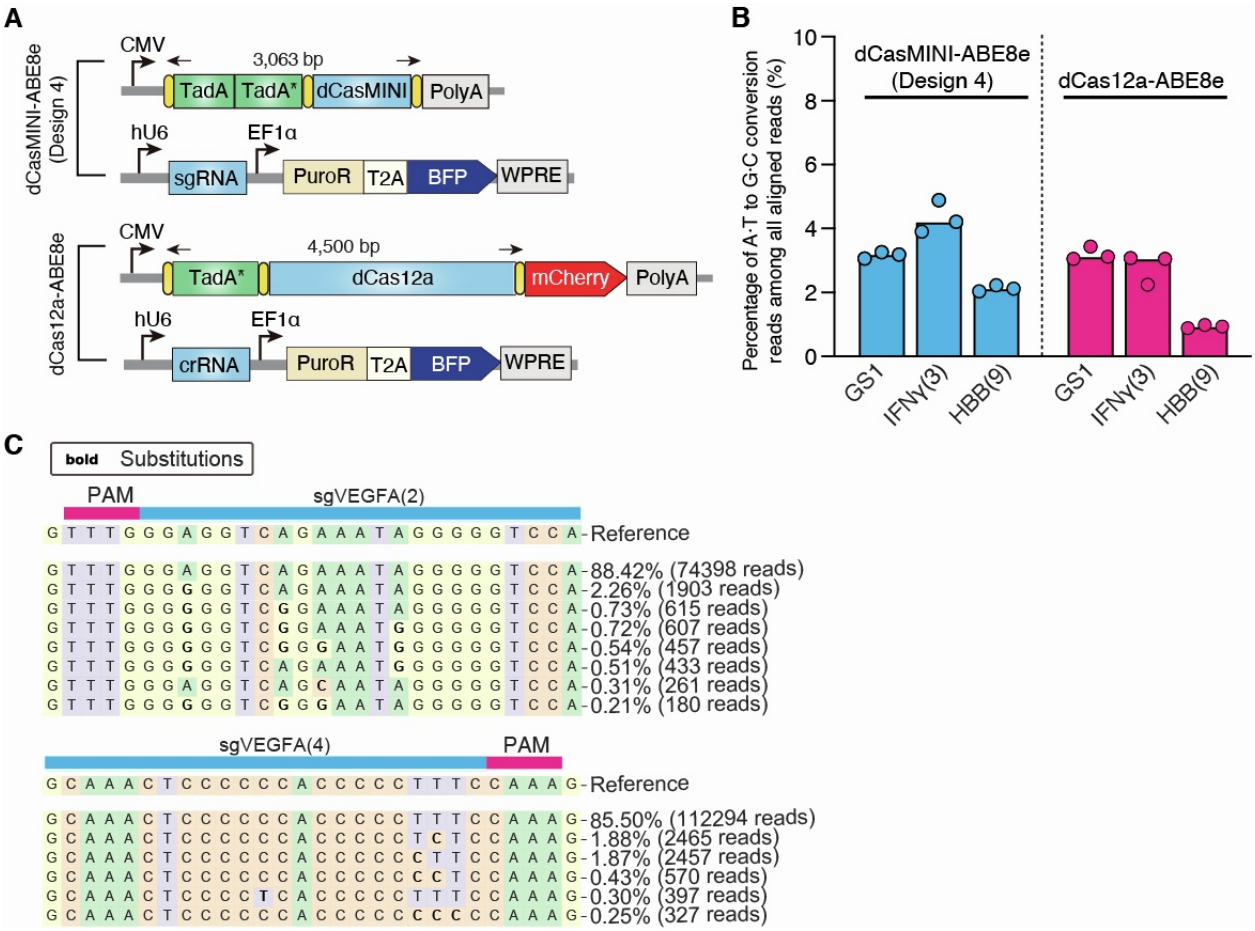


Figure S6. Comparison of dCasMINI-ABE to dCas12a-ABE and representative sequencing data for dCasMINI-ABE base editing. Related to Figure 5.

(A) Schematic of constructs for dCasMINI-ABE Design 4 and its sgRNA and constructs for dCas12a-ABE and its crRNA used for side-by-side comparison for base editing.

(B) Base editing activity using dCasMINI-ABE and dCas12a-ABE at three genomic sites using sgRNAs or crRNAs targeting the same genomic sites. GS1, genomic site 1. Bars represent mean values, and data represent three biological replicates.

(C) Raw sequencing reads from deep sequencing using dCasMINI-ABE and sgRNA targeting site 3 in the IFNγ locus or site 4 in the VEGFA locus. The sequenced reads and percentage among the total aligned reads are shown on the right. Representative variants with >0.2% of the total reads generated by CRISPResso2 are shown.

