

Novel miniature CRISPR–Cas13 systems from uncultivated microbes effective in degrading SARS-CoV-2 sequences and influenza viruses

Chunlong Xu

Chinese Academy of Sciences

Yingsi Zhou

Chinese Academy of Sciences

Qingquan Xiao

Chinese Academy of Sciences

Bingbing He

Chinese Academy of Sciences

Guannan Geng

Chinese Academy of Sciences

Zikang Wang

Chinese Academy of Sciences

Birong Cao

University of Chinese Academy of Sciences

Xiang Wang

Fudan University

Dongming Zhou

Fudan University

Tanglong Yuan

Chinese Academy of Agricultural Sciences

Xiaona Huo

Chinese Academy of Sciences

Hui Yang (✉ huiyang@ion.ac.cn)

Chinese Academy of Sciences

Research Article

Keywords: natural microbes, Cas13, computational pipeline

Posted Date: May 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-30924/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Methods on May 3rd, 2021. See the published version at <https://doi.org/10.1038/s41592-021-01124-4>.

Abstract

Recently emerging SARS-CoV-2 virus has caused a global pandemic, with millions of infections and over 200,000 deaths¹. However, development of effective anti-coronavirus treatments has lagged behind. Competitive co-evolution between microbes and viruses has led to the diversification of microbes' CRISPR/Cas defense systems against infectious viruses^{2,3}. Among class-2 single effector systems, Cas13 is effective in combating RNA phages⁴. Previous studies have discovered novel Cas9 and Cas12 systems from metagenomic sequence of natural microbes⁵⁻⁷. Here we report the identification of two additional compact Cas13 families from natural microbes that are effective in degrading RNA viruses in mammalian cells. Using metagenomic terabase data sets, we searched for previously uncharacterized Cas13 genes proximal to the CRISPR array with a customized computational pipeline, and identified two most compact families (775 to 803 amino acids) of CRISPR-Cas ribonucleases, named hereafter as CRISPR/Cas type VI-E and VI-F. Out of seven Cas13 proteins, we found that Cas13e.1 was the smallest and could be engineered for efficient RNA interference and base editing in cultured mammalian cell lines. Moreover, Cas13e.1 has a high activity for degrading SARS-CoV-2 sequences and the genome of live influenza A virus (IAV). Together with a minimal pool of 10 crRNAs, Cas13e.1 could target over 99% of all known 3,137 coronavirus genomes for achieving antiviral defense. Overall, our results demonstrated there exist untapped bacterial defense systems in natural microbes that can function efficiently in mammalian cells, thus potentially useful for preventing viral infection in humans such as COVID-19.

Main Text

We developed a fast computational pipeline (Extended Data Fig. 1a) in order to search for novel CRISPR-Cas13 systems from metagenomic data sets.

Using the CRISPR array as a search anchor, we first obtained all metagenomic assemblies from the JGI database⁸ and adapted existing algorithms for de novo CRISPR array detection⁹. This led to the identification of 340,425 putative CRISPR repeat arrays (Extended Data Fig. 1a). Up to 10 kilobases (kb) of genomic DNA sequence flanking each CRISPR array was extracted to further identify predicted protein-coding genes in the immediate vicinity. In order to identify compact Cas13 effectors for effectively *in vivo* delivery, we searched for 250,901 candidate proteins with 400-900 amino acid residues and within 10 protein-coding genes of the repeat, and found 24,959 proteins containing two predicted RxxxxH motifs of the HEPN ribonuclease domain separately located at the N- and C-terminus of the protein (Extended Data Fig. 1a). Among all RxxxxH motif-containing proteins, 64 contained two RxxxxH motifs of the three following types: RNxxxxH, RHxxxxH, and RQxxxxH. These three types were also found in the majority of previously known Cas13 (Extended Data Fig. 1b). Based on the fact that all reported CRISPR/Cas13 systems have a single crRNA with conserved stem-loop structure¹⁰, we identified thirty-one Cas13 candidates (Extended Data Fig. 1a). After excluding proteins with known functions in NCBI NR database, we obtained six candidate Cas13 proteins (Extended Data Fig. 1a). Further alignment of the six proteins back to the original pool of 24,959 proteins yielded one more candidate protein with RNxxxxH and RxxxxH motif (Extended Data Fig. 1a). Based on sequence divergence, seven candidate proteins with

exceptionally small sizes (775 to 803 amino acids) could be classified into two novel Cas13 families (Fig. 1a, b). We named them as “type VI-E” and “VI-F” families of Cas13, with two members (“Cas13e.1”, “Cas13e.2”) in VI-E, and five members (“Cas13f.1” to “Cas13f.5”) in VI-F (Fig. 1a, b).

The RNA-targeting activity of these seven novel Cas13 proteins was further screened in order to identify highly active Cas13 orthologs, using a eukaryotic cell-based mCherry reporter system (Extended Data Fig. 2a). By synthesizing the human codon-optimized version of each protein, we generated mammalian expression plasmids carrying the catalytically active or inactive proteins by mutating RxxxxH motifs^{11,12} (Extended Data Fig. 2b). Each protein was then fused with both N- and C-terminal nuclear localization signals (NLS). These VI-E and VI-F proteins were paired with two distinct forms of guide RNAs, either with a 30-nucleotide (nt) spacer flanked by two 36-nt direct repeat (DR) sequences to mimic an unprocessed guide RNA (pre-crRNA) or a 36-nt direct repeat with 30-nt spacer (crRNA) predicted to mimic mature guide RNAs (Fig. 1c, d). To determine crRNA architecture, we first tested DR position at 5’ or 3’ end of crRNA with reporter inhibition assay (Extended data Fig. 2c). The crRNA with 3’ DR instead of 5’ DR showed substantial suppression of reporter expression (Extended Data Fig. 2d), indicating that the crRNA accompanying Cas13e.1 shared similar 3’ DR structure with that of previously reported Cas13b13. We then assessed the ability of different VI-E and VI-F proteins to knock down the mCherry reporter level in cultured HEK293T cells. Two days after transfection with the plasmid expressing each of the VI-E and VI-F protein and corresponding single target-specific crRNA, we observed significant reduction of mCherry protein, with Cas13e.1 exhibiting the highest knockdown efficiency (Fig. 1d, e). In contrast, transfection with non-targeting (NT) crRNA together with each Cas13, or alternatively, crRNA with inactive Cas13, had no significant effect on the mCherry level (Fig. 1d, e, Extended Data Fig. 2d), suggesting crRNA- and HEPN-dependent knockdown. It was found that both the single DR crRNA and pre-crRNA with dual DR could mediate potent knockdown, and NLS significantly improved knockdown activity of Cas13e.1 (Fig. 1d, e). To determine the optimal spacer length for efficient Cas13e.1 targeting, we generated a series of spacer ranging from 20 to 50 nt in length (Extended Data Fig. 2e). Reporter inhibition activity dropped significantly below 30-nt spacer (Extended Data Fig. 2e), and thus crRNAs with 30-nt spacer were used for the following RNA interference experiments unless otherwise indicated.

