www.sciencemag.org/cgi/content/full/science.aav7271/DC1

# Supplementary Materials for

## Functionally diverse type V CRISPR-Cas systems

Winston X. Yan*, Pratyusha Hunnewell*, Lauren E. Alfonse, Jason M. Carte, Elise Keston-Smith, Shanmugapriya Sothiselvam, Anthony J. Garrity, Shaorong Chong, Kira S. Makarova, Eugene V. Koonin, David R. Cheng, David A. Scott†

*These authors contributed equally to this work.
†Corresponding author. Email: dscott@arbor.bio

Published 6 December 2018 on *Science* First Release
DOI: 10.1126/science.aav7271

**This PDF file includes:**

Materials and Methods
Figs. S1 to S19
Tables S1 to S10
References

## Materials and Methods

Pipeline for Class 2 CRISPR-Cas loci identification

Genome and metagenome sequences were downloaded from NCBI (*29*, *30*), NCBI whole genome sequencing (WGS), and DOE JGI Integrated Microbial Genomes (*31*) and compiled to construct a database of 293,985 putative CRISPR-Cas loci within which we searched for novel RuvC containing type V effectors. As previously described in detail (*19*), this approach to pipeline engineering performs minimal filtering in the intermediate stages in order to expand the search space for novel CRISPR effector discovery and reduce biases. Sequences for Cas12g-i and Cas12c characterized in this study are included in Table S1.

Classification of type V systems

Subtypes and effector gene names were assigned according to the multiple-criteria classification approach described previously *(9, 33)*. Specifically, the presence of a single, RuvC-like nuclease domain which is similar to TnpB and the previously described type V effectors indicates that all the CRISPR-*cas* loci described in this work belong to type V. The signature protein of type V is Cas12, with subtypes denoted by letters. Subtypes V-A through E and the corresponding signature effectors Cas12a-e have been assigned previously *(1, 11, 14)*. The recently described Cas14 proteins are (predicted) V-U effectors (**Fig. 1,** *11*; U standing for Uncharacterized) which, once characterized, may receive a dedicated subtype designation, such as the V-F suggested by Harrington and colleagues *(20)*. Accordingly, we considered the corresponding Cas12f effector designation 'in use' and denoted the three previously undescribed CRISPR-Cas effectors

studied in this work Cas12g,h,i. All these proteins show only weak sequence similarity (e > $10^{-3}$) to the currently characterized Cas12 proteins (**Fig. 1a**). There are also functional considerations that support the classification of the 3 CRISPR-Cas systems described here as distinct subtypes. In particular, Cas12g shows specific enzymatic properties that have not been identified in any previously characterized type V systems. Cas12h, unlike Cas12b, lacks an adaptation module and does not require tracrRNA, although a common ancestor of Cas12b and Cas12h cannot be ruled out. Cas12i is the most diverged of the analyzed proteins and cannot be grouped with any subtype based on sequence similarity, and lacks an adaptation module as well.

The classification tree was constructed by comparing sequence profiles extracted from multiple alignments of groups of readily alignable Cas12 proteins. Profile-profile comparisons were performed using HHsearch (*32*); scores between two profiles were normalized by the minimum of the self-scores and converted to a distance matrix on the natural log scale. The UPGMA dendrogram was reconstructed from the distance matrix. The tree at the depth of 2 distance unites (corresponding to the pairwise HHsearch score of $e^{-2D}$ = 0.02 relative to the self-score) typically reliably recovers profile similarity and can serve as a guide for subtype classification (*11*). All multiple alignments used for the tree construction are available at ftp://ftp.ncbi.nlm.nih.gov/pub/makarova/supplement/Cas12.

Gene synthesis & oligo library cloning

The effector plasmid was created by synthesizing the *E. coli* codon-optimized protein sequences of CRISPR effectors and/or accessory proteins under the transcriptional control of *lac* and IPTG-inducible T7 promoters into a pET-28a(+) plasmid backbone (EMD-Millipore, "Effector plasmid", **Fig. 1B, Table S2**). To capture the required noncoding sequences for each system, such as the tracrRNA, the noncoding sequences flanking the CRISPR array, putative effector or accessory open reading frames, as well as predicted anti-repeats indicative of tracrRNA elements, were concatenated and cloned into pACYC184 (New England Biolabs) under the transcriptional control of *lac* and IPTG-inducible T7 promoters ("Noncoding plasmid", **Fig. 1B, Table S2**). Effector mutants (e.g., dCas12g1 D513A or Cas12g1 A513D) plasmids were cloned by site directed mutagenesis: sequence changes were first introduced into PCR fragments using custom DNA primers (IDT), which were then re-assembled into a plasmid using NEBuilder HiFi DNA Assembly Master Mix or NEB Gibson Assembly Master Mix (New England Biolabs) following the manufacturer's instructions.

To generate the minimal CRISPR array for the *in vivo* bacterial screening plasmid library, we designed oligonucleotide library synthesis (OLS) pools (Agilent) with a "repeat-spacer-repeat" construction, with two full-length direct repeats flanking a spacer sequence. The "repeat" elements were derived from the consensus direct repeat sequence found in the CRISPR array associated with the effector, and the "spacer" sequences, with its length determined by the mode of spacers found in the endogenous CRISPR array, contained sequences complementary to the pACYC184 plasmid and select *E. coli*

essential genes, as well as nontargeting negative control spacers, for a total of 8900 library elements. Flanking the minimal CRISPR array were unique PCR priming sites that enabled amplification of a specific library from a larger pool of oligo synthesis.

For the minimal CRISPR array, we designed oligonucleotide library synthesis (OLS) pools comprising two direct repeats flanking natural-length spacer sequences. Spacer sequences were selected targeting the pACYC184 plasmid, select *E. coli* essential genes, and nontargeting negative control spacers for a total of 8900 elements in the array library. To test both directions of array expression, the array library was cloned into the effector plasmid in both the forward and reverse orientations under the control of a J23119 promoter (NEB Golden Gate Assembly Master Mix, New England Biolabs). The resulting effector plasmid library contained a total of ~18,000 plasmid elements, each containing the Cas effector, accessory proteins, and a unique CRISPR array. The effector plasmid library also represented the "input library" for *in vivo* negative selection screens.

Effector Plasmid and Noncoding Plasmid sequences for Cas12g-i and Cas12c are included in Table S2.

*In vivo E. coli* screen

The *in vivo* screen was performed as previously described (*19*), with the following enhancements to enable a greater ability to systematically and quantitatively characterize the activity of CRISPR-Cas systems, including: 1) inclusion of noncoding plasmid for identification of tracr RNAs or other required noncoding elements; 2) targeting of *E. coli*

essential genes in addition to pACYC184 for enhanced understanding of constraints on the interference of each system; 3) inclusion of unique molecular identifiers (UMIs) preceding the effector plasmid library CRISPR arrays to enhance the accuracy of screen analysis.

The CRISPR Effector Plasmid library and/or the separate Noncoding Plasmid were then co-transformed into electrocompetent *E. coli*. Unless otherwise indicated, E. cloni EXPRESS BL21(DE3) *E. coli* (Lucigen) were used and transformed using a Gene Pulser Xcell (Bio-rad) following the manufacturer's instructions. The cells were plated onto bioassay plates containing both Chloramphenicol (Fisher) and Kanamycin (Alfa Aesar), and grown for 11 hours at 37ºC, after which we estimated the approximate colony count to ensure sufficient library representation and harvested the cells. During outgrowth, RNA-programmable CRISPR-Cas interference would result in spacer-specific loss of plasmid-based antibiotic resistance or self-targeting of essential genes, leading to depletion of bacteria containing active CRISPR systems.

Plasmid DNA fractions were extracted from the harvested cells to create the "output library" (effector plasmid library recovered post-outgrowth) using a QIAprep Spin Miniprep Kit (Qiagen), while total RNA ≥17nt was harvested by lysing the harvested cells in Direct-zol (Zymo Research), followed by extraction using the Direct-zol RNA miniprep kit (Zymo Research).

Screens were read out by targeted next generation sequencing (NGS) (Nextseq 550, Illumina) of CRISPR array elements in the "input library" (pre-transformation effector plasmid library) and "output library". NGS libraries were prepared by performing a PCR using custom primers flanking the CRISPR array cassette and containing barcodes and handles compatible with Illumina sequencing chemistry. The NGS library was then normalized, pooled, and loaded onto a Nextseq 550 (Illumina) to evaluate the activity of the effectors.

NGS screen data were processed to identify strongly depleted CRISPR arrays conferring interference activity as described previously (*19*). At least two independent biological replicates were performed for each screen to ensure consistency, and data shown represents the intersection of two representative biological replicates. Strongly depleted spacers are described as those with greater than 3-fold decrease in normalized frequency in the output library relative to the input, referred to as the "fold depletion". The RNA fractions were prepared for small RNA sequencing and analyzed as described previously for the identification of required noncoding elements for activity, such as the mature crRNA and tracrRNA (*19*).

Purification of effector proteins

Effector vectors were transformed into *E. coli* NiCo21 (DE3) (New England BioLabs) and expressed under a T7 promoter. Transformed cells were initially grown overnight in 3mL Luria Broth (Sigma) + 50ug/mL kanamycin, followed by inoculation of 1L of Terrific Broth media (Sigma) + 50ug/mL kanamycin with 1mL of overnight

culture. Cells were grown at 37ºC until an OD600 of 1-1.5, then protein expression was induced with 0.2mM IPTG. Cultures were then grown at 20ºC for an additional 14-18 hours. Cultures were harvested and pelleted via centrifugation, then resuspended in 80mL of lysis buffer (50mM HEPES pH 7.6, 0.5M NaCl, 10mM imidazole, 14mM 2-mercaptoethanol, and 5% glycerol) + protease inhibitors (Sigma). Cells were lysed via cell disruptor (Constant System Limited), then centrifuged twice at 28,000xg for 20 minutes at 4ºC in order to clarify the lysate. The lysate was loaded onto a 5mL HisTrap FF column (GE Life Sciences), then purified via FPLC (AKTA Pure, GE Life Sciences) over an imidazole gradient from 10mM to 250mM. Cas12i1 and Cas12g1 were purified in low salt buffer (50mM HEPES-KOH pH 7.8, 500mM KCl, 10mM $MgCl_2$, 14mM 2-mercaptoethanol, and 5% glycerol) and high salt buffer (50mM HEPES-KOH pH 7.8, 1M $NH_4Cl$, 10mM $MgCl_2$, 14mM 2-mercaptoethanol, and 5% glycerol), respectively. After purification, fractions were run on SDS-PAGE gels, and fractions containing protein of the appropriate size were pooled and concentrated using 10kD Amicon Ultra-15 Centrifugal Units. Cas12g1 was further dialyzed into a buffer without imidazole (25mM HEPES-KOH pH 7.8, 1M $NH_4Cl$, 10mM $MgCl_2$, 1mM DTT, 7mM 2-mercaptoethanol, and 30% glycerol). Protein concentration was determined by Qubit protein assay (Thermo Fisher).

crRNA, tracrRNA and substrate RNA preparation

Single stranded DNA oligo templates for crRNA, tracrRNA and substrate RNA were ordered from IDT. Substrate RNA and pre-crRNA templates were PCR amplified to generate a double-stranded *in vitro* transcription (IVT) template DNA using NEBNext

High-Fidelity 2X PCR Master Mix (New England BioLabs). Double-stranded DNA templates for mature crRNA and tracrRNA were generated by annealing T7 primer with templates followed by extension using DNA Polymerase I, Large (Klenow) Fragment (New England BioLabs). Annealing was performed by incubating for 5 min at 95˚C followed by a -5˚C/min ramp down to 4˚C. *In vitro* transcription was performed by incubating the dsDNA templates with T7 RNA polymerase at 37˚C for 3 hours using HiScribe T7 Quick High Yield RNA Synthesis kit (New England BioLabs). After incubation, IVT samples were treated with Turbo DNase (Thermo Fisher Scientific) and then purified using RNA Clean & Concentrator kit (Zymo Research). For Cas12g1, equimolar amounts of mature crRNA containing 5' hydroxyl and 5' monophosphate end group was used. Mature crRNA generated from IVT was treated with Calf Intestinal Alkaline Phosphatase (Thermo Fisher Scientific) or RNA 5'-polyphoshpatase (Lucigen) for 2 hours at 37˚C to generate 5'-hydroxyl or 5'-monophosphate, respectively, followed by clean up with RNA Clean & Concentrator kit (Zymo Research). Concentrations were measured via Nanodrop 2000 (Thermo Fisher Scientific).

Pre-crRNA, tracrRNA, and mature crRNA sequences used in biochemical characterization of Cas12g and Cas12i are included in Table S3, S4, and S5, respectively. Oligonucleotide templates and primers for preparation of crRNAs are included in Table S9.

Preparation of IR800-labeled substrate RNA and DNA

RNA substrates from IVT were treated with Calf Intestinal Alkaline Phosphatase (Thermo Fisher Scientific) for 30 minutes at 37˚C to convert the 5'-triphosphate to 5' terminal hydroxyl group and purified using RNA Clean & Concentrator kit (Zymo Research). A thiol end group was added to the 5' terminal hydroxyl group of the DNA and RNA substrates via 5' EndTag Labeling Kit (Vector Labs), then substrates were labeled with IRDye 800CW Maleimide (LI-COR Biosciences). Substrates were purified using DNA Clean & Concentrator kit or RNA Clean & Concentrator kit (Zymo Research). Labeled dsDNA substrates were generated by labeling the non-spacer-complementary (NSC) strand, annealing with a primer, then extending with DNA Polymerase I, Large (Klenow) Fragment (New England BioLabs) for 15 minutes at 25˚C. These substrates were purified with DNA Clean & Concentrator kit (Zymo Research). Concentrations were measured via Nanodrop 2000 (Thermo Fisher Scientific).

RNA and DNA target, nontarget, and collateral substrate sequences used in biochemical characterization of Cas12g and Cas12i are included in Table S6-8, respectively. Oligonucleotide templates and primers for preparation of RNA substrates are included in Table S10.

Cas12g1 and Cas12i1 pre-crRNA processing assay

Pre-crRNA processing assays for Cas12g1 were performed at 37˚C for 1 hour in cleavage buffer at a final pre-crRNA concentration of 25nM. Pre-crRNA processing assays for Cas12i1 were performed at 37˚C for 30 minutes in cleavage buffer at a final pre-crRNA concentration of 100nM. The reaction was performed in NEBuffer 3 (New

England BioLabs) for Cas12g and optimized cleavage buffer (50mM Tris-HCl pH 8.0, 50mM NaCl, 1mM DTT, 10mM MgCl2, 50ug/ml BSA) for Cas12i. For Cas12g, pre-crRNA processing was performed with 5' IR800-labeled pre-crRNA to distinguish from the tracrRNA. The tracrRNA was added at equimolar concentrations to Cas12g1. Reactions were quenched with the addition of 1ug/uL of Proteinase K (Ambion) and incubated at 37˚C for 15 minutes. 50mM EDTA was added to the reactions before mixing with equal volume of 2X TBE-Urea sample buffer (Invitrogen) and denaturing at 65˚C for 3 minutes. Samples were analyzed on 15% TBE-Urea gels (Invitrogen). Gels were stained for 5 minutes with SYBR Gold nucleic acid stain (Invitrogen) and imaged on Gel Doc EZ (Bio-Rad). Gels containing labeled pre-crRNA were first imaged on Odyssey CLx (LI-COR Biosciences) prior to SYBR staining.