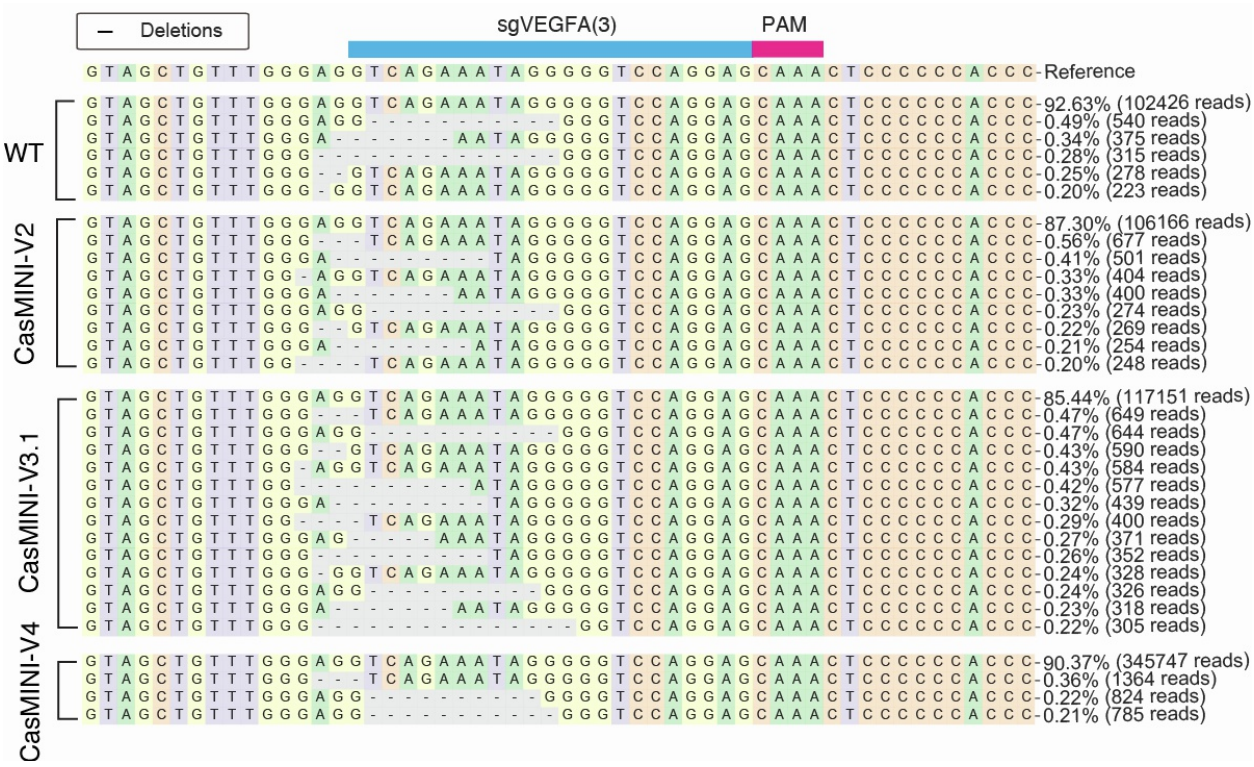


Figure S7. Representative sequencing data for nuclease active CasMINI-mediated indel patterns. Related to Figure 6. Raw sequencing reads from deep sequencing using the wildtype Cas12f, CasMINI-V2, CasMINI-V3.1, CasMINI-V4 by targeting the site 3 in the VEGFA locus. The sequenced reads and percentage among the total aligned reads are shown on the right. Representative variants with >0.2% of the total reads generated by CRISPResso2 are shown.

SUPPLEMENTAL TABLES

Table S1. Plasmids generated in the study. Related to Figures 1 – 6. BB, backbone.

Figure	Plasmid	Description	Construct
Figure 1B	pSLQ7368	Fusion #1 of dCas12f-VPR	pHR-PGK-SV40 NLS-Cas12f/D326A/D510A-VPR-SV40 NLS-mCherry-WPRE
	sgRNA in pSLQ9830	Common sgRNA BB (Direct fusion)	pHR-hU6-CasMINI sgRNA BB EF1a-Puro-T2A-BFP-WPRE
Figure 1D	pSLQ7368	Fusion #1 of dCas12f-VPR	pHR-PGK-SV40 NLS-Cas12f/D326A/D510A-VPR-SV40 NLS-mCherry-WPRE
	sgRNA in pSLQ9832	Common sgRNA BB (Design1)	pHR-hU6-CasMINI sgRNA_#1 BB EF1a-Puro-T2A-BFP-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
	sgRNA in pSLQ9836	Common sgRNA BB (Design3)	pHR-hU6-CasMINI sgRNA_#3 BB EF1a-Puro-T2A-BFP-WPRE
Figure 1F-1I	dCasMINI variants using the fusion based on pSLQ7334	Fusion #5 of different dCasMINI variants-VPR	pHR-PGK-SV40_NLS-dCasMINI variants-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure 2B-2H	pSLQ9926	dCasMINI (V4)-VPR	pHR-PGK-SV40_NLS-dCas12f/ D326A/D510A/D143R/T147R/K330R/E528R-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure 3B-3E	pSLQ9926	dCasMINI (V4)-VPR	pHR-PGK-SV40_NLS-dCas12f/ D326A/D510A/D143R/T147R/K330R/E528R-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure 3F ; Figure 3G; Figure 4A – 4D	pSLQ9926	dCasMINI (V4)-VPR	pHR-PGK-SV40_NLS-dCas12f/ D326A/D510A/D143R/T147R/K330R/E528R-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
	pSLQ7349	LbdCas12a-VPR	pHR-PGK-SV40_NLS-dLbCas12a-VPR-c-Myc_NLS-mCherry-WPRE
	crRNA in pSLQ8453	Common crRNA BB	pHR-hU6-LbdCas12a crRNA BB EF1a-Puro-T2A-BFP-WPRE
Figure 5B-5D	pSLQ9761	Design 1 of dCasMINI (V4)-ABE	CMV-BPNLS-TadA*(8e)-Linker B-SV40 NLS-Cas12f/D326A/D510A/D143R/T147R/K330R/E528R-linkerC-c-Myc_NLS-3x FLAG-polyA