We next sought to compare the knockdown efficiency of Cas13e.1 and Cas13f.1 against that of previously identified Cas13 proteins, Cas13a14, Cas13b13 and Cas13d12 (Extended Data Fig. 1c). Across three crRNA target loci in *mCherry*, Cas13e.1, Cas13f.1 and RfxCas13d overall outperformed over LwaCas13a and PspCas13b in HEK293T cells at 48 hr after transfection (Fig. 2a, Extended Data Fig. 3a). Further experiments in examining the knockdown efficiency for endogenous transcripts (3 genes, each with 3 crRNAs) showed that Cas13e.1 and RfxCas13d were similarly efficient (80.7 ± 2.1 % vs. 78.8 ± 2.9 %, mean \pm s.e.m, $p = 0.6$; Fig. 2b, Extended Data Fig. 3b). To confirm that RNA interference by Cas13e.1 is broadly applicable, we selected a panel of 12 additional human genes with diverse roles in mammalian cells, using 3 crRNA per gene. We found that Cas13e.1 consistently showed high-level knockdown activity for each gene, using any of the three crRNAs (Fig. 2c), indicating the uniformity of the Cas13e.1 system for RNA interference. Because Cas13 family is capable of processing its own CRISPR array¹², we next leveraged this property for the delivery of pre-crRNA for multiple targeting with a simple single-vector

system (Fig. 2d). We found that robust simultaneous knockdown of four RNA transcripts could be achieved by transfection of Cas13e.1 together with an array encoding four crRNAs, each tiling one mRNA (*EZH2*, *HRAS*, or *PPARG*) and a nuclear localized long non-coding RNA (lncRNA, *MALAT1*) (Fig. 2d). Furthermore, we also found that Cas13e.1 could achieve ~90% knockdown of mouse endogenous transcripts *Pten* in cultured mouse N2a cells (Extended Data Fig. 4a). Cas13e.1 targeting was specific, because transcriptome-wide RNA-seq analysis showed that more than half of top-ranked genes with altered expression were related with *Pten* (Extended Data Fig. 4b, Extended Data Table 2).

By fusion with various versions of the ADAR2 deaminase domain (ADAR2dd), Cas13 could also be engineered for RNA base editing^{15,16}, i.e., to convert adenine (A) to inosine (I) or cytosine (C) to uridine (U). We have fused dCas13e.1 with high fidelity ADAR2dd (with E488Q/T375G, referred as ADAR2dd*) to generate A-to-I RNA base editors (named as “eABE”). To test the activity of eABE, we generated an RNA-editing reporter using a mutated mCherry with a nonsense mutation [W98X (UGG to UAG)], which could functionally be repaired to the wild-type codon through A-to-I editing, and mCherry fluorescence could be detected after eABE editing (Extended Data Fig. 5a). We found that eABE indeed effectively induced mCherry fluorescence in cells transfected with mutant mCherry transcripts, together with both eABE and 50-nt crRNA, but not with either alone (Fig. 3a). To reduce the size of dCas13e.1 for efficient *in vivo* delivery, we generated various base editors by fusing the truncated dCas13e.1 (using structure-guided method) with ADAR2dd* (Fig. 3b, Extended Data Fig. 5b). We then systematically screened the editing activity of a variety of fused base editors with different dCas13e.1 truncations at either or both N- and C-terminus (Fig. 3b) in search of miniature editors, and identified the smallest and functional editor (“mini”) with 150 aa and 180 aa truncation at C- and N-terminus, respectively (Fig. 3b), suitable for packaging into commonly used adeno-associated virus (AAV). We then examined the effect of mismatched base position with 50-nt spacer on A-to-I editing efficiency for both full size and mini eABE (Extended Data Fig. 6a), and found that mismatched base position from 15 to 25 nt on crRNA sequence yielded higher editing efficiency than other positions (Extended Data Fig. 6b). The RNA editing efficiency of the full-size and mini eABE system was further examined in mammalian cells for several endogenous transcripts. We found A-to-I conversion were efficiently achieved by both editors (Fig. 3c, Extended Data Fig. 6c). To extend the base editing capability of the dCas13e.1 protein, we further generated C-to-U base editor (“eCBE”) by fusing full-length or truncated dCas13e.1 with RNA cytosine deaminase derived from evolved ADAR216, and found both full-length and mini eCBE could achieve efficient C-to-U editing as well in HEK293T cells (Fig. 3d).

To create effective and specific crRNA sequences to target and cleave SARS-CoV-2, we first performed a bioinformatics analysis by aligning published SARS-CoV-2 genomes¹⁷ and selected 30 crRNAs targeting RNA sites coding for RdRP (RNA-dependent RNA Polymerase) and E (envelop) proteins (with 15 crRNAs for each). Proof-of-concept experiments were performed on RdRP and E sequences that are conserved among SARS-CoV viruses (Fig. 4a).