Target Cleavage Assays with Cas12g

*ssRNA, ssDNA and dsDNA substrates***:** Cas12g target cleavage assays were performed for 1 hour at the specified temperature in cleavage buffer. Ternary complex was formed by incubating a 1:2:0.5 molar ratio of Cas12g:mature crRNA:tracrRNA for 10 minutes at 37˚C followed by transfer to ice. Since mature crRNA generated from processing of pre-crRNA is expected to contain a mix of 5' hydroxyl and monophosphate end groups, mature crRNA for the assay was prepared by mixing equimolar amounts of hydroxyl end group and monophosphate end group mature crRNAs. Ternary complex was further diluted on ice keeping the molar ratio of protein to RNA fixed. The complex was then added to IR800-labeled substrates and incubated for 1 hour at the specified temperature. The final concentration of all labeled substrates was 100nM. For collateral

activity assays, the final concentration of both labeled and unlabeled substrates was 100nM each unless otherwise specified. Reactions were treated with RNase cocktail (Thermo Fisher Scientific) or Turbo DNase (Thermo Fisher Scientific) to visualize DNA and RNA substrates respectively. Reactions were quenched by adding 1ug/uL of Proteinase K (Ambion) and incubating for 15 minutes at 37˚C, followed by adding 50mM EDTA. Samples were then mixed with equal volume of 2X TBE-Urea sample buffer and denatured at 65˚C for 3 minutes for RNA samples and 95˚C for 3 minutes for DNA samples. Samples were analyzed on denaturing gel electrophoresis on 15% TBE-Urea gel (Invitrogen). Gels were first imaged on Odyssey CLx (LI-COR Biosciences) followed by a 5-minute stain with SYBR Gold nucleic acid stain (Invitrogen), then imaged on Gel Doc EZ (Bio-Rad).

*RNA:DNA hybrid substrates***:** Cleavage assays for RNA:DNA hybrid substrates were performed as described above using unlabeled RNA and 5'IR800-labeled DNA. All RNA:DNA hybrid substrates were mixed at a 1:2 ratio of RNA:DNA and annealed in hybridization buffer (50mM Tris pH 8.0 and 100mM NaCl) by incubating at 70˚C for 2 minutes followed by a 5˚C/min ramp down to 4˚C. Final concentration of the labeled DNA was 100nM. Reactions were treated with RNase cocktail prior to the Proteinase K addition.

Target Cleavage Assays with Cas12i

*ssDNA*: Cas12i target cleavage assays with ssDNA were performed in optimized cleavage buffer (50mM Tris-HCl pH 8.0, 50mM NaCl, 1mM DTT, 10mM MgCl2,

50ug/ml BSA). Binary complex was formed by incubating a 1:2 molar ratio of Cas12i:pre-crRNA for 10 minutes at 37˚C, followed by transfer to ice. All further complex dilutions were done on ice keeping the protein:RNA ratio fixed. The complex was added to 100nM IR800-labeled substrates and incubated at 37˚C for 30 minutes. Reactions were treated with RNase cocktail and Proteinase K and analyzed as above.

*dsDNA***:** dsDNA target cleavage assays were set up in the optimized cleavage buffer at 37˚C for 1 hour. Binary complex was formed as described above and added to 100nM dsDNA substrate. Reactions were first treated with RNase cocktail with incubation at 37˚C for 15 minutes. Next, they were treated with Proteinase K with incubation at 37˚C for 15 minutes. To detect dsDNA cleavage products the reactions were analyzed with 15% TBE-Urea gel as described above. To detect nicking activity of Cas12i, reactions were SPRI purified after Proteinase K treatment and split into three fractions. One fraction was analyzed on a 15% TBE-Urea gel as described above. Another fraction was mixed with Hi-Density TBE Sample Buffer (5X) (Invitrogen) and analyzed on a non-denaturing 4-20% TBE gel to detect nicked dsDNA products. The last fraction was incubated with 0.01U/uL of S1 Nuclease (Thermo Fisher Scientific) at 50˚C for 1 hour to convert nicks into double-stranded breaks followed by mixing with Hi-Density TBE Sample Buffer (5X) and analyzed on a non-denaturing 4-20% TBE gel. All gels were imaged on the Odyssey CLx, SYBR stained for 5 minutes, then imaged on the Gel Doc EZ.

To identify the nicked strand, dsDNA was prepared by labeling either the crRNA spacer-complementary or non-complementary strand (same sequence as the crRNA). The

cleavage reaction was performed as described above. The labeled strands were then annealed with the corresponding primers and extended with DNA Polymerase I, Large (Klenow) Fragment (New England BioLabs) for 15 minutes at 25˚C. The dsDNA substrates were then purified via SPRI.

Thermal Melt Assays for Cas12g

Melting curves of effector proteins were determined through differential scanning fluorimetry (DSF). All reactions took place at effector concentration of 0.62uM in DSF buffer (50mM Tris-HCl pH 7.9, 100mM NaCl, 1mM DTT) with the addition of 5X SYPRO Orange Protein Gel Stain (Sigma). Binary (effector + m-crRNA) and ternary (effector + m-crRNA + tracrRNA) complexes were formed at a 2.5:1 ratio of RNA to protein. Apo and effector complexes were incubated at 37ºC for 10 minutes prior to melting to assure complex formation. The reactions were then split into four 5uL technical replicates. Protein melting was performed in the Lightcycler 480 (Roche) over a temperature range from 20ºC to 80ºC, at a rate of 0.06ºC/second, with 10 acquisitions/second. The first derivative of the raw fluorescence data was taken in order to determine the $T_m$ of the effector.

Data have been deposited in the following resources

Next-Generation Sequencing for bacterial DNA-sequencing, RNA-sequencing of *E. coli* primary screens, and RNA-sequencing of *in vivo* pre-crRNA processing will be made available on the NCBI Sequence Read Archive under the Bioproject ID, PRJNA496291.

**Figure S1**

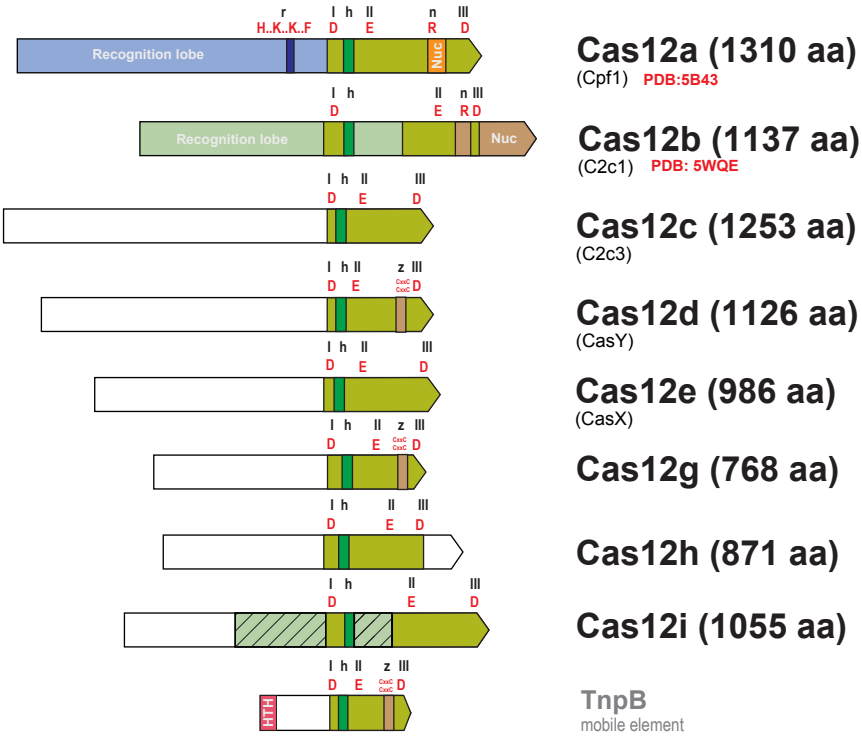

r
H..K..K..F
I h II
D E
n III
R D
Recognition lobe
Nuc
Cas12a (1310 aa)
(Cpf1) PDB:5B43
Recognition lobe
Nuc
Cas12b (1137 aa)
(C2c1) PDB: 5WQE
Cas12c (1253 aa)
(C2c3)
z
CxxC
CxxC
Cas12d (1126 aa)
(CasY)
Cas12e (986 aa)
(CasX)
Cas12g (768 aa)
Cas12h (871 aa)
Cas12i (1055 aa)
HTH
TnpB
mobile element

**Fig. S1. Domain architectures of selected Cas12 proteins and TnpB**

The proteins are shown approximately to scale. For the C-terminal RuvC domain, the catalytic residues in the three conserved sequence motifs (I, II and III) are indicated and shown to scale. Proteins aligned by location of RuvC-I motif. Bridge helix (h) and Zn-finger domain (z, CxxC..CxxC) are shown only approximately to scale for the proteins for which structure is not available. Experimentally identified residues involved in RNA cleavage (r) and additional DNA nuclease (n) are shown for Cas12a and Cas12b. The extended regions of Cas12b and Cas12i with limited sequence similarity beyond RuvC-domain are colored light green and hashed light green respectively. Amino acid sizes are of representative proteins from each subtype.

# Figure S2

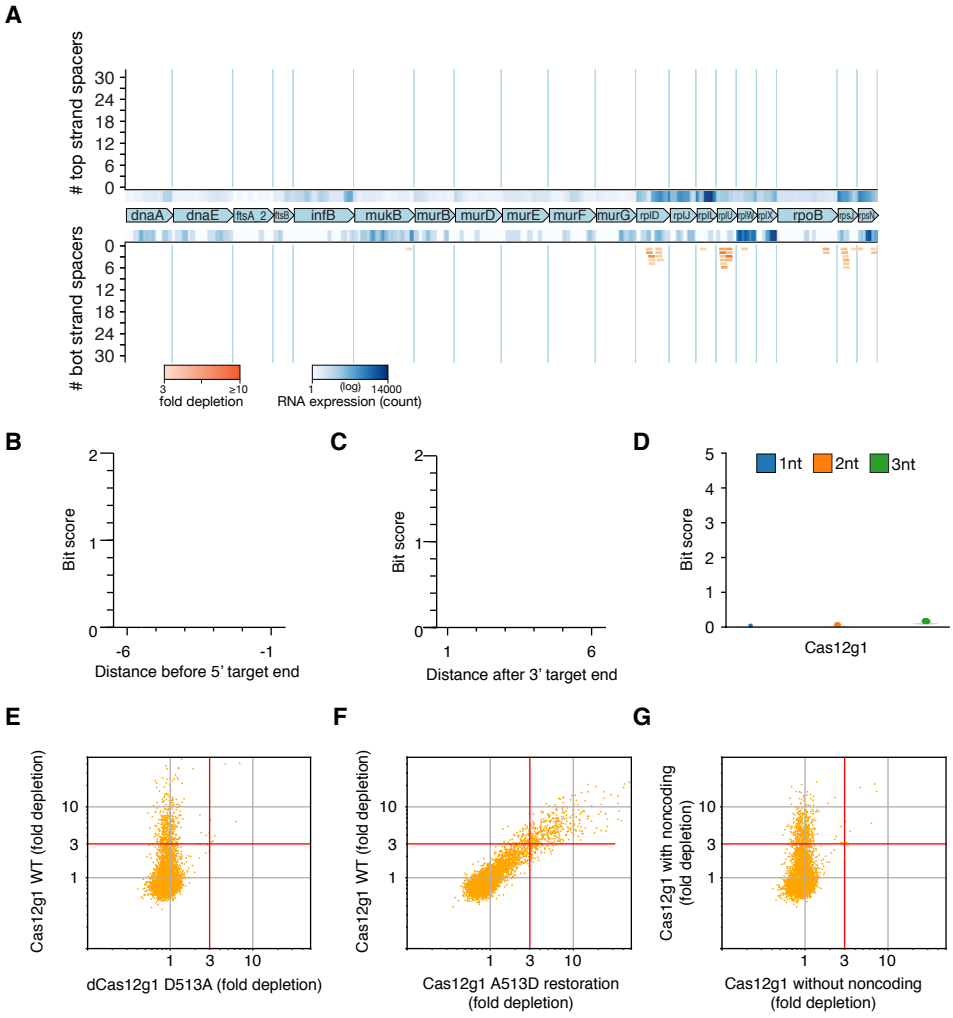

A
# top strand spacers
30
24
18
12
6
0
dnaA
dnaE
ftsA
2
ftsB
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplK
rplA
rplX
rpoB
rpsL
rpsG
# bot strand spacers
fold depletion
3
≥10
RNA expression (count)
1
(log)
14000
B
Bit score
Distance before 5' target end
-6
-1
C
Distance after 3' target end
1
6
D
1nt
2nt
3nt
Cas12g1
E
Cas12g1 WT (fold depletion)
dCas12g1 D513A (fold depletion)
F
Cas12g1 A513D restoration
(fold depletion)
G
Cas12g1 with noncoding
(fold depletion)
Cas12g1 without noncoding
(fold depletion)

**Fig. S2. Dependencies of Cas12g *in vivo* interference activity**

(A) Spatial distribution of spacers from strongly depleted CRISPR arrays from Cas12g1 *in vivo* screening over *E. coli* essential gene targets separated by target strand (top: bottom strand-targeting; bottom: top strand-targeting; color saturation: spacer fold depletion. Orange and green colors on the bars indicate the orientation of the active DR of the CRISPR arrays, with green representing the forward DR orientation and orange the reverse. (B to D) Sequence logos generated from the strongly depleted targets on the (B) left flank and (C) right flank show no sequence dependence of targeting, confirmed by calculating (D) bit scores for all possible permutations of target and flanking nucleotides up to length 3. (E to F) Comparison of *in vivo* screening activity of (E) Cas12g1 WT versus dCas12g1(D513A) RuvC mutant, (F) Cas12g1 WT versus the A513D rescue mutant cloned from dCas12g1(D513A), and (G) Cas12g1 WT with the noncoding plasmid versus without. In (E) and (F), all of the screening samples contained the noncoding plasmid. Red lines indicate threshold for "strongly depleted" (>3 fold depletion).