	pSLQ9755	Design 2 of dCasMINI (V4)-ABE	CMV-BPNLS_TadA* (8e)-SV40_NLS-Cas12f/D326A/D510A/D143R/T147R/K330R/E528R-c-Myc_NLS-mCherry-PolyA
	pSLQ9763	Design 3 of dCasMINI (V4)-ABE	CMV-SV40 NLS-Cas12f/D326A/D510A/D143R/T147R/K330R/E528R-linker C-BPNLS-TadA*(8e)-linker B-c-Myc_NLS-3x FLAG-polyA
	pSLQ9757	Design 4 of dCasMINI (V4)-ABE	CMV-BPNLS_TadA-linkerA-TadA*(8e)-Linker B-SV40 NLS-Cas12f/D326A/D510A/D143R/T147R/K330R/E528R-linkerC-c-Myc_NLS-3x FLAG-polyA
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure 6B-6E	pSLQ9821	CasMINI-V2	CMV-SV40_NLS-Cas12f/D143R/T147R-3xFlag-c-Myc_NLS-polyA
	pSLQ9824	CasMINI-V3.1	CMV-SV40_NLS-Cas12f/D143R/T147R/E151A-3xFlag-c-Myc_NLS-polyA
	pSLQ9823	CasMINI-V4	CMV-SV40_NLS-Cas12f/D143R/T147R/K330R/E528R-HA tag-c-Myc_NLS-polyA
	pSLQ9825	Cas12f	CMV-SV40_NLS-Cas12f-3xFlag-c-Myc_NLS-polyA
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure S2B	pSLQ7368	Fusion #1 of dCas12f-VPR	pHR-PGK-SV40 NLS-Cas12f/D326A/D510A-VPR-SV40 NLS-mCherry-WPRE
	sgRNA in pSLQ9830	Common sgRNA BB (Direct fusion)	pHR-hU6-CasMINI sgRNA BB EF1a-Puro-T2A-BFP-WPRE
	pSLQ7368	Fusion #1 of dCas12f-VPR	pHR-PGK-SV40 NLS-Cas12f/D326A/D510A-VPR-SV40 NLS-mCherry-WPRE
	sgRNA in pSLQ9832	Common sgRNA BB (Design1)	pHR-hU6-CasMINI sgRNA_#1 BB EF1a-Puro-T2A-BFP-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
	sgRNA in pSLQ9836	Common sgRNA BB (Design3)	pHR-hU6-CasMINI sgRNA_#3 BB EF1a-Puro-T2A-BFP-WPRE
Figure S2D	pSLQ7368	Fusion #1	pHR-PGK-SV40 NLS-Cas12f /D326A/D510A-VPR-SV40 NLS-mCherry-WPRE
	pSLQ7367	Fusion #2	pHR-PGK-Cas12f/D326A/D510A-VPR-2×SV40 NLS-mCherry-WPRE
	pSLQ7369	Fusion #3	pHR-PGK-VPR-Cas12f/D326A/D510A-2×SV40 NLS-mCherry-WPRE
	pSLQ7370	Fusion #4	pHR-PGK-SV40 NLS-VPR-Cas12f/D326A/D510A-SV40 NLS-mCherry-WPRE
	pSLQ7334	Fusion #5	pHR-PGK-SV40_NLS-Cas12f/D326A/D510A-VPR-c-Myc_NLS-mCherry-WPRE
	pSLQ7335	Fusion #6	pHR-PGK-SV40_NLS-Cas12f/D326A/D510A-VPR- c-Myc_NLS-SV40_NLS-mCherry-WPRE
	pSLQ7337	Fusion #7	pHR-PGK-SV40_NLS -Cas12f/D326A/D510A-VPR-SV40_NLS-P2A-mCherry-WPRE

	pSLQ7338	Fusion #8	pHR-PGK-Cas12f/D326A/D510A-VPR-c-Myc_NLS-SV40_NLS-mCherry-WPRE
	pSLQ7339	Fusion #9	pHR-PGK-Cas12f/D326A/D510A-VPR-2× SV40_NLS-P2A-mCherry-WPRE
	pSLQ7340	Fusion #10	pHR-PGK-c-Myc_NLS-SV40_NLS-Cas12f/D326A/D510A-VPR-SV40_NLS-mCherry-WPRE
	pSLQ7341	Fusion #11	pHR-PGK-Cas12f/D326A/D510A-VPR-2× c-Myc_NLS-SV40_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure S3	pSLQ9926	dCasMINI (V4)-VPR	pHR-PGK-SV40_NLS-dCas12f/ D326A/D510A/D143R/T147R/K330R/E528R-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure S3 & S4A	dCasMINI variants using the fusion based on pSLQ7334	Fusion #5 of different dCasMINI variants-VPR	pHR-PGK-SV40_NLS-dCasMINI variants-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
Figure S5	pSLQ9926	dCasMINI (V4)-VPR	pHR-PGK-SV40_NLS-dCas12f/ D326A/D510A/D143R/T147R/K330R/E528R-VPR-c-Myc_NLS-mCherry-WPRE
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
	pSLQ7349	LbdCas12a-VPR	pHR-PGK-SV40_NLS-dLbCas12a-VPR-c-Myc_NLS-mCherry-WPRE
	crRNA in pSLQ8453	Common crRNA BB	pHR-hU6-LbdCas12a crRNA BB EF1a-Puro-T2A-BFP-WPRE
Figure S6B and S6C	pSLQ9757	Design 4 of dCasMINI (V4)-ABE	CMV-BPNLS_TadA-linkerA-TadA*(8e)-Linker B-SV40 NLS-Cas12f/D326A/D510A/D143R/T147R/K330R/E528R-linkerC-c-Myc_NLS-3x FLAG-polyA
	sgRNA in pSLQ9834	Common sgRNA BB (Design2)	pHR-hU6-CasMINI sgRNA_#2 BB EF1a-Puro-T2A-BFP-WPRE
	pSLQ9747	LbdCa12a-ABE	CMV-BPNLS_TadA (8e)-LbdCas12a-c-Myc_NLS-mCherry-PolyA
	pSLQ8453	Common crRNA BB	pHR-hU6-LbdCas12a crRNA BB EF1a-Puro-T2A-BFP-WPRE
Figure S7	pSLQ9821	CasMINI-V2	CMV-SV40_NLS-Cas12f/D143R/T147R-3xFlag-c-Myc_NLS-polyA
	pSLQ9824	CasMINI-V3.1	CMV-SV40_NLS-Cas12f/D143R/T147R/E151A-3xFlag-c-Myc_NLS-polyA
	pSLQ9823	CasMINI-V4	CMV-SV40_NLS-Cas12f/D143R/T147R/K330R/E528R-HA tag-c-Myc_NLS-polyA
	pSLQ9825	Cas12f	CMV-SV40_NLS-Cas12f-3xFlag-c-Myc_NLS-polyA

Table S2. Spacer sequences of sgRNAs used in this study. Related to Figures 1 – 6.