The RdRP protein is the antiviral target for Remdesivir¹⁸ and E protein is critical for SARS-CoV pathogenesis¹⁹. To evaluate whether Cas13e.1 is effective for degrading SARS-CoV-2 sequences, we

created a reporter by fusing GFP with synthesized partial SARS-CoV-2 fragments of RdRP (genome coordinates 15,037-15,158 bp) and E (26,232-26,394 bp) (Fig. 4a). At 48 hours after co-transfection of HEK293T cells with the reporter and Cas13e.1/crRNAs, we observed that nearly all RdRP- and E-targeting crRNAs tested (27 out of 30) were able to support the suppression of GFP fluorescence in the cells by about 70%, as compared to that found for control transfection with the non-targeting crRNA (Fig. 4a).

To examine whether Cas13e.1 can tolerate mismatches between the crRNA and the targeted viral RNA, we calculated the knockdown activity for an example crRNA (SARS-CoV-2 crRNA_1) with 1 or 2 mismatches (Fig. 4b) and found that Cas13e.1 could well tolerate single nt mismatch at different positions on the example crRNA (Fig. 4b). Results on two-tandem mismatches revealed a critical (seed) region between 16-30 nt of the crRNA for efficient Cas13e.1-induced knockdown (Fig. 4b). We next examined the minimal number of crRNAs that required to target the majority of known coronaviruses found in both humans and animals, using a similar strategy previously described²⁰. From all known 3,137 coronavirus genomes, we identified approximately 7.1 million potential crRNA targets (Extended Data Fig. 7a). Based on the above results on the tolerance for nt mismatch, we estimated that only five 22-nt and six 30-nt crRNA with zero mismatch were able to target over 90% of coronavirus genomes (Extended Data Fig. 7b, c). With the tolerance for single-nt mismatch, we estimated that 3, 10 and 17 crRNAs could target 95.3%, 99.1% and 100% of all coronaviruses, respectively (Fig. 4c, d). The ability to use a relatively small number of crRNAs to broadly target nearly all coronavirus strains points to the uniqueness of Cas13e.1-based RNA interference approach, in contrast to traditional vaccination or pharmaceutical ones.

Next, we applied the CRISPR/Cas13e.1 strategy for inhibiting influenza RNA virus H1N1 that has a tropism for respiratory tract epithelial cells similar to SARS-CoV-2. We directly designed four crRNAs targeting at the nucleoprotein segment of the H1N1 genome that is essential for viral replication and transcription^{19,21}. To test antiviral ability of Cas13e.1 in a setting that mimics virus infection, we used an influenza H1N1 strain "A/Puerto Rico/8/1934"²² in the MDCK (Madin-Darby canine kidney) cell line (Fig. 4e). Compared with non-targeting crRNA, 3 out of 4 crRNAs showed high knockdown efficiency on the nucleoprotein transcript (Fig. 4f).

Consistently, target-specific crRNAs significantly reduced the abundance of nucleoprotein-positive H1N1 virus found in the supernatant of infected cultures, indicating effective inhibition of viral growth (Fig. 4g). Together, these results showed that Cas13e.1 system could be used to confer antiviral ability for mammalian cells.

In summary, by mining of metagenomic sequence data sets of natural uncultivated microbes, we identified two novel families of small CRISPR/Cas systems (type VI-E and VI-F), highlighting the diversity of natural microbial CRISPR systems. Furthermore, we demonstrated that Cas13e.1-based gene targeting can effectively cleave RNA sequences of SARS-CoV-2 fragments and influenza IAV in cultured cell models. Notably, Cas13e.1 exhibited good mismatch tolerance, capable of targeting over 99% coronaviruses using a minimal set of 10 crRNAs with single nucleotide mismatch. This may help to prevent the virus from escaping antiviral inhibition through mutation.

Compared with previously reported Cas13a/b/d, Cas13e.1 has a very compact size and robust RNA-guided ribonuclease activity, thus useful for *in vivo* RNA editing-based research and therapeutic applications²³⁻²⁶.

Methods

Computational identification of novel CRISPR/Cas13 systems

Metagenome sequences were downloaded from DOE JGI Integrated Microbial Genomes⁸. A computational pipeline was used to produce an expanded database of class 2 CRISPR-Cas systems from metagenomic sources. CRISPR arrays were identified using Piler-CR²⁷, with all default parameters. Proteins were predicted with Prodigal²⁸ in anon mode on all contigs at least 5 kb in length, and de-duplicated (i.e., removing identical protein sequences) to construct a database. Proteins with length between 400 to 900 residues were obtained. RNAfold (<http://rna.tbi.univie.ac.at/>) were used to predict the secondary structure of direct repeat sequences. For functional characterization of the candidate Cas proteins, protein family databases NR database, and Cas proteins in NCBI were used to annotate the candidate Cas proteins. Multiple sequence alignment was then conducted for each candidate Cas effector proteins using MAFFT²⁹. MEGA30 were used to construct the phylogenetic tree. I-TASSER³¹ were used to perform the protein structure prediction.

Plasmid constructions

Retrieved coding sequences of Cas13e and Cas13f were human-codon optimized and synthesized for cloning into Sall/NotI digested pCX539 backbone with Gibson Assembly method. Predicted direct repeat sequences for each Cas13 variants were synthesized as oligos for cloning downstream of human U6 promoter for expression in mammalian cells. A G>A amber mutation was introduced in the mCherry coding sequence to generate mutant fluorescence protein as RNA base-editing reporter. All primers and Cas13 sequences used in this study were provided in Extended Data table 1 and supplementary sequence 1-14.

Cell culture, transfection and flow cytometry analysis

Mammalian cell lines used in the study were HEK293T and N2A. Media for culturing cells were prepared by supplementing DMEM with 10% FBS, GlutMAX, Sodium Pyruvate, and Penciline/Streptomycin. Transfection of HEK293T and N2A cells was conducted with lipofectamine 3000 following manufacture manual, and cells were analyzed by BD FACSAria II or sorted by MoFlo XDP at 48 hr after transfection. Flow cytometry results were analyzed with FlowJo X (v10.0.7).