**Figure S3**

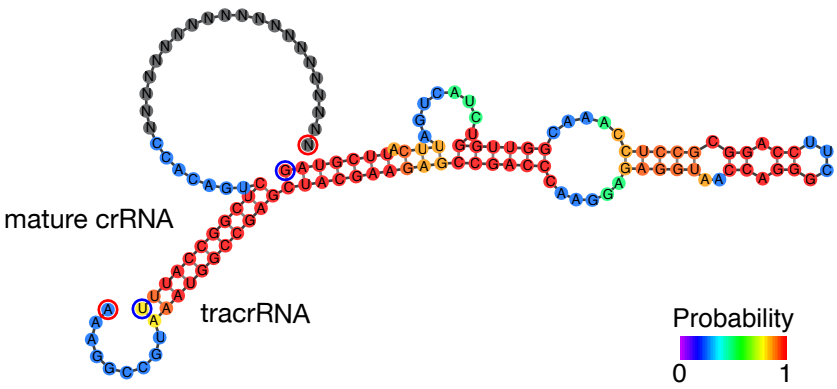


**Fig. S3. Co-fold of the Cas12g1 mature crRNA and tracrRNA**

Co-fold of the mature crRNA and tracrRNA with sequences determined by small RNAseq of *E. coli* expressing the Cas12g1 system from the *in vivo* screen. Colors indicate probability of folding, with gray corresponding to variable spacer sequences. Blue and red circles outline the 5' and 3' ends of the RNA, respectively.

**Figure S4**

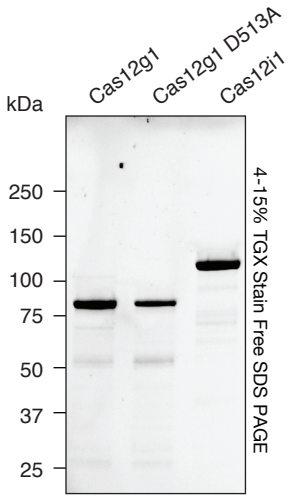

Cas12g1
Cas12g1 D513A
Cas12i1
kDa
250
150
100
75
50
37
25
4-15% TGX Stain Free SDS PAGE

**Fig. S4. Purification of Cas12g1 protein and D513A mutant, and Cas12i1**

Stain Free 4-15% SDS-Page (Bio-rad) gel of effector proteins post His-tag purification, fraction pooling, and concentration.

# Figure S5

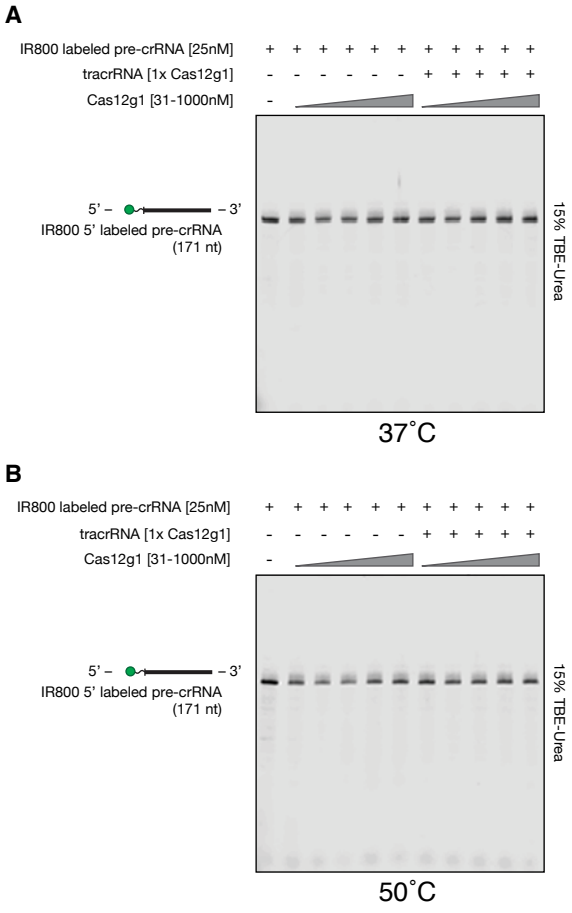

A
IR800 labeled pre-crRNA [25nM]
tracrRNA [1x Cas12g1]
Cas12g1 [31-1000nM]
5' –
– 3'
IR800 5' labeled pre-crRNA
(171 nt)
15% TBE-Urea
37˚C
B
IR800 labeled pre-crRNA [25nM]
tracrRNA [1x Cas12g1]
Cas12g1 [31-1000nM]
5' –
– 3'
IR800 5' labeled pre-crRNA
(171 nt)
15% TBE-Urea
50˚C

**Fig. S5. *In vitro* crRNA processing of Cas12g1**

(A and B) Denaturing gel showing lack of *in vitro* pre-crRNA processing by Cas12g1 effector at (A) 37ºC and (B) 50ºC. Pre-crRNA with and without tracrRNA was incubated with Cas12g1 protein for 1 hour at 37ºC and analyzed on a 15% TBE-Urea gel.

# Figure S6

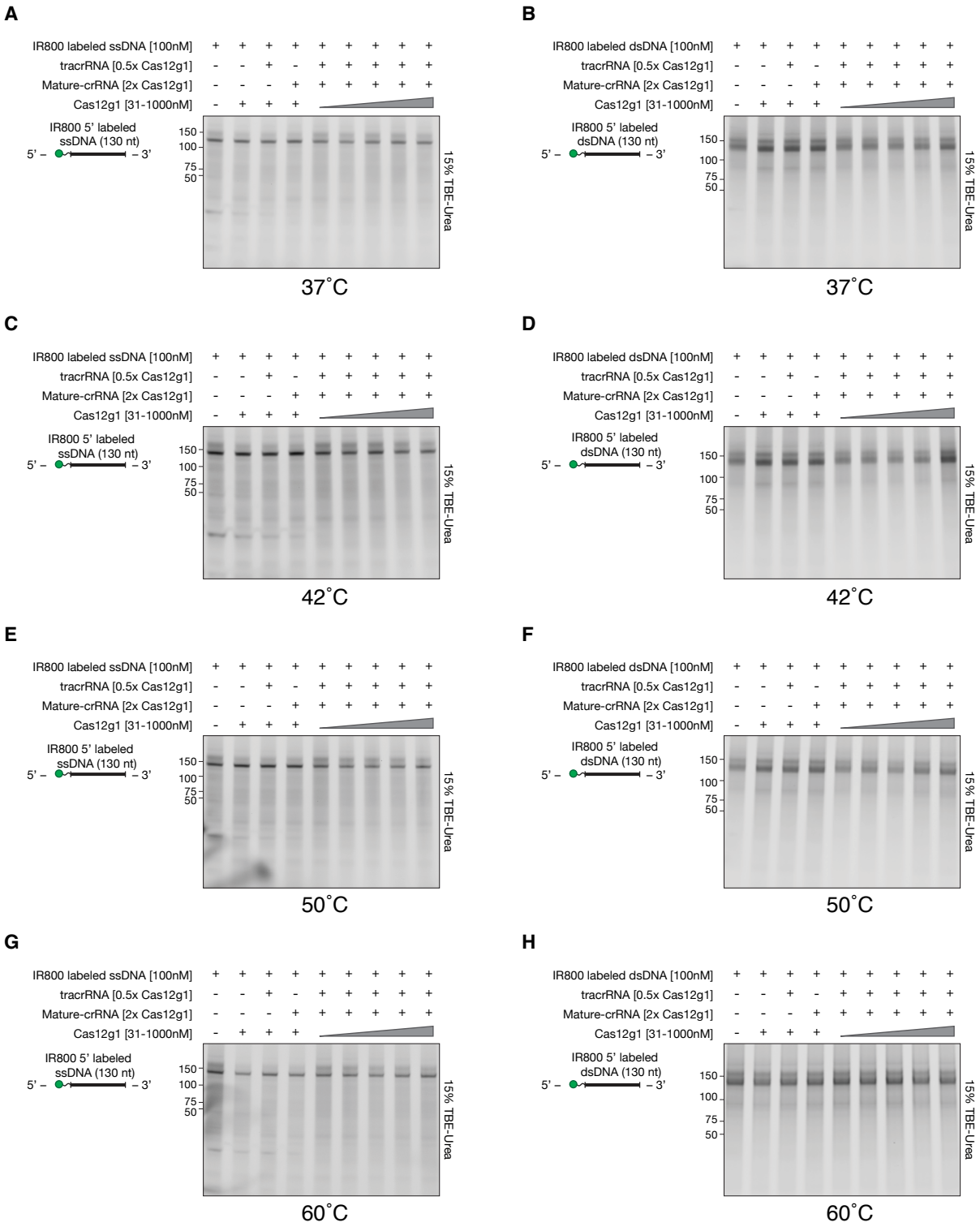

A
IR800 labeled ssDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled ssDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
37˚C
B
IR800 labeled dsDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled dsDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
37˚C
C
IR800 labeled ssDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled ssDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
42˚C
D
IR800 labeled dsDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled dsDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
42˚C
E
IR800 labeled ssDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled ssDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
50˚C
F
IR800 labeled dsDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled dsDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
50˚C
G
IR800 labeled ssDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled ssDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
60˚C
H
IR800 labeled dsDNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-1000nM]
IR800 5' labeled dsDNA (130 nt)
5' – 3'
150
100
75
50
15% TBE-Urea
60˚C

**Fig. S6. No detectable nuclease activity of Cas12g1 on ssDNA and dsDNA targets**

(A to H) Denaturing gels demonstrating the absence of activity for Cas12g1 effector protein ternary complex against ssDNA and dsDNA targets. Cas12g1 was tested for cleavage activity against ssDNA and dsDNA substrates at 37ºC (A, B), 42ºC (C, D), 50ºC (E, F) and 60ºC (G, H). ssDNA target is 5' end labeled with IR800 dye, and dsDNA target is labeled with IR800 dye on the 5' end of the non-spacer-complementary strand. For all reactions, the Cas12g1 ternary complex was formed by pre-incubating Cas12g1 protein, mature crRNA and tracrRNA for 10 minutes at 37ºC prior to adding the substrates and incubating for 1 hour at the specified temperature.

# Figure S7

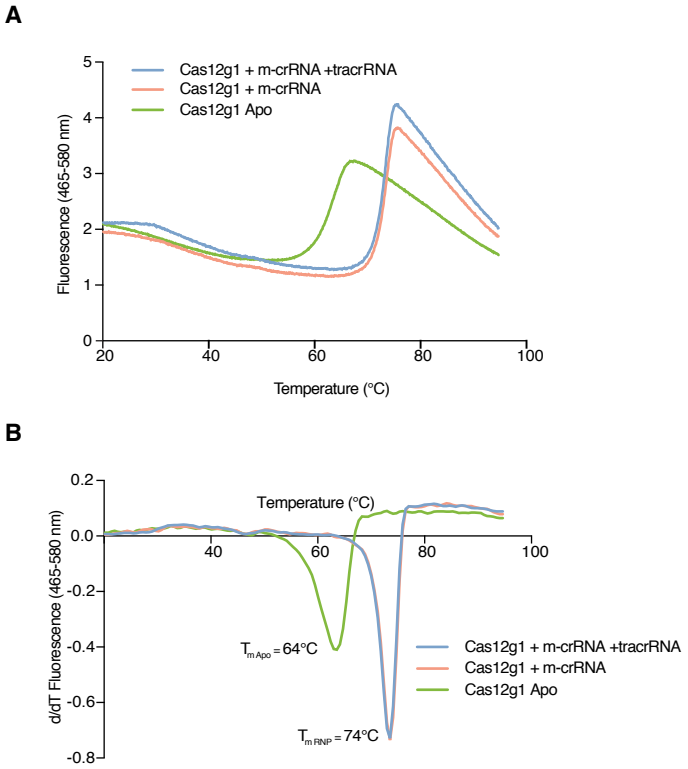

A
Cas12g1 + m-crRNA +tracrRNA
Cas12g1 + m-crRNA
Cas12g1 Apo
Fluorescence (465-580 nm)
Temperature (°C)
B
d/dT Fluorescence (465-580 nm)
Temperature (°C)
T m Apo = 64°C
T m RNP = 74°C
Cas12g1 + m-crRNA +tracrRNA
Cas12g1 + m-crRNA
Cas12g1 Apo

**Fig. S7. Cas12g1 apo and binary complexes are thermostable**

(A) Melting curve of Cas12g1 displaying the thermostability of apo Cas12g1 and increased thermostability of binary (Cas12g1 + crRNA) and ternary (Cas12g1 + crRNA + tracrRNA) complex of Cas12g1. Melting curve was obtained using differential scanning fluorimetry (DSF). (B) First derivative plots of the thermal melts from (A) showing the melting temperatures of apo, binary, and ternary complexes of Cas12g1. The first derivative traces of the Cas12g1 binary and ternary complexes are nearly overlapping.