Relevant Gene or Site	Guide Name	Sequence (5'-3')	5'PAM
NT (non-targeting control)	sgLacZ (sgNT)	CGAATACGCCCACGCGATGGGTA	——
NT (non-targeting control)	crLacZ (crNT)	CGAATACGCCCACGCGATGGGT	——
GFP	sgTet/crTet	CTCCCTATCAGTGATAGAGAACG	TTTA
CD2	sgCD2-1	CAAAAGAGTGATCCTTAGTGATC	TTTA
	sgCD2-2	TGAGTGTGTGTTTTCTTGCTGCA	TTTA
	sgCD2-3/crCD2-3	TGTTACTGTAAAAGATGTAAAGA	TTTA
	sgCD2-4/crCD2-4	CATCTTTTACAGTAACATAAAAC	TTTA
	sgCD2-5	CCTATATTTCTATGTGGTCTTGT	TTTA
	sgCD2-6	CATTCTGCTATTGGCTTGTGAAC	TTTG
	sgCD2-7	GCTTCTTGTTTACAAAAGAGTGAT	TTTG
	sgCD2-8	TGTGAGAATTAAAATGCAGCAAG	TTTG
	sgCD2-9	CCAAAGCAGATGTGTTTATGAGT	TTTG
	sgCD2-10	GCAAAGGAGCACATCAGAAGGGC	TTTG
CXCR4	sgCXCR4-1/crCXCR4-1	GCAAGGATGGACGCGCCACAGAG	TTTA
	sgCXCR4-2	TAAAAGTCCGGCCGCGGCCAGAA	TTTA
	sgCXCR4-3	TAAAAACACGCTCCGAGCGCGGC	TTTA
	sgCXCR4-4	GAGGCGGAGGGCGGCGTGCCTGG	TTTA
	sgCXCR4-5	ACTTCGGGGTTAAGCGCCTGGTG	TTTA
	sgCXCR4-6/crCXCR4-6	AGGGAAGCGGGATGCGCCTGAAG	TTTG
	sgCXCR4-7	CGGGTGGTCGGTAGTGAGTCCGG	TTTG
	sgCXCR4-8	TTGGCTGCGGCAGCAGGTAGCAA	TTTG
	sgCXCR4-9	TCATAAATGTACAAACGTTTGAA	TTTG
	sgCXCR4-10	AACTTAGAGCGCAGCCCCTCTCC	TTTG
IFNγ	sgIFNG-1	ACCGCATTCTTTCCTTGCTTTCT	TTTA
	sgIFNG-2	AAAGGCTGCCCCTTTGTAAAGGT	TTTA
	sgIFNG-3/crIFNG-3	CCAGGGCGAAGTGGGGAGGTACA	TTTA
	sgIFNG-4	CTTCACACCATTCAAGGACTGGA	TTTA
	sgIFNG-5/crIFNG-5	AGATGAGATGGTGACAGATAGGC	TTTA
	sgIFNG-6	AGAGGCCCTAGAATTTCGTTTTT	TTTG
	sgIFNG-7	TGAATGAAGAGTCAACATTTTAC	TTTG
	sgIFNG-8	TACCTCCCCACTTCGCCCTGGTA	TTTG
	sgIFNG-9	TATTAATAACTAAGGTTTTGTGG	TTTG
	sgIFNG-10	GACCTGATCAGCTTGATACAAGA	TTTG
HBB	sgHBB-1	TTTATTTGTATTTTTGACTGCAT	TTTA
	sgHBB-2	TCTCTTGTTTCCCAAAACCTAAT	TTTA
	sgHBB-3	TTCTATTTTTAGACATAATTTAT	TTTA

	sgHBB-4/crHBB-4	GACATAATTTATTAGCATGCATG	TTTA
	sgHBB-5	TTAGCATGCATGAGCAAATTAAG	TTTA
	sgHBB-6	ATCCAAATAAGGAGAAGATATGC	TTTA
	sgHBB-7	GTGCATCAACTTCTTATTTGTGT	TTTA
	sgHBB-8	CGTAATATTTGGAATCACAGCTT	TTTA
	sgHBB-9/crHBB-9	GTAGCAATTTGTACTGATGGTAT	TTTA
	sgHBB-10	AGGAGACCAATAGAAACTGGGCA	TTTA
	sgHBB-11	ACACCACTGATTACCCCATTGAT	TTTG
	sgHBB-12	TCTACCATAATTCAGCTTTGGGA	TTTG
	sgHBB-13	CATATTCTGGAGACGCAGGAAGA	TTTG
	sgHBB-14	TGTAATAAGAAAATTGGGAAAAC	TTTG
	sgHBB-15	GAATCACAGCTTGGTAAGCATAT	TTTG
	sgHBB-16	CAAGTGTATTTACGTAATATTTG	TTTG
	sgHBB-17	TACTGATGGTATGGGGCCAAGAG	TTTG
	sgHBB-18	AAGTCCAACTCCTAAGCCAGTGC	TTTG
	sgHBB-19	CTTCTGACACAACTGTGTTCACT	TTTG
	sgHBB-20	AGGTTGCTAGTGAACACAGTTGT	TTTG
IL1RN	sgIL1RN-1	GGTAAGCTCCTTCCACTCTCATT	TTTA
	sgIL1RN-2	TGGGCAGCAGCTCAGTTGAGTTA	TTTA
	sgIL1RN-3	CAGGAGGGTGACTCAGGCTAGCA	TTTC
	sgIL1RN-4	TGCTAGCCTGAGTCACCCTCCTG	TTTC
	sgIL1RN-5	GTTTCTGCTAGCCTGAGTCACCC	TTTG
	sgIL1RN-6	AATGAATGTGTGCACACATGCAT	TTTG
	sgIL1RN-7	TTAGAGCGTTGGGGACCTTGTCT	TTTG
	sgIL1RN-8	TTTGCTAACTTGTTTCTTGTCTG	TTTA
	sgIL1RN-9	CCTATAGCAATATGCCCTATGAA	TTTC
	sgIL1RN-10	GGGAAAGTGGCCAGGGAAGCCCT	TTTA
HBG	sgHBG-1	ACAGAGGAGGACAAGGCTACTAT	TTTC
	sgHBG-2	TTCTTCATCCCTAGCCAGCCGCC	TTTA
	sgHBG-3	CCTTGTCAAGGCTATTGGTCAAG	TTTG
	sgHBG-4	GCCAGGGACCGTTTCAGACAGAT	TTTA
	sgHBG-5	AGACAGATATTTGCATTGAGATA	TTTC
	sgHBG-6	CATTGAGATAGTGTGGGGAAGGG	TTTG
	sgHBG-7	TAGCCTTTGCCTTGTTCCGATTC	TTTA
	sgHBG-8	CCTTGTTCCGATTCAGTCATTCC	TTTG
	sgHBG-9	TCTAATTTATTCTTCCCTTTAGC	TTTC
	sgHBG-10	AACTACAGGCCTCACTGGAGCTA	TTTA
ASCL1	sgASCL1-1	TTATTTTGTTAACTCCCTTCCTT	TTTA
	sgASCL1-2	ACTCGCCCTCCCTGGCCGGATCC	TTTC
	sgASCL1-3	AATGGGACACCCAGCCCCACGCG	TTTC
	sgASCL1-4	GGGAGTGGGTGGGAGGAAGAGGT	TTTA