RNA editing and sequencing analysis

To analyze A-to-I or C-to-U base editing efficiency of dCas13e, successfully transfected cells were sorted for RNA extraction. RNA was extracted with RNA-easy Isolation Reagent according to the manufacturer's protocol. The cDNAs were reverse transcribed from RNAs by HiScript II One Step RT-PCR Kit, and crRNA target sites were amplified from cDNAs with Phanta Max Super-Fidelity DNA Polymerase for Sanger or deep sequencing methods.

Deep sequencing library were prepared with Nextera XT DNA Library Prep Kit according the manufacturer's manual and sequenced on Hiseq. Sequencing data were first de-multiplexed by Cutadapt (v2.8)³² based on sample barcodes. The de-multiplexed reads were then processed by CRISPResso2³³ for the quantification of A-to-I or C-to-U conversion efficiency at each target site. Sanger sequencing results were analyzed with EditR³⁴ to quantify A-to-I or C-to-U conversion efficiency at each target site.

qRT-PCR, RNA-seq and analysis

To quantify RNA knockdown efficiency of Cas13 effectors, RNAs were extracted from successfully transfected cells and reverse-transcribed to cDNAs with HiScript II One Step RT-PCR Kit (Vazyme, Biotech). qPCR was performed with the cDNA for each sample on Roche 480 II-A, using AceQ Universal SYBR qPCR Master Mix (Vazyme, Biotech). qPCR results were analyzed with $-\Delta\Delta CT$ method.

To analyze functional specificity of Cas13 effectors, RNAs were extracted with TRIZOL (Ambion)-based method, fragmented and reverse transcribed to cDNAs with HiScript II One Step RT-PCR Kit according to the manufacturer's protocol. RNA-seq library was generated with TruSeq Stranded Total RNA library preparation kit using the standard protocol. The transcriptome libraries were sequenced using 150 bp paired-end Illumina Xten platform. RNA-seq data were analyzed as previously described²⁵ and presented as the mean of all repeats. After filtering the low-quality reads with SolexaQA (V3.1.7.1)³⁵, RNA-seq reads were aligned to the mm10 reference genome with Hisat2 (V2.0.4)³⁶. All uniquely mapped reads were used to calculate the read counts with htseq-count (v0.11.2)³⁷. DEseq2 (1.24.0)³⁸ were used to calculate differentially expressed genes. Genes with fold change >2 and FDR < 0.05 were treated as differentially expressed genes.

Bioinformatics analysis of coronaviruses

We downloaded 3,137 coronavirus genomes from NCBI database. MAFFT29 were used to align other coronavirus genomes to the SARS-CoV-2 genome to calculate the identity and depth with 100-bp window size. To calculate the minimal numbers of crRNAs to target all the coronavirus genomes. We first used Prodigal28 to perform the gene prediction. BLASTP39 was used to conduct pairwise alignment between the S proteins, and BLASTP alignment results with similarity > 50% were obtained. MCL was then used to further cluster S protein based on the BLASTP results in order to create clusters of S proteins. We select one representative genome from resulting S protein groups and produced 78 representative genomes in total. Next, we used a sliding window method to extract all unique 30-nt or 22-nt sequences in the coronavirus genomes, yielding 7,132,478 and 5,931,021 unique sequences for each. The unique sequences were sorted in descending order by the number of coronavirus genomes it appears. We used the 78 representative genomes to create a barcode for each unique sequences. The unique sequences that appears in the most genomes were added to the minimal crRNA pool, and all of the genomes targeted by the pool were removed. Using this approach, forty-two 30-crRNAs and twenty-two 22-nt crRNAs were identified that were able to target all 3,137 Coronavirinae virus genomes. We next used the same way calculate the minimal numbers of crRNAs with no more than one mismatch with the coronavirus genomes and seventeen 30-nt crRNA with single mismatches were identified.

Antiviral experimental and analysis method

Madin-Darby canine kidney (MDCK) cells were seeded onto 96-well plates and incubated with Dulbecco's modified Eagle medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% penicillin-streptomycin. The cells were further infected with influenza A virus H1N1 (A/Puerto Rico/8/1934)22 at the dose of 100TCID₅₀. At 1 hr post-infection, the medium was replaced with DMEM containing 0.1% BSA and 1 µg/ml of TPCK-trypsin. At 48 hrs post-infection, supernatant and cell lysis were collected for measuring virus titers. Total RNA was extracted from supernatants of virus-infected MDCK cells, and quantitative reverse transcription PCR (qRT-PCR) was performed using influenza virus-specific primers for determination of relative levels of viral loads. All primers used in this study were provided in supplementary table 1.

Statistical analysis

All values are shown as mean ± s.e.m. Unpaired Student's t-test (two-tailed) was used for comparisons and $P < 0.05$ was considered to be statistically significant. Details of statistical values are provided in Supplementary Tables. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Data availability

All the sequencing data have been deposited in the NCBI SRA under project accession numbers PRJNA629900. All materials are available upon reasonable request.

Code availability

The authors declare that all code used in this study are available within the article and its Extended Data or from the corresponding author upon reasonable request.

Declarations

Acknowledgements We thank Dr. Mu-ming Poo for discussions and comments on this manuscript; FACS facility, H. Wu and L. Quan in ION. This work was supported by R&D Program of China (2017YFC1001300 and 2018YFC2000100), CAS Strategic Priority Research Program (XDB32060000), National Natural Science Foundation of China (31871502, 31925016, 91957122, 31901047), Basic Frontier Scientific Research Program of Chinese Academy of Sciences From 0 to 1 original innovation project (ZDBS-LY-SM001), Shanghai Municipal Science and Technology Major Project (2018SHZDZX05), Shanghai City Committee of Science and Technology Project (18411953700, 18JC1410100, 19XD1424400, 19YF1455100), and International Partnership Program of Chinese Academy of Sciences (153D31KYSB20170059).