# Figure S8

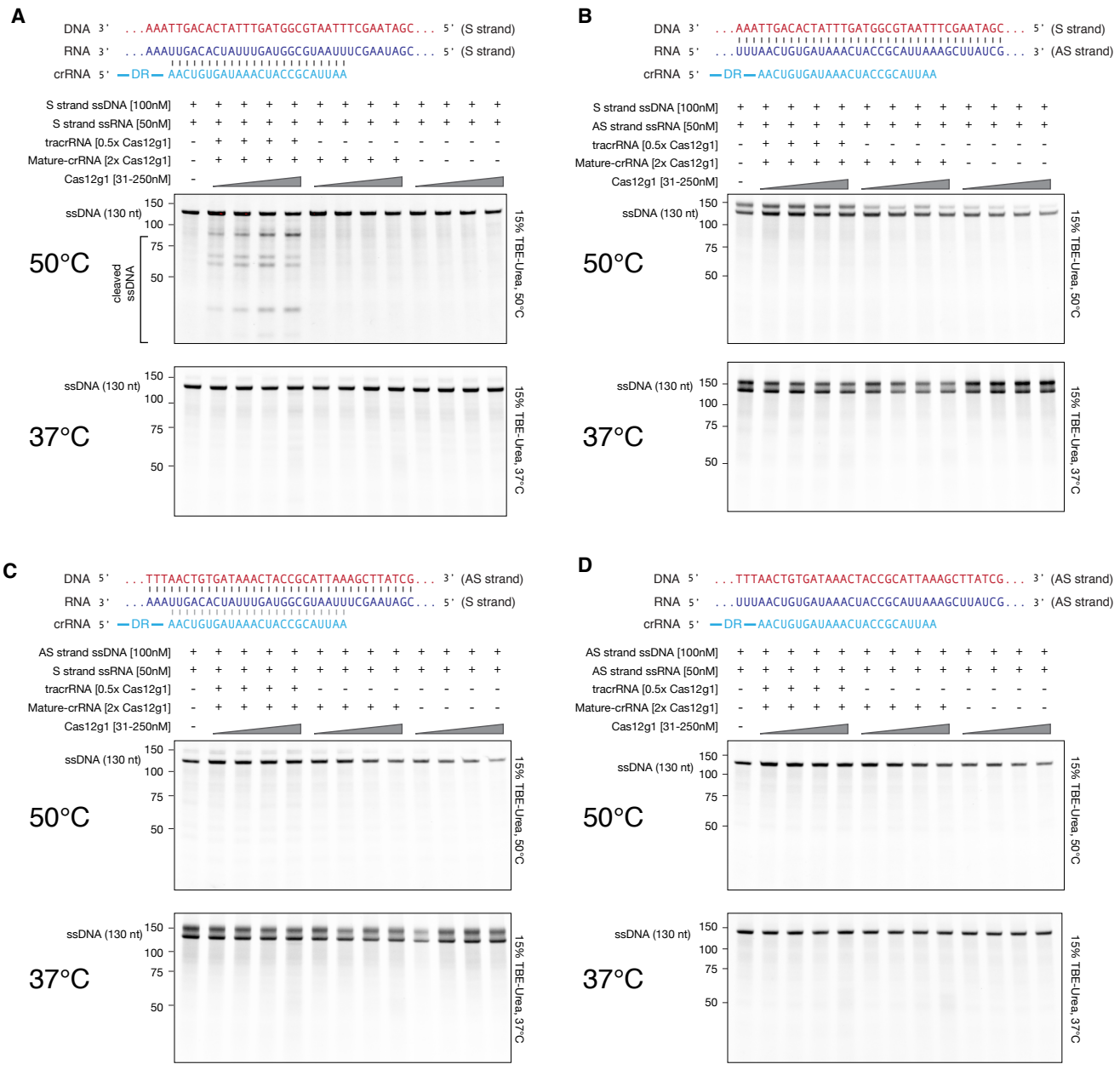

A
DNA 3' ...AAATTGACACTATTTGATGGCGTAATTTCGAATAGC... 5' (S strand)
RNA 3' ...AAAUUGACACUAUUUGAUGGCGUAAUUUCGAAUAGC... 5' (S strand)
crRNA 5' —DR— AACUGUGAUAAACUACCGCAUUAA
S strand ssDNA [100nM]
S strand ssRNA [50nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
ssDNA (130 nt)
cleaved ssDNA
50°C
37°C
15% TBE-Urea, 50°C
15% TBE-Urea, 37°C
B
DNA 3' ...AAATTGACACTATTTGATGGCGTAATTTCGAATAGC... 5' (S strand)
RNA 5' ...UUUAACUGUGAUAAACUACCGCAUUAAAGCUUAUCG... 3' (AS strand)
crRNA 5' —DR— AACUGUGAUAAACUACCGCAUUAA
AS strand ssDNA [100nM]
AS strand ssRNA [50nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
ssDNA (130 nt)
50°C
37°C
15% TBE-Urea, 50°C
15% TBE-Urea, 37°C
C
DNA 5' ...TTTAACTGTGATAAACTACCGCATTAAAGCTTATCG... 3' (AS strand)
RNA 3' ...AAAUUGACACUAUUUGAUGGCGUAAUUUCGAAUAGC... 5' (S strand)
crRNA 5' —DR— AACUGUGAUAAACUACCGCAUUAA
AS strand ssDNA [100nM]
S strand ssRNA [50nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
ssDNA (130 nt)
50°C
37°C
15% TBE-Urea, 50°C
15% TBE-Urea, 37°C
D
DNA 5' ...TTTAACTGTGATAAACTACCGCATTAAAGCTTATCG... 3' (AS strand)
RNA 5' ...UUUAACUGUGAUAAACUACCGCAUUAAAGCUUAUCG... 3' (AS strand)
crRNA 5' —DR— AACUGUGAUAAACUACCGCAUUAA
AS strand ssDNA [100nM]
AS strand ssRNA [50nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
ssDNA (130 nt)
50°C
37°C
15% TBE-Urea, 50°C
15% TBE-Urea, 37°C

**Fig. S8. Nuclease activity of Cas12g1 against ssDNA in presence of RNA transcripts at 50ºC and 37ºC**

(A to D) Ternary complex of Cas12g1 effector protein was tested on (A) sense ssDNA (including crRNA spacer-complementary target) and sense ssRNA, (B) sense ssDNA and antisense ssRNA, (C) antisense ssDNA and sense ssRNA, and (D) antisense ssDNA and antisense ssRNA at 50ºC and 37ºC. (S and AS denote sense and antisense, respectively.) For all reactions, ssDNA and ssRNA substrates were hybridized prior to adding the effector ternary complex, yielding a DNA-RNA hybrid for the sense-antisense pairings of ssDNA and ssRNA. Cas12g1 ternary complex was formed by pre-incubating Cas12g1 protein and mature crRNA and tracrRNA for 10 minutes at 37ºC prior to adding the substrates and incubating for 1 hour at the specified temperature. All reactions were RNase treated and visualized using SYBR Gold nucleic acid staining.

# Figure S9

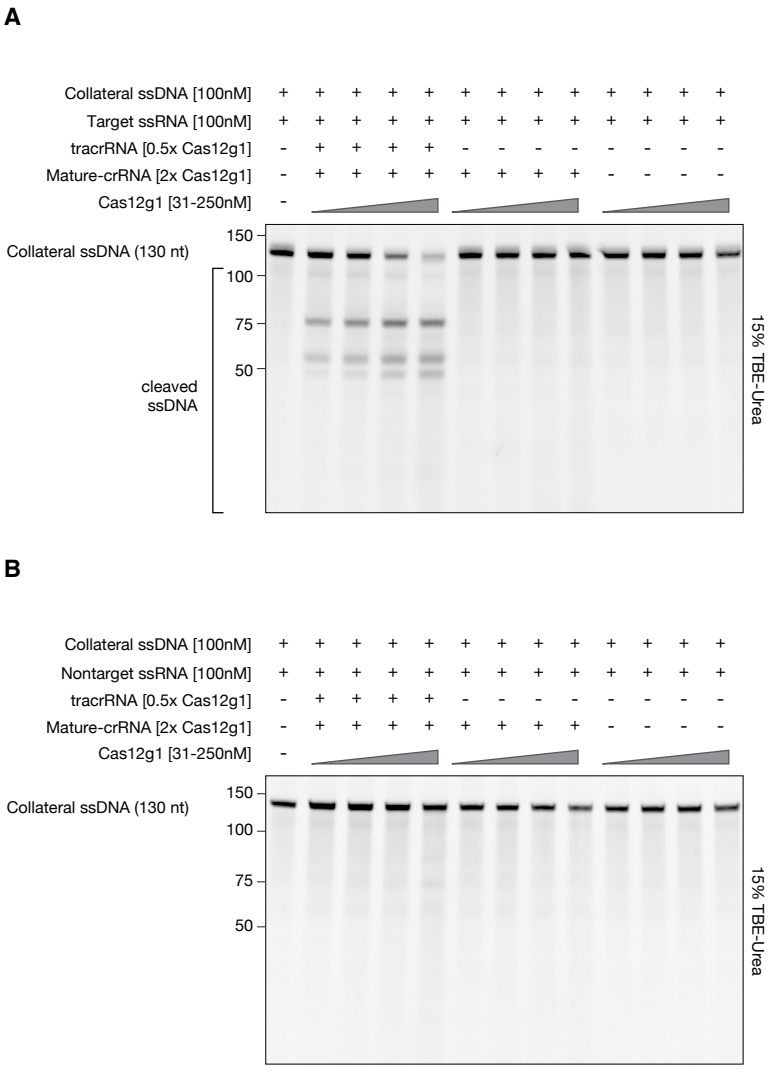

A
Collateral ssDNA [100nM]
Target ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
Collateral ssDNA (130 nt)
cleaved ssDNA
150
100
75
50
15% TBE-Urea
B
Collateral ssDNA [100nM]
Nontarget ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
Collateral ssDNA (130 nt)
150
100
75
50
15% TBE-Urea

**Fig. S9. Collateral activity of Cas12g1 on ssDNA**

Denaturing gels displaying collateral cleavage activity of Cas12g1 effector protein ternary complex against a nontarget ssDNA substrate upon recognition of the target ssRNA. (A and B) Collateral ssDNA substrate is cleaved at 50ºC by Cas12g1 ternary complex only in the presence of (A) target ssRNA and not (B) in the presence of nontarget ssRNA. Cas12g1 ternary complex was formed by pre-incubating Cas12g1 protein and mature crRNA and tracrRNA for 10 minutes at 37ºC prior to adding the substrates and incubating for 1 hour at 37ºC. All reactions were RNase treated and visualized using SYBR Gold nucleic acid staining.

# Figure S10

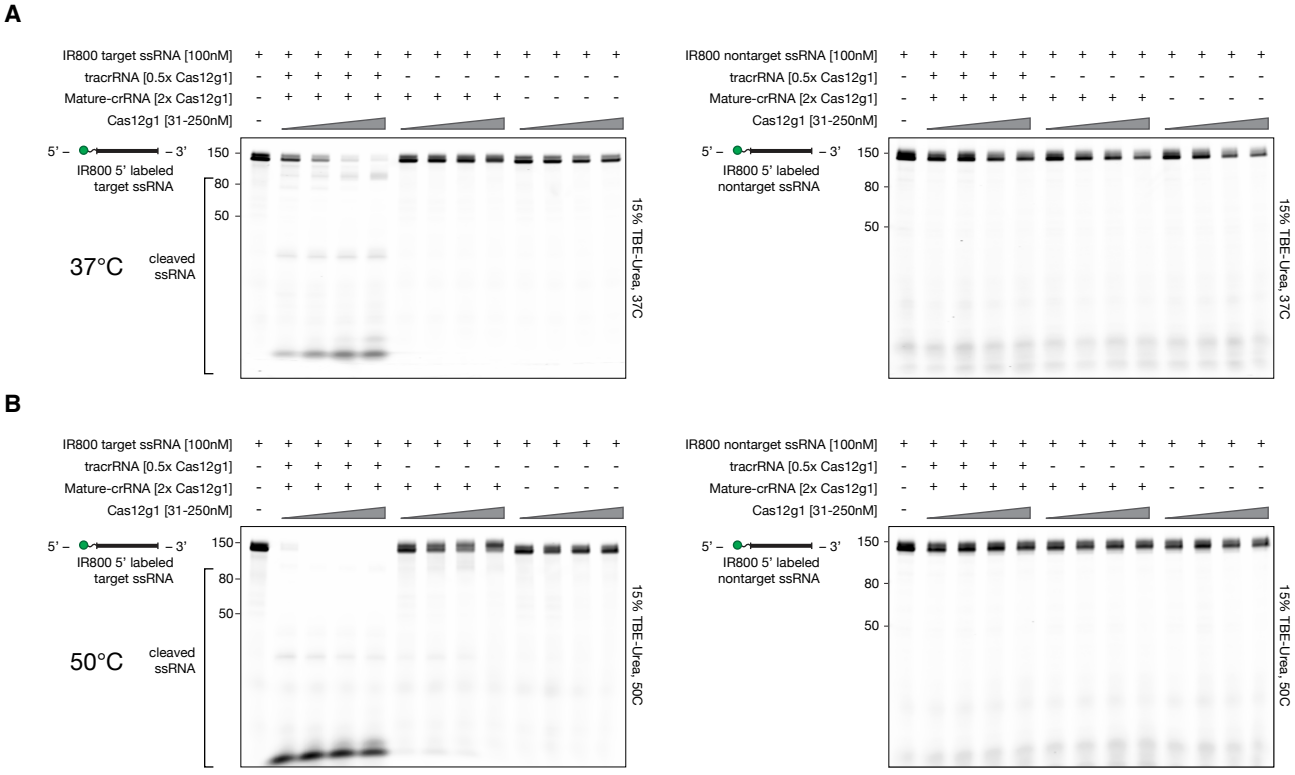

A
IR800 target ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
5' – – 3'
IR800 5' labeled target ssRNA
37°C
cleaved ssRNA
150
80
50
15% TBE-Urea, 37C
IR800 nontarget ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
5' – – 3'
IR800 5' labeled nontarget ssRNA
150
80
50
15% TBE-Urea, 37C
B
IR800 target ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
5' – – 3'
IR800 5' labeled target ssRNA
50°C
cleaved ssRNA
150
80
50
15% TBE-Urea, 50C
IR800 nontarget ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
5' – – 3'
IR800 5' labeled nontarget ssRNA
150
80
50
15% TBE-Urea, 50C

**Fig. S10. Targeted ssRNA cleavage by Cas12g1**

Denaturing gels displaying targeted ssRNA cleavage activity of Cas12g1 effector protein ternary complex. (A and B) Cleavage reactions performed at (A) 37ºC and (B) 50ºC show the successful cleavage of IR800 dye-labeled target ssRNA substrate only in the presence of a matching crRNA. Cas12g1 ternary complex was formed by pre-incubating Cas12g1 protein with mature crRNA and tracrRNA for 10 minutes at 37ºC prior to adding the substrates and incubating for 1 hour at the indicated temperature.

# Figure S11

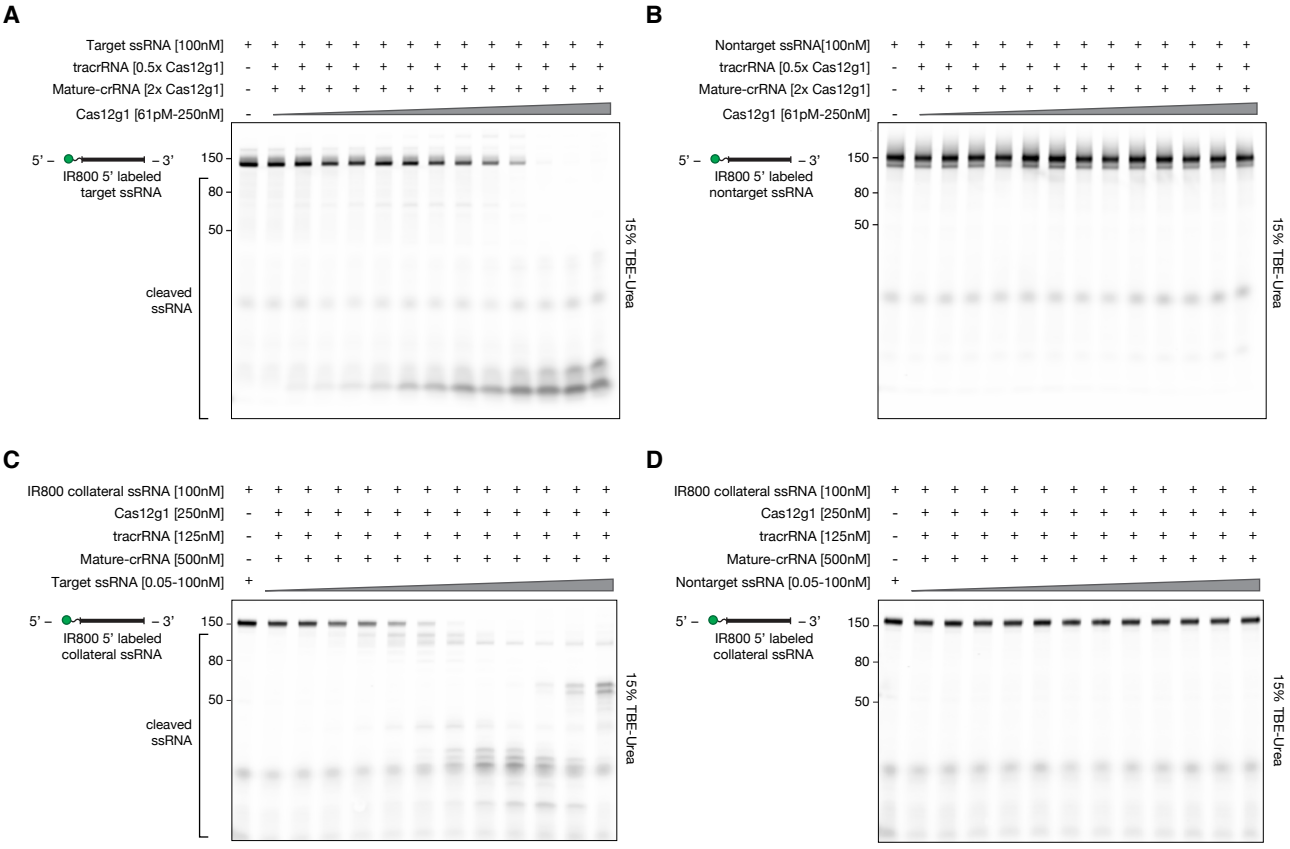

A
Target ssRNA [100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [61pM-250nM]
5' – – 3'
IR800 5' labeled target ssRNA
150
80
50
cleaved ssRNA
15% TBE-Urea
B
Nontarget ssRNA[100nM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [61pM-250nM]
5' – – 3'
IR800 5' labeled nontarget ssRNA
150
80
50
15% TBE-Urea
C
IR800 collateral ssRNA [100nM]
Cas12g1 [250nM]
tracrRNA [125nM]
Mature-crRNA [500nM]
Target ssRNA [0.05-100nM]
5' – – 3'
IR800 5' labeled collateral ssRNA
150
80
50
cleaved ssRNA
15% TBE-Urea
D
IR800 collateral ssRNA [100nM]
Cas12g1 [250nM]
tracrRNA [125nM]
Mature-crRNA [500nM]
Nontarget ssRNA [0.05-100nM]
5' – – 3'
IR800 5' labeled collateral ssRNA
150
80
50
15% TBE-Urea