	sgASCL1_5	CAAGGAGCGGGAGAAAGGAACGG	TTTG
	sgASCL1_6	TCCCGCTCCTTGCAAACTCTCCA	TTTC
	sgASCL1_7	TTGTTGCAGTGCGTGCGCCTGGC	TTTG
	sgASCL1_8	TTCAGCCGGGAGTCCGGCACGCG	TTTA
	sgASCL1_9	GGAAGGGGGTGGGGGGCGTCACA	TTTA
	sgASCL1_10	TCCCTCCTGTGACGCCCCCCACC	TTTG
VEGFA	sgVEGFA_1	GGACTGGAGTTGCTTCATGTACA	TTTG
	sgVEGFA_2	GGAGGTCAGAAATAGGGGGTCCA	TTTG
	sgVEGFA_3	CTCCTGGACCCCCTATTTCTGAC	TTTG
	sgVEGFA_4	GAAAGGGGGTGGGGGGAGTTTGC	TTTG
	sgVEGFA_5	GCCAGAGCCGGGGTGTGCAGACG	TTTA
GS0	sgGS0	CAGGGCCAGCGGGCTGGAAAATT	TTTA
GS1	sgGS1/crGS1	GAAGCACATCAAGGACATTCTAA	TTTA
GS2	sgGS2	ATTAATCCACAACCACCTCATCT	TTTA
GS3	sgGS3	GTTTAAACACACCGGGTTAATAA	TTTG

Table S3. Designed full sgRNA sequences. Related to Figure 1.

Name	sgRNA sequence
Direct fusion	5'- GGGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGG GATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAG TGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTTTCCTCTCCAATTCT GCACAAGAAAGTTGCAGAACCCGAATAGACGAATGAAGGAATGCAACNNNNN NNNNNNNNNNNNNNNNNN-3'
Design 1	5'- GGGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGG GATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAG TGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTGTTCCTCTCCAATTCT GCACAAGAAAGTTGCAGAACCCGAATAGACGAATGAAGGAATGCAACNNNNN NNNNNNNNNNNNNNNNNN-3
Design 2	5'- GGGCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGG GATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAG TGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGC AACNNNNNNNNNNNNNNNNNNNN-3'
Design 3	5'- GCTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGA TTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTG CTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTGAATGAAGGAATGCAA CNNNNNNNNNNNNNNNNNNNN-3'

Table S4. CasMINI protein sequences for best activities for gene activation, base editing, and gene editing. Related to Figures 1 - 6.

Plasmid	Gene	Description
pSLQ9926	dCasMINI (V4)-VPR	pHR-PGK-SV40_NLS-Cas12f/ D326A/D510A/D143R/T147R/K330R/E528R-VPR-c-Myc_NLS-mCherry-WPRE

TACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTG
GCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTT
TGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTC
GTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAGCACCGCTGAGCAATG
GAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAGGCTGG
GAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGGTCCTCC
GGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGC
CTTTCGACCTGCAGCCCAAGCTTACGCGTCTGCAGGATATCAAGCTTGCGGTACCGCGGGCCCGGGCCACC
ATGGGACCCAAGAAAAAACGCAAGGTGGGAAGCGGATCCGCCAAAAACACCATTACCAAAACACTGAAACTG
CGTATTGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATTGTTGCCGACGAAAAAAACAACCGCGAAAAAA
TCGCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCTGCAGCAAACATCTGAAAGTTGCAGCATATTGTAC
CACACAGGTTGAACGTAATGCATGCCTGTTTTGTAAAGCACGTAAACTGGATGACAAATTCTACCAAAAACTGC
GTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGAAATCAGCGAAATTTTTCGCCAGCTGCAGAAACAGGCAGC
AGAAATCTATAATCAGAGCCTGATCGAACTGTACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCA
GCAGCGTTGAACATTATCTGAGTAGAGTTTGTTATAGACGTGCAGCAGAACTGTTTAAAAACGCAGCAATTGC
AAGCGGTCTGCGTAGCAAAATCAAAAGCAATTTTCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCG
ACCACCAAAAGCGATAATTTTCCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAG
CAATCATAATAGCGACTTCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATC
GTCCGTGGGAGAAATTTGACTTTGAACAGGTTCAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCC
AGCGTCGTAAACGTAATAAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGG
CGATTATCAGACCAGCTACATTGAAGTTAAACGTGGCAGCAAAATCTGTGAAAAAAGCGCATGGATGCTGAAT
CTGAGCATTGATGTTCCGAAAATTGATAAAGGTGTGGATCCGAGCATTATTGGTGGTATTGCAGTTGGTGTTA
GATCACCGCTGGTTTGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACTTC
AACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCATGGTGCAA
AAAACAAACTGAAACCGATCACCATTCTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGATTGAACGTTG
GGCATGTGAAATCGCGGATTTCTTCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAATCTGGAAAGCATG
AAACGTAAAGAGGACAGCTATTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCAGAAATGCAGAACAAAA
TCGAATTCAAACTGAAGCAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATAATACCAGCAAAACCTGTAG
CAAATGTGGCCATCTGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAACAAATTCCCGCACTTTAAATGCG
AAAAATGCAACTTCAAAGAAAACGCCGCGTATAATGCAGCCCTGAATATTTCAAACCCGAAACTGAAAAGCAC
CAAAGAGAGACCGGCCTATCCCTATGACGTGCCCGATTATGCCAGCCTGGGCAGCGGCGACGGCATTGGTA
GTGGGAGCAACGGCAGCAGCCTCGATGCTTTAGACGATTTTGACTTAGATATGCTTGGTTCAGACGCGTTAGA
CGACTTCGACCTAGACATGTTAGGCTCAGATGCATTGGACGACTTCGATTTAGATATGTTGGGCTCCGATGCC
CTAGATGACTTTGATCTAGATATGCTAGGTAGTGGCGGCAGCGGATCCCAGTATCTGCCCGACACAGATGATA
GACACCGAATCGAAGAGAAACGCAAGCGAACGTATGAAACCTTCAAATCGATCATGAAGAAATCGCCCTTCTC
GGGTCCGACCGATCCCAGGCCCCCACCGAGAAGGATTGCGGTCCCGTCCCGCTCGTCGGCCAGCGTGCCG
AAGCCTGCGCCGCAGCCCTACCCCTTCACGTCGAGCCTGAGCACAATCAATTATGACGAGTTCCCGACGATG
GTGTTCCCCTCGGGACAAATCTCACAAGCCTCGGCGCTCGCACCAGCGCCTCCCCAAGTCCTTCCGCAAGC
GCCTGCCCCAGCGCCTGCACCGGCAATGGTGTCCGCCCTCGCACAGGCCCCTGCGCCCGTCCCCGTGCTC
GCGCCTGGACCGCCCCAGGCGGTCGCTCCACCGGCTCCGAAGCCGACGCAGGCCGGAGAGGGAACACTCT
CCGAAGCACTTCTTCAACTCCAGTTTGATGACGAGGATCTTGGAGCACTCCTTGGAAACTCGACAGACCCTGC
GGTGTTTACCGACCTCGCGTCAGTAGATAACTCCGAATTTCAGCAGCTTTTGAACCAGGGTATCCCGGTCGC
GCCACATACAACGGAGCCCATGTTGATGGAATACCCCGAAGCAATCACGAGACTTGTGACGGGAGCGCAGC
GGCCTCCCGATCCCGCACCCGCACCTTTGGGGGCACCTGGCCTCCCTAACGGACTTTTGAGCGGCGACGAG
GATTTCTCCTCCATCGCCGATATGGATTTCTCAGCCTTGCTGTCACAGATTTCCAGCGGCTCTGGCAGCGGCA
GCCGGGATTCCAGGGAAGGGATGTTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAG
GGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCC
ACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCAC
CAGTCCCTCAGCCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGAT
GAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGC
TGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACT
TGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTT
CCTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTT

CATATGGGTGGAGGCTCCGGGGAAGATCCTGCTGCCAAACGCGTTAAACTAGACATGGGAAGCGGAGTGAG
CAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCG
TGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAA
GCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTC
CAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTG
GGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGC
GAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACC
ATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGA
GGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTG
CAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTG
GAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAA

Plasmid	Gene	Description
pSLQ9757	Design 4 of dCasMINI (V4)-ABE	CMV-BPNLS_TadA-linkerA-TadA*(8e)-Linker B-SV40 NLS-Cas12f/D326A/D510A/D143R/T147R/K330R/E528R-linkerC-c-Myc_NLS-3x FLAG-polyA

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAA
TGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC
TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT
CGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT
TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT
CGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT
GGTTTAGTGAACCGTCAGATCCGCTAGCCCGGGCCACCATGGGAAAACGGACAGCCGACGGAAGCGAGTTC
GAGTCACCAAAGAAGAAGCGGAAAGTCTCTGAAGTCGAGTTTAGCCACGAGTATTGGATGAGGCACGCACTG
ACCCTGGCAAAGCGAGCATGGGATGAAAGAGAAGTCCCCGTGGGCGCCGTGCTGGTGCACAACAATAGAGT
GATCGGAGAGGGATGGAACAGGCCAATCGGCCGCCACGACCCTACCGCACACGCAGAGATCATGGCACTGA
GGCAGGGAGGCCTGGTCATGCAGAATTACCGCCTGATCGATGCCACCCTGTATGTGACACTGGAGCCATGC
GTGATGTGCGCAGGAGCAATGATCCACAGCAGGATCGGAAGAGTGGTGTTCGGAGCACGGGACGCCAAGAC
CGGCGCAGCAGGCTCCCTGATGGATGTGCTGCACCACCCCGGCATGAACCACCGGGTGGAGATCACAGAGG
GAATCCTGGCAGACGAGTGCGCCGCCCTGCTGAGCGATTTCTTTAGAATGCGGAGACAGGAGATCAAGGCC
CAGAAGAAGGCACAGAGCTCCACCGACTCTGGAGGATCTAGCGGAGGATCCTCTGGAAGCGAGACACCAGG
CACAAGCGAGTCCGCCACACCAGAGAGCTCCGGCGGCTCCTCCGGAGGATCCTCTGAGGTGGAGTTTTCCC
ACGAGTACTGGATGAGACATGCCCTGACCCTGGCCAAGAGGGCACGGGATGAGAGGGAGGTGCCTGTGGG
AGCCGTGCTGGTGCTGAACAATAGAGTGATCGGCGAGGGCTGGAACAGAGCCATCGGCCTGCACGACCCAA
CAGCCCATGCCGAAATTATGGCCCTGAGACAGGGCGGCCTGGTCATGCAGAACTACAGACTGATTGACGCCA
CCCTGTACGTGACATTCGAGCCTTGCGTGATGTGCGCCGGCGCCATGATCCACTCTAGGATCGGCCGCGTG
GTGTTTGGCGTGAGGAACTCAAAAAGAGGCGCCGCAGGCTCCCTGATGAACGTGCTGAACTACCCCGGCAT
GAATCACCGCGTCGAAATTACCGAGGGAATCCTGGCAGATGAATGTGCCGCCCTGCTGTGCGATTTCTATCG
GATGCCTAGACAGGTGTTCAATGCTCAGAAGAAGGCCCAGAGCTCCATCAACTCCGGAGGATCTAGCGGAGG
ATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGC
GGCAGCCCCAAGAAAAAACGCAAGGTGGGAAGCGGATCCGCCAAAAACACCATTACCAAAACACTGAAACTG
CGTATTGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATTGTTGCCGACGAAAAAAACAACCGCGAAAAAA
TCGCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCTGCAGCAAACATCTGAAAGTTGCAGCATATTGTAC
CACACAGGTTGAACGTAATGCATGCCTGTTTTGTAAAGCACGTAAACTGGATGACAAATTCTACCAAAAACTGC
GTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGAAATCAGCGAAATTTTTCGCCAGCTGCAGAAACAGGCAGC
AGAAATCTATAATCAGAGCCTGATCGAACTGTACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCA
GCAGCGTTGAACATTATCTGAGTAGAGTTTGTTATAGACGTGCAGCAGAACTGTTTAAAAACGCAGCAATTGC
AAGCGGTCTGCGTAGCAAAATCAAAAGCAATTTTCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCG
ACCACCAAAAGCGATAATTTTCCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAG
CAATCATAATAGCGACTTCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATC
GTCCGTGGGAGAAATTTGACTTTGAACAGGTTCAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCC
AGCGTCGTAAACGTAATAAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGG
CGATTATCAGACCAGCTACATTGAAGTTAAACGTGGCAGCAAAATCTGTGAAAAAAGCGCATGGATGCTGAAT
CTGAGCATTGATGTTCCGAAAATTGATAAAGGTGTGGATCCGAGCATTATTGGTGGTATTGCAGTTGGTGTTA
GATCACCGCTGGTTTGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACTTC
AACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCATGGTGCAA
AAAACAAACTGAAACCGATCACCATTCTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGATTGAACGTTG
GGCATGTGAAATCGCGGATTTCTTCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAATCTGGAAAGCATG
AAACGTAAAGAGGACAGCTATTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCAGAAATGCAGAACAAAA
TCGAATTCAAACTGAAGCAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATAATACCAGCAAAACCTGTAG

CAAATGTGGCCATCTGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAACAAATTCCCGCACTTTAAATGCG
AAAAATGCAACTTCAAAGAAAACGCCGCGTATAATGCAGCCCTGAATATTTCAAACCCGAAACTGAAAAGCAC
CAAAGAGAGACCGGCCTATCCCTATGACGTGCCCGATTATGCCAGCCTGGGCAGCGGCGACGGCATTGGTA
GTGGGAGCAACGGCAGCAGCCTCCATATGGGTGGAGGCTCCGGGGAAGATCCTGCTGCCAAACGCGTTAAA
CTAGACGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGT
GAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGT
CTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAA
TAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG

Plasmid	Gene	Description
pSLQ9824	CasMINI-V3.1	CMV-SV40_NLS-Cas12f/D143R/T147R/E151A-3xFlag-c-Myc_NLS-polyA

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCC
GCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAA
TGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAAC
TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGG
CCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT
CGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT
TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT
CGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT
GGTTTAGTGAACCGTCAGATCCGCTAGCCCGGGCCACCATGGGACCCAAGAAAAAACGCAAGGTGGGAAGC
GGATCCGCCAAAAACACCATTACCAAAACACTGAAACTGCGTATTGTGCGTCCGTATAATAGCGCAGAAGTGG
AAAAAATTGTTGCCGACGAAAAAAACAACCGCGAAAAAATCGCACTGGAAAAGAACAAAGACAAAGTGAAAGA
AGCCTGCAGCAAACATCTGAAAGTTGCAGCATATTGTACCACACAGGTTGAACGTAATGCATGCCTGTTTTGT
AAAGCACGTAAACTGGATGACAAATTCTACCAAAAACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAG
AAATCAGCGAAATTTTTCGCCAGCTGCAGAAACAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAACTGTA
CTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTAGAGTTTGTT
ATAGACGTGCAGCAGCTCTGTTTAAAAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAATCAAAAGCAATTT
TCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATTTTCCGATTCCGCTG
GTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAGCAATCATAATAGCGACTTCATCATCAAGATTCC
GTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAGAAATTTGACTTTGAACAGGTT
CAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTAAACGTAATAAAGGTTGGAGCAAA
GATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTATCAGACCAGCTACATTGAAGTTAAAC
GTGGCAGCAAAATCTGTGAAAAAAGCGCATGGATGCTGAATCTGAGCATTGATGTTCCGAAAATTGATAAAGG
TGTGGATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTAAATCACCGCTGGTTTGCGCAATTAACAATGCAT
TTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACTTCAACAAGAAAATGTTTGCACGTCGTCGTATCCTG
CTGAAAAAAAACCGTCATAAACGTGCAGGTCATGGTGCAAAAAACAAACTGAAACCGATCACCATTCTGACCG
AAAAAAGTGAACGTTTTCGCAAAAAGCTGATTGAACGTTGGGCATGTGAAATCGCGGATTTCTTCATTAAAAAC
AAAGTTGGCACCGTGCAGATGGAAAATCTGGAAAGCATGAAACGTAAAGAGGACAGCTATTTTAACATTCGCC
TGCGTGGCTTTTGGCCGTATGCAGAAATGCAGAACAAAATCGAATTCAAACTGAAGCAGTATGGCATCGAAAT
TCGTAAAGTTGCACCGAATAATACCAGCAAAACCTGTAGCAAATGTGGCCATCTGAACAACTATTTCAACTTCG
AGTACCGCAAGAAAAACAAATTCCCGCACTTTAAATGCGAAAAATGCAACTTCAAAGAAAACGCCGATTATAAT
GCAGCCCTGAATATTTCAAACCCGAAACTGAAAAGCACCAAAGAGGAACCGGCCGACTATAAGGACCACGAC
GGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGAGCCTGGGCAGCGGCCCTGCTGCCC
AAACGCGTTAAACTAGACTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTG
TTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGA
GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG

Table S5. qPCR primers used in this study. Related to Figures 2 – 3.

Gene	Primer name	Primer sequence (5'-3')	Size (bp)
HBB	q_HBB-F	GCACGTGGATCCTGAGAACT	176
	q_HBB-R	ATTGGACAGCAAGAAAGCGAG	
IL1RN	q_IL1RN-F	GGAATCCATGGAGGGAAGAT	100
	q_IL1RN-R	TGTTCTCGCTCAGGTCAGTG	
HBG	q.HBG-F	GCTGAGTGAACTGCACTGTGA	106
	q.HBG-R	GAATTCTTTGCCGAAATGGA	
IFNG	q.IFNG-F	GAGTGTGGAGACCATCAAGGA	129
	q.IFNG-R	TGTATTGCTTTGCGTTGGAC	
ASCL1	q.ASCL1-F	GGGCTCTTACGACCCGCTCA	127
	q.ASCL1-R	AGGTTGTGCGATCACCCTGCTT	
GAPDH	q_GAPDH-F	CAATGACCCCTTCATTGACC	159
	q_GAPDH-R	TTGATTTTGGAGGGATCTCG	

Table S6. Primers used for high-throughput sequencing in this study. Related to Figure 5 & Figure 6. "N" represents random nucleotides, and "n" represents Illumina indices used for deep sequencing.