Author contributions C.X., Y.Z. and H.Y. conceived the project. C.X., Y.Z., Q.X., B.H. and G.G. designed and conducted experiments. Y.Z. performed bioinformatics analysis. Z.W. and B.C. assisted with plasmids construction and RNA analysis. Y.T. assisted with cell experiments. X.W., D.Z. and X.H. assisted with virus experiments. H.Y. designed experiments and supervised the whole project. C.X., Y.Z. and H.Y. wrote the paper.

Competing interests The authors declare no competing interests.

References

1. Organization, W. H. Coronavirus disease 2019 (COVID-19) Situation Report *World Health Organization* **8** (2020).
2. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science (New York, Y.)* **356**, doi:10.1126/science.aal5056 (2017).
3. Faure, *et al.* CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nature reviews. Microbiology* **17**, 513-525, doi:10.1038/s41579-019-0204-7 (2019).

4. Makarova, K. *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature reviews. Microbiology* **18**, 67-83, doi:10.1038/s41579-019-0299-x (2020).
5. Burstein, *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237-241, doi:10.1038/nature21059 (2017).
6. Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science (New York, Y.)* **362**, 839-842, doi:10.1126/science.aav4294 (2018).
7. Yan, X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science (New York, N.Y.)* **363**, 88-91, doi:10.1126/science.aav7271 (2019).
8. Markowitz, M. *et al.* IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic acids research* **40**, D115-122, doi:10.1093/nar/gkr1044 (2012).
9. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell* **60**, 385-397, doi:10.1016/j.molcel.2015.10.008 (2015).
10. O'Connell, M. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. *Journal of molecular biology* **431**, 66-87, doi:10.1016/j.jmb.2018.06.029 (2019).
11. Abudayyeh, O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (New York, N.Y.)* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016).
12. Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676.e614, doi:10.1016/j.cell.2018.02.033 (2018).
13. Smargon, A. *et al.* Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular cell* **65**, 618-630.e617, doi:10.1016/j.molcel.2016.12.023 (2017).
14. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284, doi:10.1038/nature24049 (2017).
15. Cox, D. B. *et al.* RNA editing with CRISPR-Cas13. *Science (New York, N.Y.)* **358**, 1019-1027, doi:10.1126/science.aaq0180 (2017).
16. Abudayyeh, O. O. *et al.* A cytosine deaminase for programmable single-base RNA editing. *Science (New York, Y.)* **365**, 382-386, doi:10.1126/science.aax7063 (2019).
17. Hulo, *et al.* ViralZone: a knowledge resource to understand virus diversity. *Nucleic acids research* **39**, D576-582, doi:10.1093/nar/gkq901 (2011).
18. de Wit, E. *et al.* Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 6771-6776, doi:10.1073/pnas.1922083117 (2020).
19. Ruch, R. & Machamer, C. E. The coronavirus E protein: assembly and beyond. *Viruses* **4**, 363-382, doi:10.3390/v4030363 (2012).
20. Abbott, R. *et al.* Development of CRISPR as an Antiviral Strategy to Combat SARS- CoV-2 and Influenza. *Cell*, doi:10.1016/j.cell.2020.04.020 (2020).

21. Freije, A. *et al.* Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Molecular cell* **76**, 826-837.e811, doi:10.1016/j.molcel.2019.09.013 (2019).
22. Harding, T., Heaton, B. E., Dumm, R. E. & Heaton, N. S. Rationally Designed Influenza Virus Vaccines That Are Antigenically Stable during Growth in Eggs. *mBio* **8**, doi:10.1128/mBio.00669-17 (2017).
23. Zhou, *et al.* Glia-to-Neuron Conversion by CRISPR-CasRx Alleviates Symptoms of Neurological Disease in Mice. *Cell*, doi:10.1016/j.cell.2020.03.024 (2020).
24. Zhou, C. *et al.* CasRx-mediated RNA targeting prevents choroidal neovascularization in a mouse model of age-related macular degeneration. *National Science Review*, doi:10.1093/nsr/nwaa033 (2020).
25. He, *et al.* Modulation of metabolic functions through Cas13d-mediated gene knockdown in liver. *Protein & cell*, doi:10.1007/s13238-020-00700-2 (2020).
26. Kushawah, G. *et al.* CRISPR-Cas13d induces efficient mRNA knock-down in animal embryos. 2020.2001.2013.904763, doi:10.1101/2020.01.13.904763 %J bioRxiv (2020).
27. Edgar, C. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC bioinformatics* **8**, 18, doi:10.1186/1471-2105-8-18 (2007).
28. Hyatt, *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).
29. Katoh, K. & Standley, M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772- 780, doi:10.1093/molbev/mst010 (2013).
30. Kumar, , Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* **33**, 1870-1874, doi:10.1093/molbev/msw054 (2016).
31. Zhang, I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **9**, 40, doi:10.1186/1471-2105-9-40 (2008).
32. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 %J EMBnet.journal, doi:10.14806/ej.17.1.200 (2011).
33. Clement, *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature biotechnology* **37**, 224-226, doi:10.1038/s41587-019-0032- 3 (2019).
34. Kluesner, M. *et al.* EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR journal* **1**, 239-250, doi:10.1089/crispr.2018.0014 (2018).
35. Cox, P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics* **11**, 485, doi:10.1186/1471-2105-11-485 (2010).
36. Kim,, Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
37. Anders, S., Pyl, T. & Huber, W. HTSeq—a Python framework to work with high- throughput sequencing data. *Bioinformatics (Oxford, England)* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

38. Love, I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
39. Altschul, F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410, doi:10.1016/s0022-2836(05)80360-2 (1990).