**Fig. S11. Cas12g1 is a robust and sensitive RNA-guided RNase**

(A) Target ssRNA cleavage of Cas12g over a log-2 dose titration of Cas12g ternary complex concentrations assayed at 50ºC. Cas12g1 ternary complex showed complete cleavage of the target ssRNA down to nucleotides at concentrations above 32nM, and detectable RNase activity on target ssRNA at complex concentrations as low as 125pM. (B) Cas12g1 ternary complex displays no RNase activity with a nontarget ssRNA over the same dose range with complex concentrations as high as 250nM. (C) Cas12g1 ternary complex cleaves collateral ssRNA at 50˚C over a range of target ssRNA concentrations as low as 100pM. (D) Over an equivalent concentration range of nontarget ssRNA, Cas12g1 ternary complex does not cleave collateral ssRNA. Labeled ssRNA species were tagged at the 5' end with IR800 dye. Cas12g1 ternary complex was formed by pre-incubating Cas12g1 protein with mature crRNA and tracrRNA for 10 minutes at 37ºC prior to adding the substrates and incubating for 1 hour at 50ºC.

# Figure S12

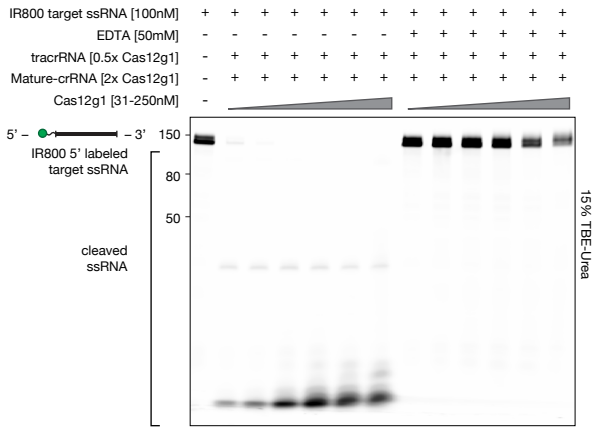

IR800 target ssRNA [100nM]
EDTA [50mM]
tracrRNA [0.5x Cas12g1]
Mature-crRNA [2x Cas12g1]
Cas12g1 [31-250nM]
5' –
– 3'
IR800 5' labeled target ssRNA
cleaved ssRNA
150
80
50
15% TBE-Urea

**Fig. S12. Nuclease activity of Cas12g1 on target ssRNA substrates is inhibited by EDTA**

Comparison of target ssRNA cleavage by the Cas12g1 ternary complex in the presence or absence of EDTA at 50°C. EDTA was added at an excess to the magnesium concentration in the cleavage buffer.

# Figure S13

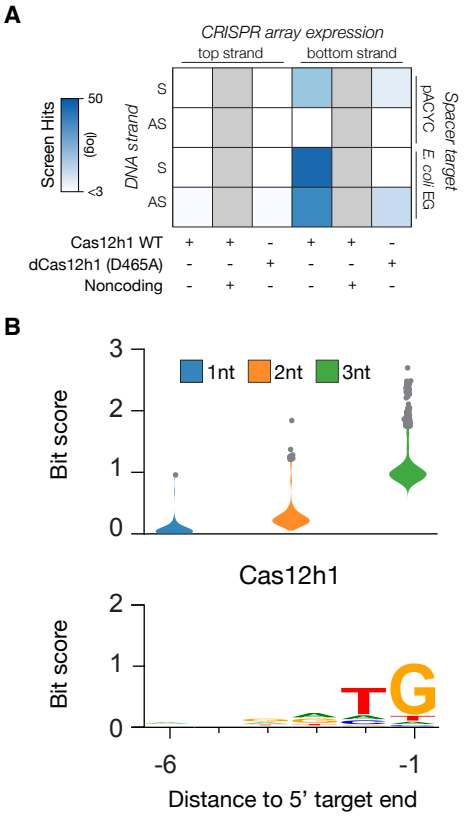

A
CRISPR array expression
top strand
bottom strand
Screen Hits
50
(log)
<3
DNA strand
S
AS
S
AS
pACYC
E. coli EG
Spacer target
Cas12h1 WT + + - + + -
dCas12h1 (D465A) - - + - - +
Noncoding - + - - + -
B
1nt
2nt
3nt
Bit score
3
2
1
0
Cas12h1
Bit score
2
1
0
-6
-1
Distance to 5' target end

**Fig. S13. Cas12h *in vivo* screening activity summary and sequence dependencies**

(A) Evaluation of the minimal active system for Cas12h1, with the heatmap summarizing the number of strongly depleted CRISPR arrays (screen hits) detected for Cas12h1 compositions as follows: 1) Cas12h1 with no noncoding (minimal active system), 2) Cas12h1 with noncoding, which was cytotoxic and unable to be screened in *E. coli* (grey boxes), 3) dCas12h1. Systems are additionally separated by CRISPR array expression direction, and strongly depleted targets for each system are separated by DNA strand (S, sense; AS, antisense; EG, essential genes). (B) (Top) Distribution of bit scores for all permutations of 1-3nt motifs within the target and 15nt flanking sequences corresponding to strongly depleted *in vivo* arrays, calculated as in *(19)*. (Bottom) Web logos from target flanking sequences.

# Figure S14

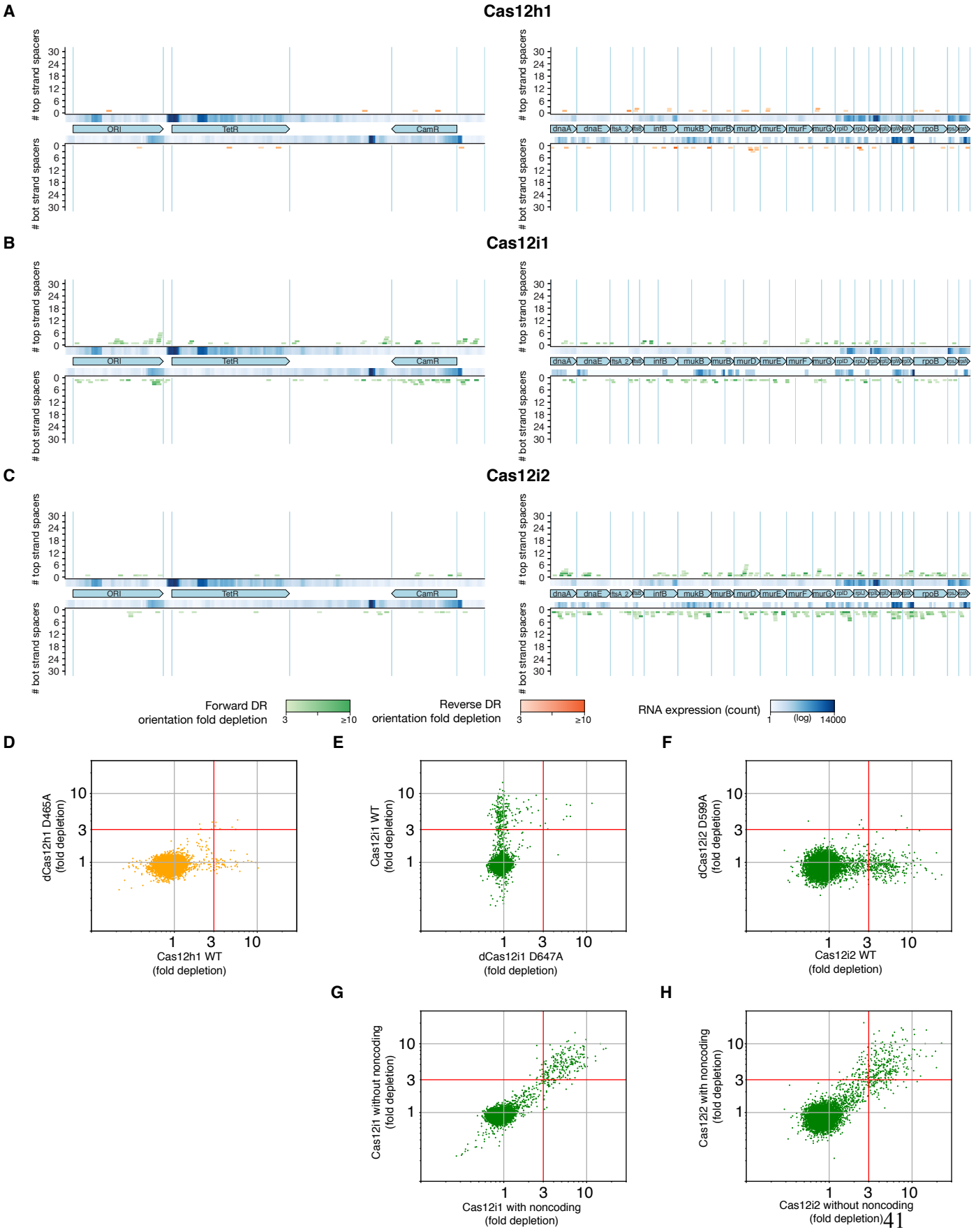

A
Cas12h1
B
Cas12i1
C
Cas12i2
# top strand spacers
# bot strand spacers
ORI
TetR
CamR
dnaA
dnaE
ftsA
ftsZ
infB
mukB
murB
murD
murE
murF
murG
rplD
rplJ
rplL
rplK
rplA
rpoB
Forward DR orientation fold depletion
Reverse DR orientation fold depletion
RNA expression (count)
D
dCas12h1 D465A (fold depletion)
Cas12h1 WT (fold depletion)
E
Cas12i1 WT (fold depletion)
dCas12i1 D647A (fold depletion)
F
dCas12i2 D599A (fold depletion)
Cas12i2 WT (fold depletion)
G
Cas12i1 without noncoding (fold depletion)
Cas12i1 with noncoding (fold depletion)
H
Cas12i2 with noncoding (fold depletion)
Cas12i2 without noncoding (fold depletion)

**Fig. S14. Dependencies of Cas12h and Cas12i *in vivo* interference activity**

(A to C) Spatial distribution of spacers from strongly depleted CRISPR arrays from *in vivo* screening of Cas12h and Cas12i and their noncoding plasmids mapped over pACYC184 plasmid and *E. coli* essential gene targets, separated by target strand (top: bottom strand-targeting; bottom: top strand-targeting; color saturation: spacer fold depletion.) Orange and green colors on the bars indicate the orientation of the active DR of the CRISPR arrays, with green representing the forward DR orientation and orange the reverse. (D to H) Comparison of *in vivo* screening activity of (D to F) WT effector versus the RuvC mutant, (G and H) WT effector with the noncoding plasmid versus without. As with the *in vivo* screen spatial distribution map, the coloring of the scatterplot represents the orientation of the active DR. Red lines indicate threshold for "strongly depleted" (>3 fold depletion).

# Figure S15

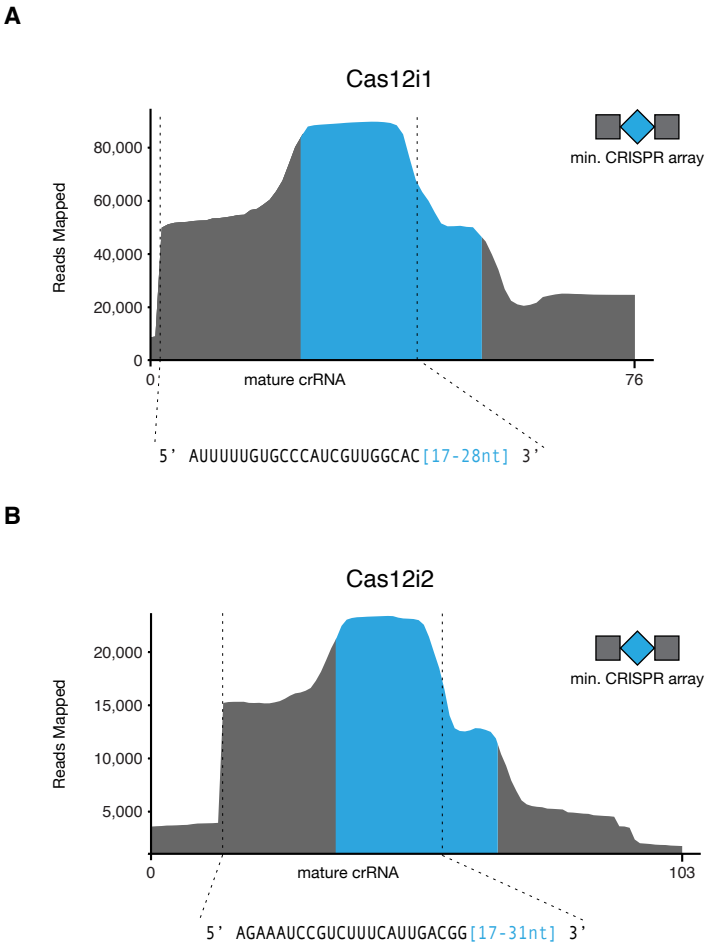

A
Cas12i1
Reads Mapped
80,000
60,000
40,000
20,000
0
min. CRISPR array
0
mature crRNA
76
5' AUUUUUGUGCCCAUCGUUGGCAC[17-28nt] 3'
B
Cas12i2
Reads Mapped
20,000
15,000
10,000
5,000
0
min. CRISPR array
mature crRNA
103
5' AGAAAUCCGUCUUUCAUUGACGG[17-31nt] 3'

**Fig. S15. Mature crRNA forms of Cas12i**

(A and B) Small RNA sequencing of *in vivo* screening samples of the minimal Cas12i systems revealed the mature crRNA of (A) Cas12i1 and (B) Cas12i2 systems.