Primers for amplifying genomic sites		
Genomic Site	**HTS Forward Primer (5'-3')**	**HTS Reverse Primer (5'-3')**
GS0	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGCGTATGACATCATCAGATATTC	CAGACGTGTGCTCTTCCGATCTTAACAAGACCTGGCTGAGCTAAC
GS1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCATTCTTATCTTAAAACCTTGTCACACAC	CAGACGTGTGCTCTTCCGATCTAGTGAGGAGAAGGCAGGAGG
GS2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTTATGATCCAATGAGACAGATGAGGTGG	CAGACGTGTGCTCTTCCGATCTCCCAAAGGACATACGGGGAG
GS3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTGCGTGCTTCTTACATGCC	CAGACGTGTGCTCTTCCGATCTCCAAGTATGCCTTAAGCAGAACAA
IFNγ-1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCCAACCACAAGCAAATGATCAATG	CAGACGTGTGCTCTTCCGATCTCAAGTTTTTTAAGATGAGATGGTGACAG
IFNγ-2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAAATGCCACAAAACCTTAGTTATTAATAC	CAGACGTGTGCTCTTCCGATCTGAAAAGCCAAGATATAACTTGTATATTTC
HBB	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGAAAACGATCTTCAATATGCTTACC	CAGACGTGTGCTCTTCCGATCTGCCCTGACTTTTATGCCCAGCCCTG
VEGFA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGAGGGACACACTGTGGCCCCTGTG	CAGACGTGTGCTCTTCCGATCTGCCCTCAACCCCACACGCACACACTCAC
Primers for adding Illumina indices		
	Index 2 adapter (5'-3')	**Index 1 adapter (5'-3')**
	AATGATACGGCGACCACCGAGATCTACACnnnnnnnn ACACTCTTTCCCTACACGACG	CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Table S7. Amplicons for high-throughput sequencing. Related to Figure 5 & Figure 6.

Site #	Amplicon (5'-3')
GS0	GTGCGTATGACATCATCAGATATTCTGCACTTGTTTGCAGCTATTCAGGCTGGCCC GCCCCGCAGTCTATGCTTTGTGTTCCAGTTTCCTTTACAGGGCCAGCGGGCTGGA AAATTACCACATGCTGTCACAGTTAGCTCAGCCAGGTCTTGTTA
GS1	CCATTCTTATCTTAAAACCTTGTCACACACAATGAAACTTTGCTGTTCACTGTCAGT TATAACTTACATGAGGTGACCCATTTCCATTCAAGGGTTTTAGAAGCACATCAAGG ACATTCTAAGGATGATTGACTTACACAATGATCTCTGAACATGCCTCCTGCCTTCTC CTCACT
GS2	GTTATGATCCAATGAGACAGATGAGGTGGTTGTGGATTAATTAAATATTCATAAAAA AGCAAATTAAGTGCTATAATGTTAGAATTACAGAAAAATAAAGGATGTAATTTTGGG ACTTTAATTTTTTAAATACTTATATTCACTTTTATAACGAAGAACTCTTTGTGGAAAAT GGTAATTTTCTGTTACCATTTGGGATAAGCACAGTTTTAAATAGTTCTGGAATTATA GAGGCACCTCCCCGTATGTCCTTTGGG
GS3	GTGCGTGCTTCTTACATGCCTTATTAACCCGGTGTGTTTAAACCAAACACTGTTCAT ATTTTTCCAGGAGGAAAACAAAACAATAAAAAACATTATTCAGATAAAATATTATAG GTTTATTTAAAACTTAATTCTCACCTTGAGTATGCAAAATACAAACTCCACAAAATGT TCATTTTACTTTGTAGTTTACAAATATACAAAATAGACGTTTGCTTAAATTTATATTAC ATATTTATTAAGGCAAGGAACTATATAGAAAAACACATTTGTTCTGCTTAAGGCATA CTTGG
IFNγ-1	CCCAACCACAAGCAAATGATCAATGTGCTTTGTGAATGAAGAGTCAACATTTTACC AGGGCGAAGTGGGGAGGTACAAAAAAATTTCCAGTCCTTGAATGGTGTGAAGTAA AAGTGCCTTCAAAGAATCCCACCAGAATGGCACAGGTGGGCATAATGGGTCTGTC TCATCGTCAAAGGACCCAAGGAGTCTAAAGGAAACTCTAACTACAACACCCAAATG CCACAAAACCTTAGTTATTAATACAAACTATCATCCCTGCCTATCTGTCACCATCTC ATCTTAAAAAACTTG
IFNγ-2	CAAATGCCACAAAACCTTAGTTATTAATACAAACTATCATCCCTGCCTATCTGTCAC CATCTCATCTTAAAAAACTTGTGAAAATACGTAATCCTCAGGAGACTTCAATTAGGT ATAAATACCAGCAGCCAGAGGAGGTGCAGCACATTGTTCTGATCATCTGAAGATCA GCTATTAGAAGAGAAAGATCAGTTAAGTCCTTTGGACCTGATCAGCTTGATACAAG AACTACTGATTTCAACTTCTTTGGCTTAATTCTCTCGGAAACGATGAAATATACAAG TTATATCTTGGCTTTTC
HBB	GGGAAAACGATCTTCAATATGCTTACCAAGCTGTGATTCCAAATATTACGTAAATAC ACTTGCAAAGGAGGATGTTTTTAGTAGCAATTTGTACTGATGGTATGGGGCCAAGA GATATATCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAACTCCTAAGCCAGTGCC AGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCCTGTGGAGCC ACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGG CTGGGCATAAAAGTCAGGGC
VEGFA	CAGAGGGACACACTGTGGCCCCTGTGCCCAGCCCTGGGCTCTCTGTACATGAAG CAACTCCAGTCCCAAATATGTAGCTGTTTGGGAGGTCAGAAATAGGGGGTCCAGG AGCAAACTCCCCCCACCCCCTTTCCAAAGCCCATTCCCTCTTTAGCCAGAGCCGG GGTGTGCAGACGGCAGTCACTAGGGGGCGCTCGGCCACCACAGGGAAGCTGGG TGAATGGAGCGAGCAGCGTCTTCGAGAGTGAGGACGTGTGTGTCTGTGTGGGTG AGTGAGTGTGTGCGTGTGGGGTTGAGGGC