Figures

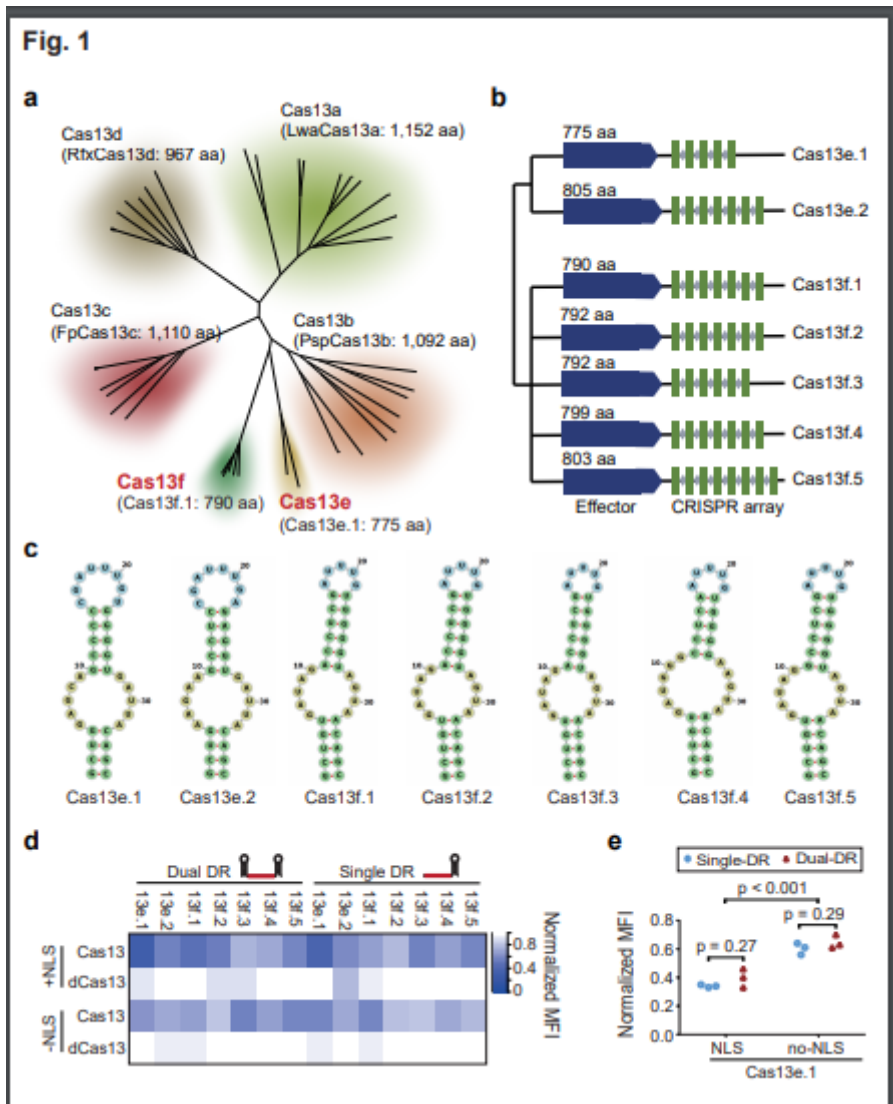


Figure 1

Computational identification and characterization of novel type VI-E and VI-F CRISPR systems. a, Maximum-likelihood tree of newly identified Cas13e, Cas13f and previously reported Cas13a, Cas13b, Cas13c and Cas13d. Commonly used family members are shown in parentheses. b, Maximum-likelihood phylogenetic tree of Cas13e and Cas13f proteins identified in this study, with the full Cas13e and Cas13f CRISPR loci drawn along with conserved HEPN RNase domains. Blue and green rectangles indicate Cas13 proteins and CRISPR direct repeats (DRs), respectively. Grey diamonds denote spacer sequences. c, Predicted secondary RNA structure of direct repeat sequence for Cas13e and Cas13f proteins. d,

Heatmap of mCherry protein knockdown activity in a Cas13e and Cas13f orthologs screen in HEK293T cells using pre-crRNA or crRNA. e, Effect of pre-crRNA vs crRNA and NLS on knockdown activity of Cas13e.1. Normalized MFI, mean fluorescent intensity relative to the non-targeting condition (n = 3). All Values shown are mean \pm s.e.m. P value, two-sided unpaired t-test.