**Figure S16**

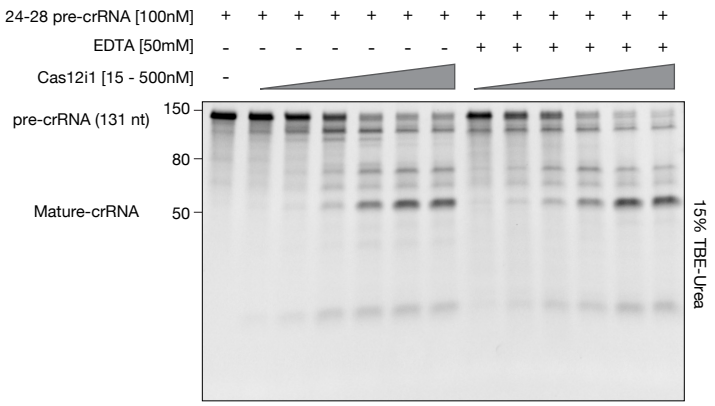

24-28 pre-crRNA [100nM]
EDTA [50mM]
Cas12i1 [15 - 500nM]
pre-crRNA (131 nt)
Mature-crRNA
150
80
50
15% TBE-Urea

**Fig. S16. Autonomous crRNA biogenesis by Cas12i1**

Denaturing gel showing pre-crRNA processing by Cas12i1 effector protein. EDTA-independent processing of pre-crRNA expressed from a minimal CRISPR array (repeat-spacer-repeat-spacer-repeat) with a 24nt repeat and 28nt spacer by Cas12i1 (24-28 pre-crRNA in figure label). Pre-crRNA was incubated with Cas12i1 protein for 30 minutes at 37ºC and analyzed on a 15% TBE-Urea gel.

# Figure S17

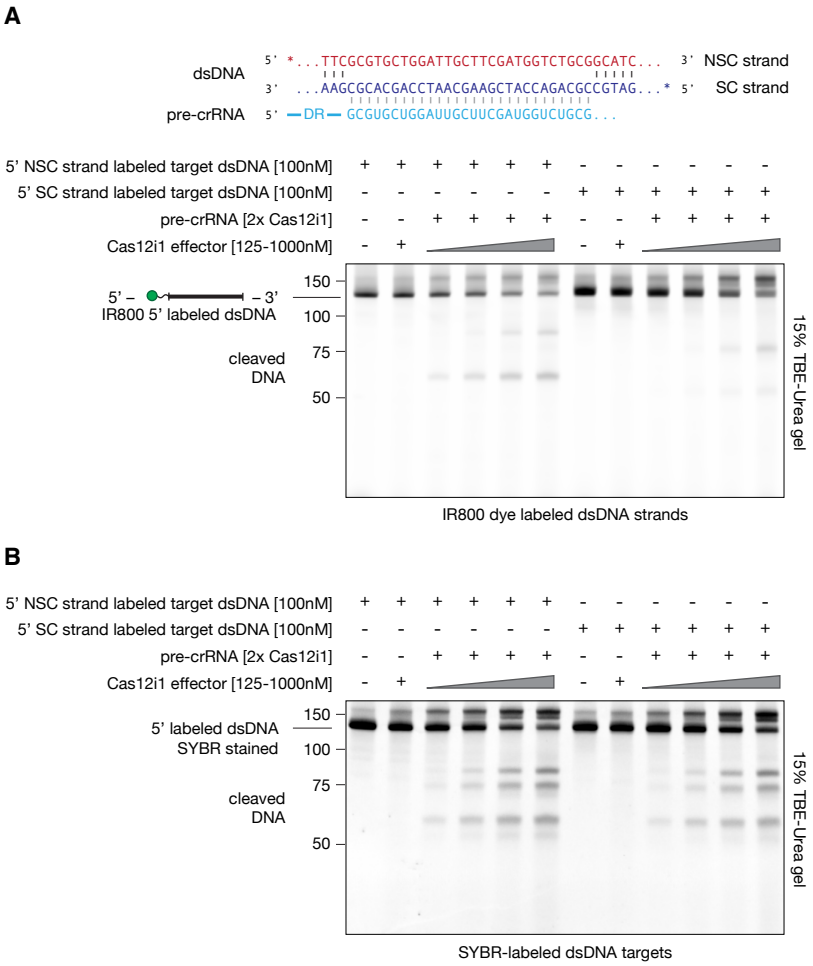
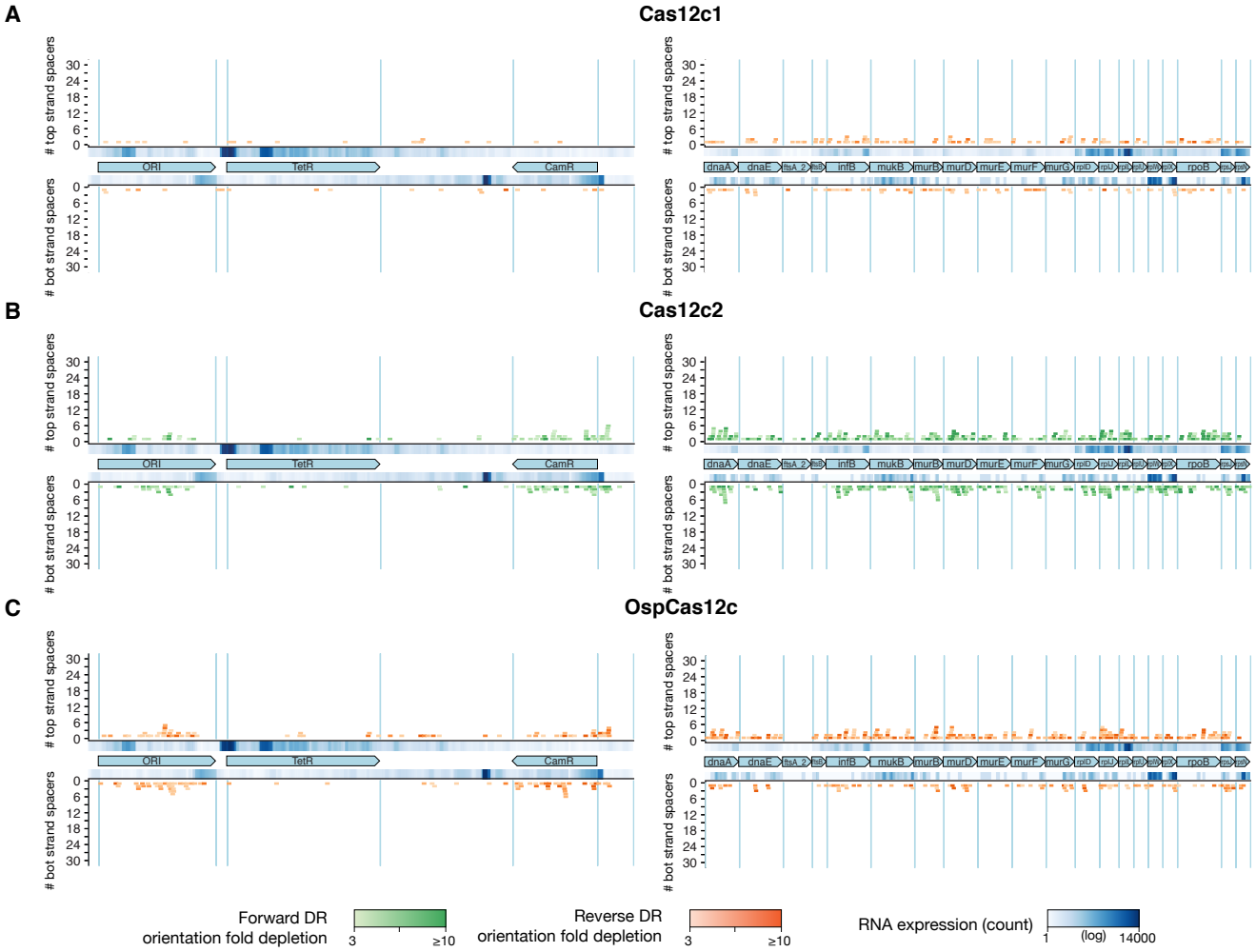

A
dsDNA
5' *...TTCGCGTGCTGGATTGCTTCGATGGTCTGCGGCATC... 3' NSC strand
3' ...AAGCGCACGACCTAACGAAGCTACCAGACGCCGTAG...* 5' SC strand
pre-crRNA
5' —DR— GCGUGCUGGAUUGCUUCGAUGGUCUGCG...
5' NSC strand labeled target dsDNA [100nM]
5' SC strand labeled target dsDNA [100nM]
pre-crRNA [2x Cas12i1]
Cas12i1 effector [125-1000nM]
5' – – 3'
IR800 5' labeled dsDNA
cleaved DNA
150
100
75
50
15% TBE-Urea gel
IR800 dye labeled dsDNA strands
B
5' NSC strand labeled target dsDNA [100nM]
5' SC strand labeled target dsDNA [100nM]
pre-crRNA [2x Cas12i1]
Cas12i1 effector [125-1000nM]
5' labeled dsDNA
SYBR stained
cleaved DNA
150
100
75
50
15% TBE-Urea gel
SYBR-labeled dsDNA targets

**Fig. S17. Cas12i1 displays reduced cleavage efficiency of the dsDNA spacer-complementary (SC) versus non-spacer-complementary (NSC) strand**

(A and B) Denaturing gels imaged by either (A) IR800 (labeled DNA only) or (B) SYBR Gold stain (total DNA) showing dsDNA cleavage or nicking of substrates containing a 5' IR800-labeled non-spacer-complementary strand (NSC, left) or 5' IR800-labeled spacer-complementary strand (SC, right) at increasing concentrations of Cas12i1 effector binary complex. The same gel was imaged in (A) and (B), with the SYBR Gold stain revealing approximately similar amounts of dsDNA cleavage products, but the IR800 image revealing a bias in strand cleavage inferred by the strand labeling. Cas12i1 complex was formed by pre-incubating Cas12i1 protein with pre-crRNA for 10 minutes at 37˚C prior to adding to the substrates and incubating for 1 hour at 37˚C.

# Figure S18


A
Cas12c1
B
Cas12c2
C
OspCas12c
# top strand spacers
# bot strand spacers
ORI
TetR
CamR
dnaA
dnaE
ftsA
infB
mukB
murB
murD
murE
murF
murG
rpoB
Forward DR orientation fold depletion
Reverse DR orientation fold depletion
RNA expression (count)
(log)

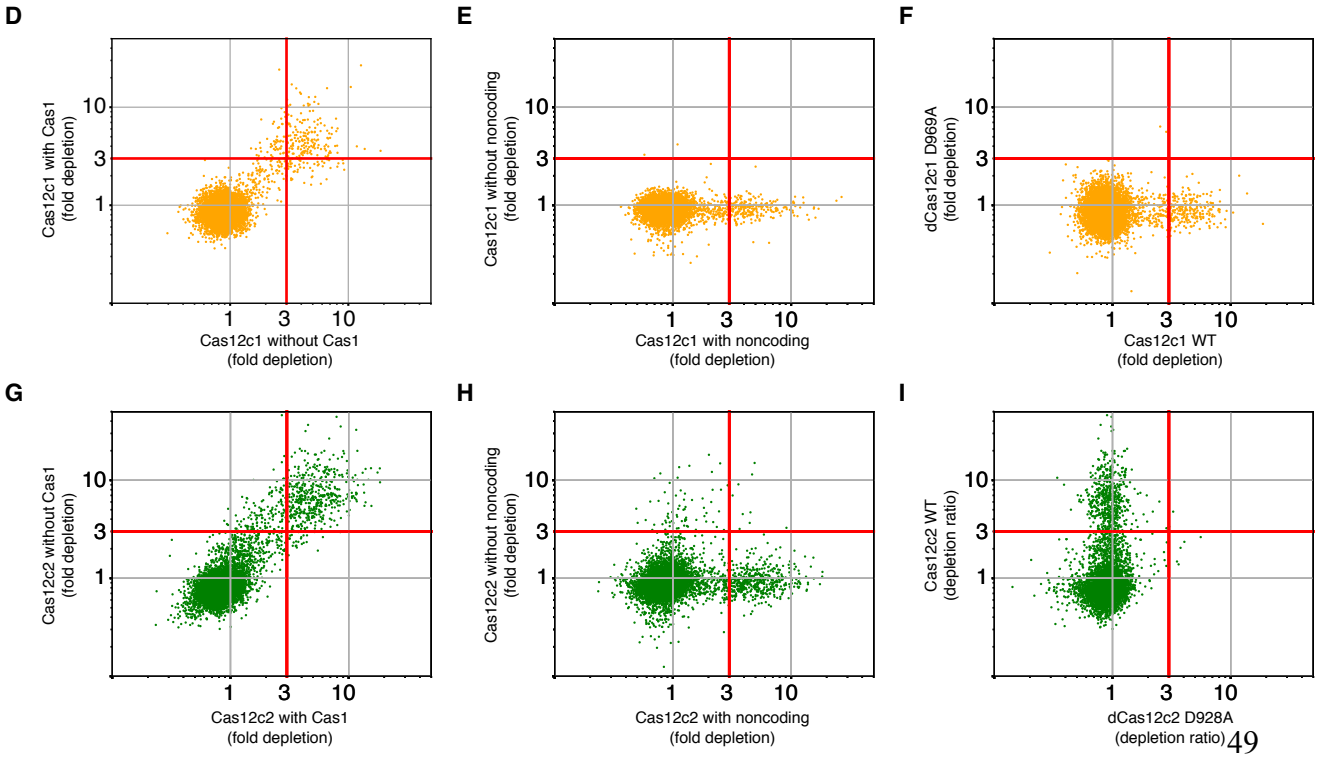

D
Cas12c1 with Cas1 (fold depletion)
Cas12c1 without Cas1 (fold depletion)
E
Cas12c1 without noncoding (fold depletion)
Cas12c1 with noncoding (fold depletion)
F
dCas12c1 D969A (fold depletion)
Cas12c1 WT (fold depletion)
G
Cas12c2 without Cas1 (fold depletion)
Cas12c2 with Cas1 (fold depletion)
H
Cas12c2 without noncoding (fold depletion)
Cas12c2 with noncoding (fold depletion)
I
Cas12c2 WT (depletion ratio)
dCas12c2 D928A (depletion ratio)

**Fig. S18. Dependencies of Cas12c *in vivo* interference activity**

(A to C) Spatial distribution of spacers from strongly depleted CRISPR arrays from *in vivo* screening of Cas12c and its noncoding plasmid mapped over pACYC184 plasmid and *E. coli* essential gene targets, separated by target strand (top: bottom strand-targeting; bottom: top strand-targeting; color saturation: spacer fold depletion.) Orange and green colors on the bars indicate the orientation of the active DR of the CRISPR arrays, with green representing the forward DR orientation and orange the reverse. (D to I) Comparison of Cas12c *in vivo* screening activity with different system components, showing minimal effect from removal of Cas1 from (D) Cas12c1 and (G) Cas12c2 systems, but significantly reduced interference activity from the removal of the noncoding plasmid in (E) Cas12c1 and (H) Cas12c2, as well as from the RuvC mutation in (F) Cas12c1 and (I) Cas12c2. As with the *in vivo* screen spatial distribution map, the coloring of the scatterplot represents the orientation of the active DR. Red lines indicate threshold for "strongly depleted" (>3 fold depletion).

# Figure S19

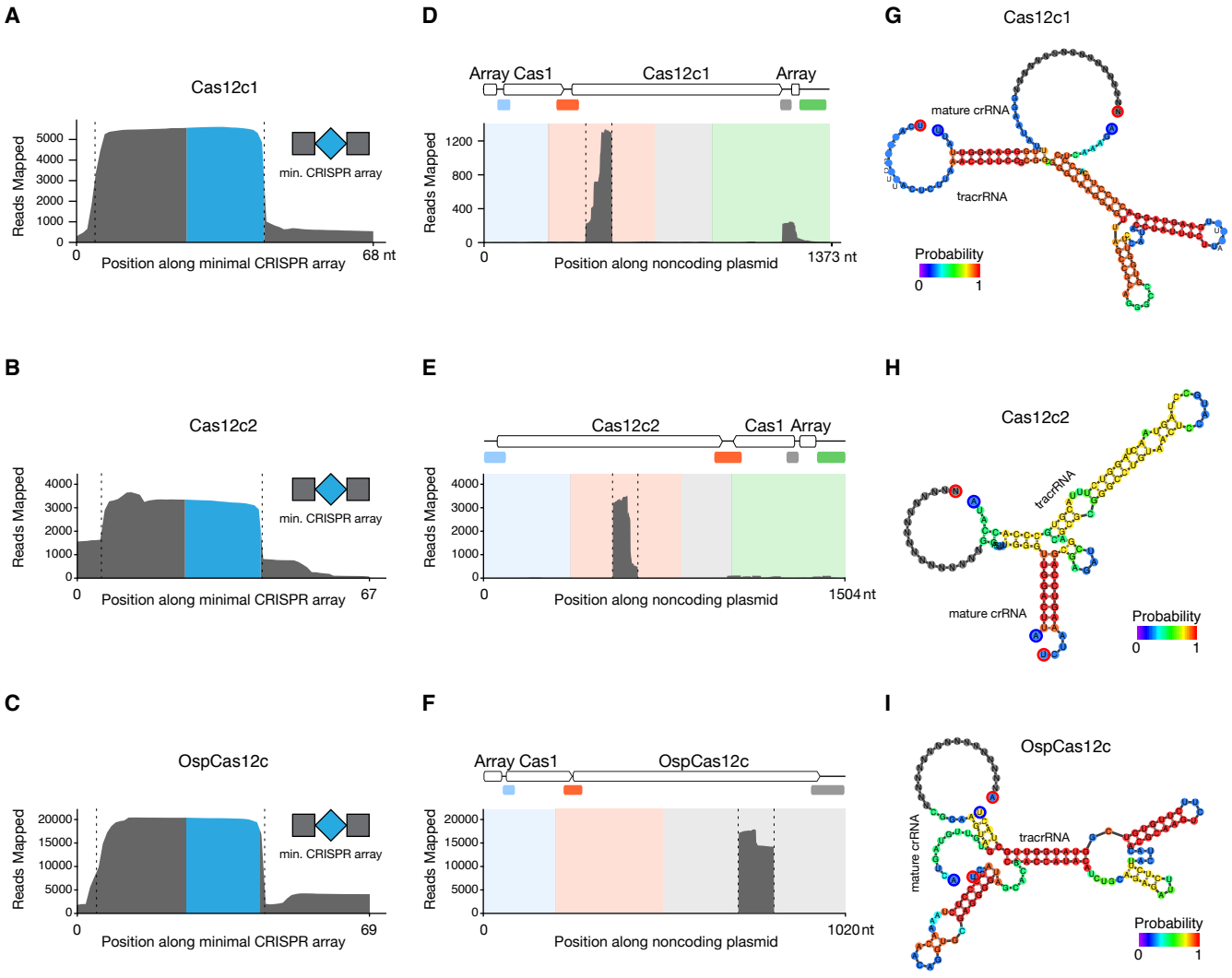

A
Cas12c1
Reads Mapped
5000
4000
3000
2000
1000
0
min. CRISPR array
0
Position along minimal CRISPR array
68 nt
B
Cas12c2
Reads Mapped
4000
3000
2000
1000
0
min. CRISPR array
0
Position along minimal CRISPR array
67
C
OspCas12c
Reads Mapped
20000
15000
10000
5000
0
min. CRISPR array
0
Position along minimal CRISPR array
69
D
Array Cas1
Cas12c1
Array
Reads Mapped
1200
800
400
0
0
Position along noncoding plasmid
1373 nt
E
Cas12c2
Cas1 Array
Reads Mapped
4000
3000
2000
1000
0
0
Position along noncoding plasmid
1504 nt
F
Array Cas1
OspCas12c
Reads Mapped
20000
15000
10000
5000
0
0
Position along noncoding plasmid
1020 nt
G
Cas12c1
mature crRNA
tracrRNA
Probability
0
1
H
Cas12c2
tracrRNA
mature crRNA
Probability
0
1
I
OspCas12c
mature crRNA
tracrRNA
Probability
0
1

**Fig. S19. Cas12c mature crRNA and tracrRNA**

(A to C) The mature crRNA and (D-F) tracrRNA as determined by small RNAseq of *E. coli* expressing the Cas12c1, Cas12c2, and OspCas12c minimal systems in the *in vivo* screen. The schematics above the tracrRNAs show sampling of the noncoding plasmid sequences from the native Cas12c loci. (G to I) Co-fold of the mature crRNA and tracrRNA. Colors indicate probability of folding, with gray corresponding to variable spacer sequences. Blue and red circles outline the 5' and 3' ends of the RNA, respectively.

**Table S1: Amino Acid Sequences of Cas12 effectors**

| System | Amino Acid Sequence |
|---|---|
| Cas12g1 | MAQASSTPAVSPRPRPRYREERTLVRKLLPRPGQSKQEFRENVKKLRKAFLQFN<br>ADVSGVCQWAIQFRPRYGKPAEPTETFWKFFLEPETSLPPNDSRSPEFRRLQAF<br>EAAAGINGAAALDDPAFTNELRDSILAVASRPKTKEAQRLFSRLKDYQPAHRMI<br>LAKVAAEWIESRYRRAHQNWERNYEEWKKEKQEWEQNHPELTPEIREAFNQIFQ<br>QLEVKEKRVRICPAARLLQNKDNCQYAGKNKHSVLCNQFNEFKKNHLQGKAIKF<br>FYKDAEKYLRCGLQSLKPNVQGPFREDWNKYLRYMNLKEETLRGKNGGRLPHCK<br>NLGQECEFNPHTALCKQYQQQLSSRPDLVQHDELYRKWRREYWREPRKPVFRYP<br>SVKRHSIAKIFGENYFQADFKNSVVGLRLDSMPAGQYLEFAFAPWPRNYRPQPG<br>ETEISSVHLHFVGTRPRIGFRFRVPHKRSRFDCTQEELDELRSRTFPRKAQDQK<br>FLEAARKRLLETFPGNAEQELRLLAVDLGTDSARAAFFIGKTFQQAFPLKIVKI<br>EKLYEQWPNQKQAGDRRDASSKQPRPGLSRDHVGRHLQKMRAQASEIAQKRQEL<br>TGTPAPETTTDQAAKKATLQPFDLRGLTVHTARMIRDWARLNARQIIQLAEENQ<br>VDLIVLESLRGFRPPGYENLDQEKKRRVAFFAHGRIRRKVTEKAVERGMRVVTV<br>PYLASSKVCAECRKKQKDNKQWEKNKKRGLFKCEGCGSQAQVDENAARVLGRVF<br>WGEIELPTAIP** |
| Cas12h1 | MKVHEIPRSQLLKIKQYEGSFVEWYRDLQEDRKKFASLLFRWAAFGYAAREDDG<br>ATYISPSQALLERRLLLGDAEDVAIKFLDVLFKGGAPSSSCYSLFYEDFALRDK<br>AKYSGAKREFIEGLATMPLDKIIERIRQDEQLSKIPAEEWLILGAEYSPEEIWE<br>QVAPRIVNVDRSLGKQLRERLGIKCRRPHDAGYCKILMEVVARQLRSHNETYHE<br>YLNQTHEMKTKVANNLTNEFDLVCEFAEVLEEKNYGLGWYVLWQGVKQALKEQK<br>KPTKIQIAVDQLRQPKFAGLLTAKWRALKGAYDTWKLKKRLEKRKAFPYMPNWD<br>NDYQIPVGLTGLGVFTLEVKRTEVVVDLKEHGKLFCSHSHYFGDLTAEKHPSRY<br>HLKFRHKLKLRKRDSRVEPTIGPWIEAALREITIQKKPNGVFYLGLPYALSHGI<br>DNFQIAKRFFSAAKPDKEVINGLPSEMVVGAADLNLSNIVAPVKARIGKGLEGP<br>LHALDYGYGELIDGPKILTPDGPRCGELISLKRDIVEIKSAIKEFKACQREGLT<br>MSEETTTWLSEVESPSDSPRCMIQSRIADTSRRLNSFKYQMNKEGYQDLAEALR<br>LLDAMDSYNSLLESYQRMHLSPGEQSPKEAKFDTKRASFRDLLRRRVAHTIVEY<br>FDDCDIVFFEDLDGPSDSDSRNNALVKLLSPRTLLLYIRQALEKRGIGMVEVAK<br>DGTSQNNPISGHVGWRNKQNKSEIYFYEDKELLVMDADEVGAMNILCRGLNHSV<br>CPYSFVTKAPEKKNDEKKEGDYGKRVKRFLKDRYGSSNVRFLVASMGFVTVTTK<br>RPKDALVGKRLYYHGGELVTHDLHNRMKDEIKYLVEKEVLARRVSLSDSTIKSY<br>KSFAHV** |
| Cas12i1 | MSNKEKNASETRKAYTTKMIPRSHDRMKLLGNFMDYLMDGTPIFFELWNQFGGG<br>IDRDIISGTANKDKISDDLLLAVNWFKVMPINSKPQGVSPSNLANLFQQYSGSE<br>PDIQAQEYFASNFDTEKHQWKDMRVEYERLLAELQLSRSDMHHDLKLMYKEKCI<br>GLSLSTAHYITSVMFGTGAKNNRQTKHQFYSKVIQLLEESTQINSVEQLASIIL<br>KAGDCDSYRKLRIRCSRKGATPSILKIVQDYELGTNHDDEVNVPSLIANLKEKL<br>GRFEYECEWKCMEKIKAFLASKVGPYYLGSYSAMLENALSPIKGMTTKNCKFVL<br>KQIDAKNDIKYENEPFGKIVEGFFDSPYFESDTNVKWVLHPHHIGESNIKTLWE<br>DLNAIHSKYEEDIASLSEDKKEKRIKVYQGDVCQTINTYCEEVGKEAKTPLVQL<br>LRYLYSRKDDIAVDKIIDGITFLSKKHKVEKQKINPVIQKYPSFNFGNNSKLLG<br>KIISPKDKLKHNLKCNRNQVDNYIWIEIKVLNTKTMRWEKHHYALSSTRFLEEV<br>YYPATSENPPDALAARFRTKTNGYEGKPALSAEQIEQIRSAPVGLRKVKKRQMR |

| | |
|---|---|
| | LEAARQQNLLPRYTWGKDFNINICKRGNNFEVTLATKVKKKKEKNYKVVLGYDA<br>NIVRKNTYAAIEAHANGDGVIDYNDLPVKPIESGFVTVESQVRDKSYDQLSYNG<br>VKLLYCKPHVESRRSFLEKYRNGTMKDNRGNNIQIDFMKDFEAIADDETSLYYF<br>NMKYCKLLQSSIRNHSSQAKEYREEIFELLRDGKLSVLKLSSLSNLSFVMFKVA<br>KSLIGTYFGHLLKKPKNSKSDVKAPPITDEDKQKADPEMFALRLALEEKRLNKV<br>KSKKEVIANKIVAKALELRDKYGPVLIKGENISDTTKKGKKSSTNSFLMDWLAR<br>GVANKVKEMVMMHQGLEFVEVNPNFTSHQDPFVHKNPENTFRARYSRCTPSELT<br>EKNRKEILSFLSDKPSKRPTNAYYNEGAMAFLATYGLKKNDVLGVSLEKFKQIM<br>ANILHQRSEDQLLFPSRGGMFYLATYKLDADATSVNWNGKQFWVCNADLVAAYN<br>VGLVDIQKDFKKK** |
| Cas12i2 | MSSAIKSYKSVLRPNERKNQLLKSTIQCLEDGSAFFFKMLQGLFGGITPEIVRF<br>STEQEKQQQDIALWCAVNWFRPVSQDSLTHTIASDNLVEKFEEYYGGTASDAIK<br>QYFSASIGESYYWNDCRQQYYDLCRELGVEVSDLTHDLEILCREKCLAVATESN<br>QNNSIISVLFGTGEKEDRSVKLRITKKILEAISNLKEIPKNVAPIQEIILNVAK<br>ATKETFRQVYAGNLGAPSTLEKFIAKDGQKEFDLKKLQTDLKKVIRGKSKERDW<br>CCQEELRSYVEQNTIQYDLWAWGEMFNKAHTALKIKSTRNYNFAKQRLEQFKEI<br>QSLNNLLVVKKLNDFFDSEFFSGEETYTICVHHLGGKDLSKLYKAWEDDPADPE<br>NAIVVLCDDLKNNFKKEPIRNILRYIFTIRQECSAQDILAAAKYNQQLDRYKSQ<br>KANPSVLGNQGFTWTNAVILPEKAQRNDRPNSLDLRIWLYLKLRHPDGRWKKHH<br>IPFYDTRFFQEIYAAGNSPVDTCQFRTPRFGYHLPKLTDQTAIRVNKKHVKAAK<br>TEARIRLAIQQGTLPVSNLKITEISATINSKGQVRIPVKFDVGRQKGTLQIGDR<br>FCGYDQNQTASHAYSLWEVVKEGQYHKELGCFVRFISSGDIVSITENRGNQFDQ<br>LSYEGLAYPQYADWRKKASKFVSLWQITKKNKKKEIVTVEAKEKFDAICKYQPR<br>LYKFNKEYAYLLRDIVRGKSLVELQQIRQEIFRFIEQDCGVTRLGSLSLSTLET<br>VKAVKGIIYSYFSTALNASKNNPISDEQRKEFDPELFALLEKLELIRTRKKKQK<br>VERIANSLIQTCLENNIKFIRGEGDLSTTNNATKKKANSRSMDWLARGVFNKIR<br>QLAPMHNITLFGCGSLYTSHQDPLVHRNPDKAMKCRWAAIPVKDIGDWVLRKLS<br>QNLRAKNIGTGEYYHQGVKEFLSHYELQDLEEELLKWRSDRKSNIPCWVLQNRL<br>AEKLGNKEAVVYIPVRGGRIYFATHKVATGAVSIVFDQKQVWVCNADHVAAANI<br>ALTVKGIGEQSSDEENPDGSRIKLQLTS** |
| Cas12c1 | MQTKKTHLHLISAKASRKYRRTIACLSDTAKKDLERRKQSGAADPAQELSCLKT<br>IKFKLEVPEGSKLPSFDRISQIYNALETIEKGSLSYLLFALILSGFRIFPNSSA<br>AKTFASSSCYKNDQFASQIKEIFGEMVKNFIPSELESILKKGRRKNNKDWTEEN<br>IKRVLNSEFGRKNSEGSSALFDSFLSKFSQELFRKFDSWNEVNKKYLEAAELLD<br>SMLASYGPFDSVCKMIGDSDSRNSLPDKSTIAFTNNAEITVDIESSVMPYMAIA<br>ALLREYRQSKSKAAPVAYVQSHLTTTNGNGLSWFFKFGLDLIRKAPVSSKQSTS<br>DGSKSLQELFSVPDDKLDGLKFIKEACEALPEASLLCGEKGELLGYQDFRTSFA<br>GHIDSWVANYVNRLFELIELVNQLPESIKLPSILTQKNHNLVASLGLQEAEVSH<br>SLELFEGLVKNVRQTLKKLAGIDISSSPNEQDIKEFYAFSDVLNRLGSIRNQIE<br>NAVQTAKKDKIDLESAIEWKEWKKLKKLPKLNGLGGGVPKQQELLDKALESVKQ<br>IRHYQRIDFERVIQWAVNEHCLETVPKFLVDAEKKKINKESSTDFAAKENAVRF<br>LLEGIGAAARGKTDSVSKAAYNWFVVNNFLAKKDLNRYFINCQGCIYKPPYSKR<br>RSLAFALRSDNKDTIEVVWEKFETFYKEISKEIEKFNIFSQEFQTFLHLENLRM<br>KLLLRRIQKPIPAEIAFFSLPQEYYDSLPPNVAFLALNQEITPSEYITQFNLYS<br>SFLNGNLILLRRSRSYLRAKFSWVGNSKLIYAAKEARLWKIPNAYWKSDEWKMI<br>LDSNVLVFDKAGNVLPAPTLKKVCEREGDLRLFYPLLRQLPHDWCYRNPFVKSV<br>GREKNVIEVNKEGEPKVASALPGSLFRLIGPAPFKSLLDDCFFNPLDKDLRECM |