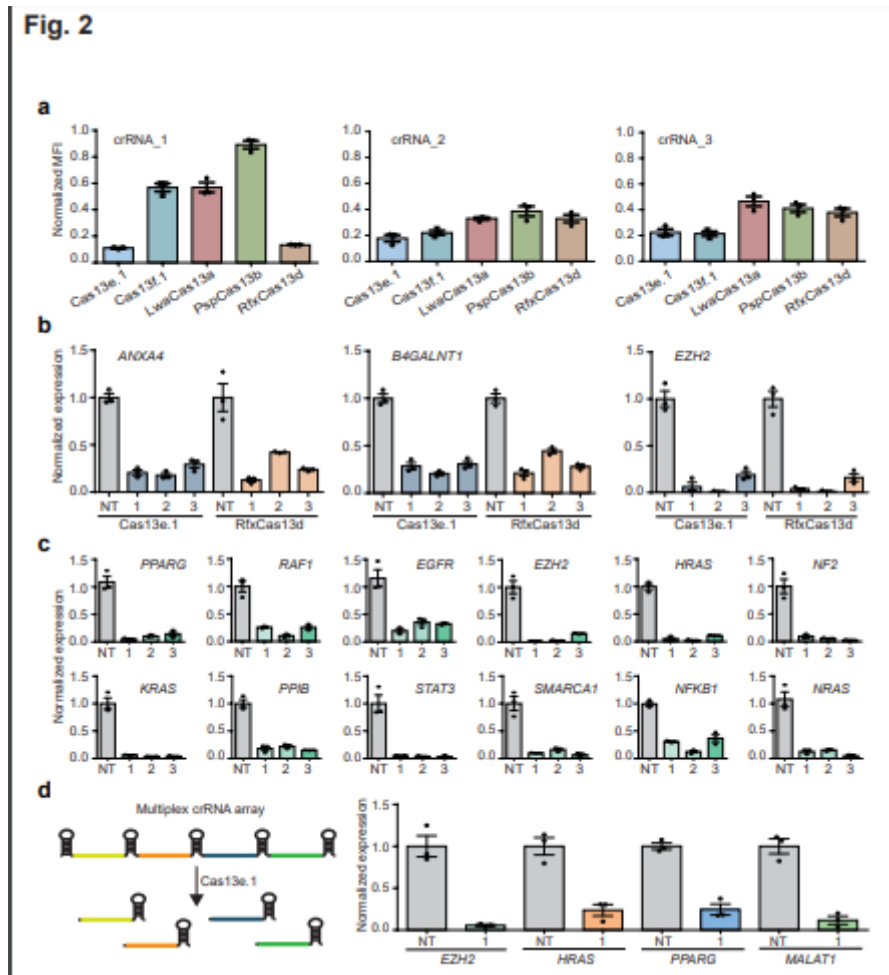


Figure 2

Efficient and specific interference activity of Cas13e.1 against transcripts in HEK293 cells. a, Reporter inhibition assay results of comparing activity among Cas13e.1, Cas13f.1, LwaCas13a, PspCas13b and RfxCas13d for three different mCherry-targeting crRNAs in HEK293T. Normalized MFI, mean fluorescent intensity (MFI) relative to the non-targeting condition (n = 3). b, Relative target RNA knockdown by individual position- matched Cas13e.1 and RfxCasRx crRNAs. c, Cas13e.1 targeting of 12 endogenous transcripts, each with 3 guides and a non-targeting (NT) crRNA in HEK293T cells. d, Arrays of 4 guides, each mediates target knockdown by Cas13e.1 in HEK293T cells via transient transfection. Knockdown relative to a NT crRNA was determined by qPCR. NT, non-targeting crRNA. All Values shown are mean \pm s.e.m (n = 3).

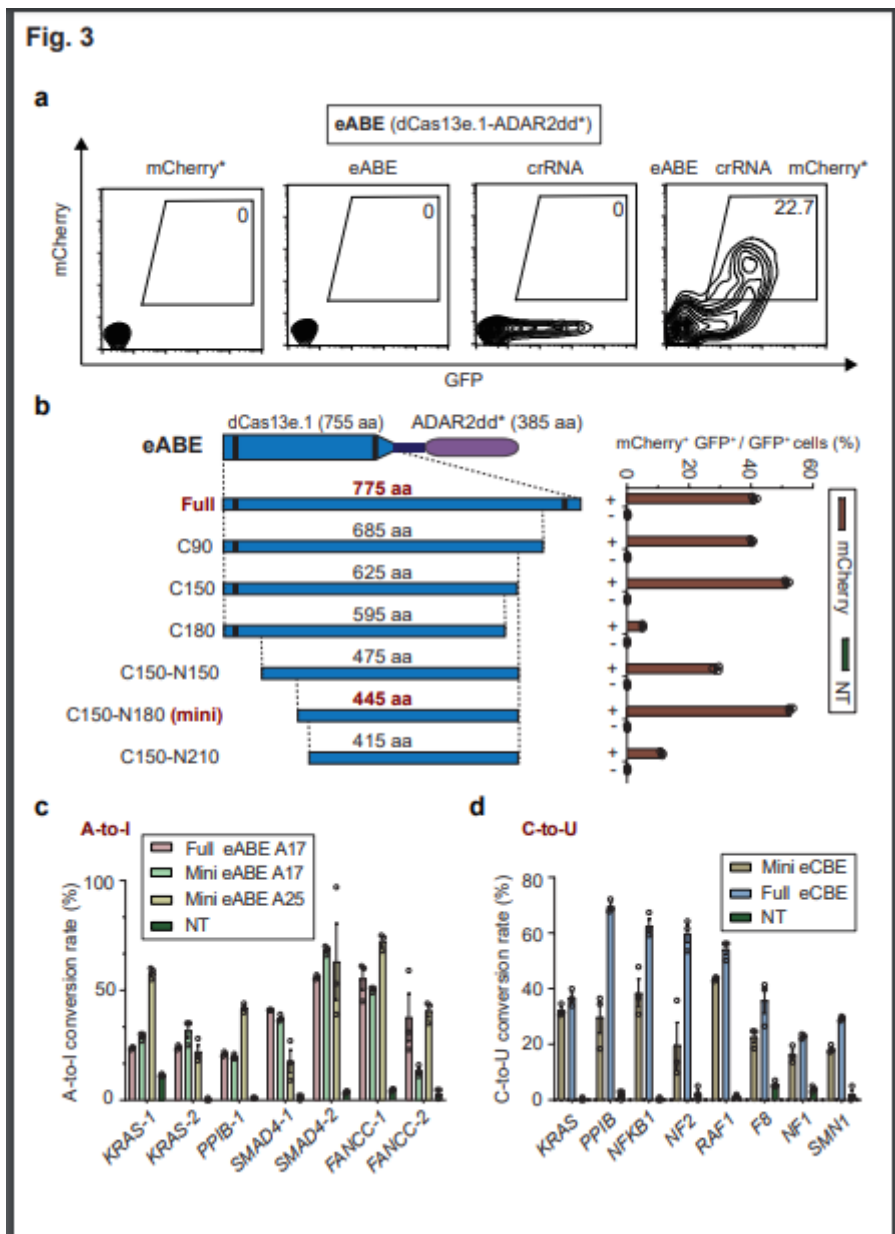


Figure 3

Truncated dCas13e.1 with ADAR2dd variants for efficient RNA base editing. a, Restored expression of mutant reporter by eABE RNA base editor as measured with flow cytometry. b, Screening results for a variety of truncated eABE variants analyzed by reporter assay. c, Comparison results between editing efficiency of full-length and mini eABE on endogenous transcripts in HEK293T analyzed with Sanger sequencing. d, Comparison results between editing efficiency of full-length and mini eCBE on endogenous transcripts in HEK293T analyzed with deep sequencing. NT, non-targeting crRNA. All values are presented as mean \pm s.e.m (n = 3).

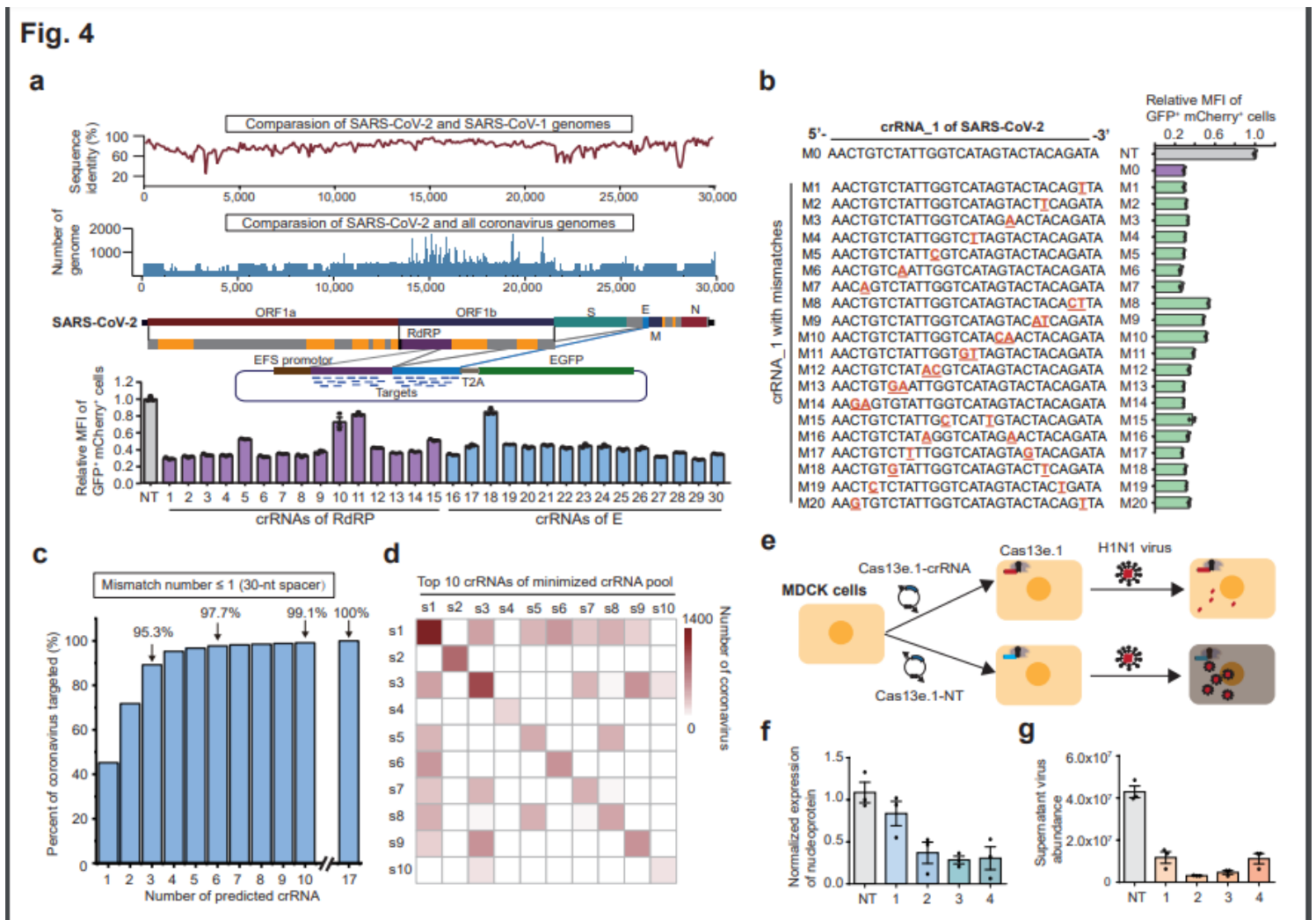


Figure 4

Antiviral activity of Cas13e.1 in mammalian cells. a, Top, Alignment of currently available SARS-CoV-2 and SARS-CoV-1 sequences. Middle, Schematic diagram of the reporter consists of EFS promoter, GFP and the synthesized RdRP and E fragment sequences. RdRP, RNA-dependent RNA Polymerase; E, envelop. Bottom, GFP expression after co-transfection of the reporter and Cas13e.1/crRNA, as measured by flow cytometry. Mean GFP fluorescent intensity changes of the reporter caused by 30 different targeting crRNAs, relative to non-targeting (NT) crRNA. b, Left, crRNA nucleotide identity or mismatch types (single-nt mismatch or two tandem nt mismatch). Right, Changes of GFP fluorescence intensity caused by co-transfection of reporter, Cas13e.1 together with each of 20 different versions of crRNA_1 with mismatched mutation, relative to NT crRNA. c, A histogram showing the predicted minimum number of crRNAs with single-nt mismatch for targeting all available 3,137 coronavirus genomes. d, A heatmap showing distribution pattern for the predicted minimum number of crRNAs with single-nucleotide mismatch to target all sequenced 3,137 coronavirus genomes. e, Procedure for testing Cas13e.1-mediated anti-H1N1 activity in H1N1 IAV-infected MDCK cells. f, IAV RNA knockdown efficiency resulting from transfection of NP- targeting Cas13e.1/crRNA. Transcript levels are relative to NT crRNA control. g,

Changes of IAV abundance following Cas13e.1/crRNA transfection, analyzed by absolute qRT-PCR of supernatant from infected cultures. All values in a, b, f and g are shown as mean \pm s.e.m (n = 3).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [Extendeddatafigures.pdf](#)