| | |
|---|---|
| | LIVDQEISQKVEAQKVEASLESCTYSIAVPIRYHLEEPKVSNQFENVLAIDQGE<br>AGLAYAVFSLKSIGEAETKPIAVGTIRIPSIRRLIHSVSTYRKKKQRLQNFKQN<br>YDSTAFIMRENVTGDVCAKIVGLMKEFNAFPVLEYDVKNLESGSRQLSAVYKAV<br>NSHFLYFKEPGRDALRKQLWYGGDSWTIDGIEIVTRERKEDGKEGVEKIVPLKV<br>FPGRSVSARFTSKTCSCCGRNVFDWLFTEKKAKTNKKFNVNSKGELTTADGVIQ<br>LFEADRSKGPKFYARRKERTPLTKPIAKGSYSLEEIERRVRTNLRRAPKSKQSR<br>DTSQSQYFCVYKDCALHFSGMQADENAAINIGRRFLTALRKNRRSDFPSNVKIS<br>DRLLDN** |
| Cas12c2 | MTKHSIPLHAFRNSGADARKWKGRIALLAKRGKETMRTLQFPLEMSEPEAAAIN<br>TTPFAVAYNAIEGTGKGTLFDYWAKLHLAGFRFFPSGGAATIFRQQAVFEDASW<br>NAAFCQQSGKDWPWLVPSKLYERFTKAPREVAKKDGSKKSIEFTQENVANESHV<br>SLVGASITDKTPEDQKEFFLKMAGALAEKFDSWKSANEDRIVAMKVIDEFLKSE<br>GLHLPSLENIAVKCSVETKPDNATVAWHDAPMSGVQNLAIGVFATCASRIDNIY<br>DLNGGKLSKLIQESATTPNVTALSWLFGKGLEYFRTTDIDTIMQDFNIPASAKE<br>SIKPLVESAQAIPTMTVLGKKNYAPFRPNFGGKIDSWIANYASRLMLLNDILEQ<br>IEPGFELPQALLDNETLMSGIDMTGDELKELIEAVYAWVDAAKQGLATLLGRGG<br>NVDDAVQTFEQFSAMMDTLNGTLNTISARYVRAVEMAGKDEARLEKLIECKFDI<br>PKWCKSVPKLVGISGGLPKVEEEIKVMNAAFKDVRARMFVRFEEIAAYVASKGA<br>GMDVYDALEKRELEQIKKLKSAVPERAHIQAYRAVLHRIGRAVQNCSEKTKQLF<br>SSKVIEMGVFKNPSHLNNFIFNQKGAIYRSPFDRSRHAPYQLHADKLLKNDWLE<br>LLAEISATLMASESTEQMEDALRLERTRLQLQLSGLPDWEYPASLAKPDIEVEI<br>QTALKMQLAKDTVTSDVLQRAFNLYSSVLSGLTFKLLRRSFSLKMRFSVADTTQ<br>LIYVPKVCDWAIPKQYLQAEGEIGIAARVVTESSPAKMVTEVEMKEPKALGHFM<br>QQAPHDWYFDASLGGTQVAGRIVEKGKEVGKERKLVGYRMRGNSAYKTVLDKSL<br>VGNTELSQCSMIIEIPYTQTVDADFRAQVQAGLPKVSINLPVKETITASNKDEQ<br>MLFDRFVAIDLGERGLGYAVFDAKTLELQESGHRPIKAITNLLNRTHHYEQRPN<br>QRQKFQAKFNVNLSELRENTVGDVCHQINRICAYYNAFPVLEYMVPDRLDKQLK<br>SVYESVTNRYIWSSTDAHKSARVQFWLGGETWEHPYLKSAKDKKPLVLSPGRGA<br>SGKGTSQTCSCCGRNPFDLIKDMKPRAKIAVVDGKAKLENSELKLFERNLESKD<br>DMLARRHRNERAGMEQPLTPGNYTVDEIKALLRANLRRAPKNRRTKDTTVSEYH<br>CVFSDCGKTMHADENAAVNIGGKFIADIEK** |
| OspCas12c | MTKLRHRQKKLTHDWAGSKKREVLGSNGKLQNPLLMPVKKGQVTEFRKAFSAYA<br>RATKGEMTDGRKNMFTHSFEPFKTKPSLHQCELADKAYQSLHSYLPGSLAHFLL<br>SAHALGFRIFSKSGEATAFQASSKIEAYESKLASELACVDLSIQNLTISTLFNA<br>LTTSVRGKGEETSADPLIARFYTLLTGKPLSRDTQGPERDLAEVISRKIASSFG<br>TWKEMTANPLQSLQFFEEELHALDANVSLSPAFDVLIKMNDLQGDLKNRTIVFD<br>PDAPVFEYNAEDPADIIIKLTARYAKEAVIKNQNVGNYVKNAITTTNANGLGWL<br>LNKGLSLLPVSTDDELLEFIGVERSHPSCHALIELIAQLEAPELFEKNVFSDTR<br>SEVQGMIDSAVSNHIARLSSSRNSLSMDSEELERLIKSFQIHTPHCSLFIGAQS<br>LSQQLESLPEALQSGVNSADILLGSTQYMLTNSLVEESIATYQRTLNRINYLSG<br>VAGQINGAIKRKAIDGEKIHLPAAWSELISLPFIGQPVIDVESDLAHLKNQYQT<br>LSNEFDTLISALQKNFDLNFNKALLNRTQHFEAMCRSTKKNALSKPEIVSYRDL<br>LARLTSCLYRGSLVLRRAGIEVLKKHKIFESNSELREHVHERKHFVFVSPLDRK<br>AKKLLRLTDSRPDLLHVIDEILQHDNLENKDRESLWLVRSGYLLAGLPDQLSSS<br>FINLPIITQKGDRRLIDLIQYDQINRDAFVMLVTSAFKSNLSGLQYRANKQSFV<br>VTRTLSPYLGSKLVYVPKDKDWLVPSQMFEGRFADILQSDYMVWKDAGRLCVID<br>TAKHLSNIKKSVFSSEEVLAFLRELPHRTFIQTEVRGLGVNVDGIAFNNGDIPS |

| | |
|---|---|
| | LKTFSNCVQVKVSRTNTSLVQTLNRWFEGGKVSPPSIQFERAYYKKDDQIHEDA<br>AKRKIRFQMPATELVHASDDAGWTPSYLLGIDPGEYGMGLSLVSINNGEVLDSG<br>FIHINSLINFASKKSNHQTKVVPRQQYKSPYANYLEQSKDSAAGDIAHILDRLI<br>YKLNALPVFEALSGNSQSAADQVWTKVLSFYTWGDNDAQNSIRKQHWFGASHWD<br>IKGMLRQPPTEKKPKPYIAFPGSQVSSYGNSQRCSCCGRNPIEQLREMAKDTSI<br>KELKIRNSEIQLFDGTIKLFNPDPSTVIERRRHNLGPSRIPVADRTFKNISPSS<br>LEFKELITIVSRSIRHSPEFIAKKRGIGSEYFCAYSDCNSSLNSEANAAANVAQ<br>KFQKQLFFEL** |

**Table S2: Plasmid Links**

| | |
|---|---|
| pET28a-mH6-Cas12g1 | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-zD08YEFj-pet28a-mh6-cas12g1/edit |
| pACYC-Cas12g1-noncoding | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-I2rcuanp-pacyc-cas12g1-noncoding/edit |
| pET28a-mH6-Cas12h1 | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-FYp6Zgxw-pet28a-mh6-cas12h1/edit |
| pET28a-mH6-Cas12i1 | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-4wVTebpc-pet28a-mh6-cas12i1/edit |
| pET28a-mH6-Cas12i2 | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-Aesci1Wf-pet28a-mh6-cas12i2/edit |
| pET28a-mH6-Cas12c1 | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-HAZpFEJC-pet28a-mh6-cas12c1/edit |
| pACYC-Cas12c1-noncoding | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-zFkbYnID-pacyc-cas12c1-noncoding/edit |
| pET28a-mH6-Cas12c2 | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-0ipkQtdj-pet28a-mh6-cas12c2/edit |
| pACYC-Cas12c2-noncoding | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-o9UThBIE-pacyc-cas12c2-noncoding/edit |
| pET28a-mH6-OspCas12c | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-8Ae4l863-pet28a-mh6-ospcas12c/edit |
| pACYC-OspCas12c-noncoding | https://benchling.com/arbor/f/j7c7QlCh-yan-et-al-2018b-type-v/seq-RB26TA6N-pacyc-ospcas12c-noncoding/edit |

**Table S3: Pre-crRNAs used in this study for *in vitro* biochemistry**

| Name | Sequence | DR | Spacer1 | Spacer2 | Target | Fig |
|---|---|---|---|---|---|---|
| Cas12g1 pre-crRNA 1 | gggGGUGGAAAGGGCCGGAGAUUUACCGGCUCUGACACCCCAUUCAUCCGCUUAUUAUCACUUAUUCAGGGUGGAAAGGGCCGGAGAUUUACCGGCUCUGACACCGCUUGGAGGAGCGCAGUCACCAAAACUUGUGGUGGAAAGGGCCGGAGAUUUACCGGCUCUGACACC | GGUGGAAAGGGCCGGAGAUUUACCGGCUCUGACACC | CCAUUCAUCCGCUUAUUAUCACUUAUUCAG | GCUUGGAGGAGCGCAGUCACCAAAACUUGU | Cas12g1 Target 1 | Fig S5 |
| Cas12i1 pre-crRNA 1 | gggAAUUUUUGUGCCCAUCGUUGGCACCCUAAUGCGGAAGUAGUGGGUAACCCGGAAUUUUUGUGCCCAUCGUUGGCACUCCGCAAGAAUUGAUUGGCUCCAAUUCUAAUUUUUGUGCCCAUCGUUGGCAC | AAUUUUUGUGCCCAUCGUUGGCAC | CCUAAUGCGGAAGUAGUGGGUAACCCGG | UCCGCAAGAAUUGAUUGGCUCCAAUUCU | Cas12i1 Target 1 | Fig S16 |
| Cas12i1 pre-crRNA 2 | gggAAUUUUUGUGCCCAUCGUUGGCACAGGCAUCAUCAGCAUUAACCACGCAAACAAUUUUUGUGCCCAUCGUUGGCACGCGUGCUGGAUUGCUUCGAUGGUCUGCGAAUUUUUGUGCCCAUCGUUGGCAC | AAUUUUUGUGCCCAUCGUUGGCAC | AGGCAUCAUCAGCAUUAACCACGCAAAC | GCGUGCUGGAUUGCUUCGAUGGUCUGCG | Cas12i1 Target 2 | Fig 3C-3F, Fig S17 |

**Table S4: TracrRNAs used in this study for *in vitro* biochemistry**

| Name | Sequence | Fig |
|---|---|---|
| Cas12g1 tracr 1 | gggGAUGCUUACUUAGUCAUCUGGUUGGCAAACCUCCGCGGACCUUCGGGACCAA UGGAGAGGAACCCAGCCGAGAAGCAUCGAGCCGGUAAAUGCCGGAAAUUUUU | Fig S5 |
| Cas12g1 tracr 2 | gggGAUGCUUACUUAGUCAUCUGGUUGGCAAACCUCCGCGGACCUUCGGGACCAA UGGAGAGGAACCCAGCCGAGAAGCAUCGAGCCGGUAAAUGCCGGAAA | Fig 2E-H, Fig S6, Fig S8, Fig S9, Fig S10, Fig S11, Fig S12 |
| Cas12g1 tracr 3 | gggGAUGCUUACUUAGUCAUCUGGUUGGCAAACCUCCGCGGACCUUCGGGACCAA UGGAGAGGAACCCAGCCGAGAAGCAUCGAGCCGGUAAAU | Fig S7 |

**Table S5: Mature crRNAs used in this study for *in vitro* biochemistry**

| Name | Sequence | Processed DR | Processed Spacer | Target | Fig |
|---|---|---|---|---|---|
| Cas12g1m crRNA 1 | gggUUUACCGGCUCUGACACCGCUUGGAGGAGCGCAGUCACCAAA | UUUACCGGCUCUGACACC | GCUUGGAGGAGCGCAGUCACCAAA | Cas12g1 ssRNA 1 | Fig S7 |
| Cas12g1m crRNA 2 | gggUUUACCGGCUCUGACACCAACUGUGAUAAACUACCGCAUUAA | UUUACCGGCUCUGACACC | AACUGUGAUAAACUACCGCAUUAA | Cas12g1 ssRNA 2 | Fig 2E-H, Fig S8, Fig S9, Fig S10, Fig S11C-D, Fig S12 |
| Cas12g1m crRNA 3 | gggUUUACCGGCUCUGACACCCUUCCUUAGCUCCUGAAAAUCUCG | UUUACCGGCUCUGACACC | CUUCCUUAGCUCCUGAAAAUCUCG | Cas12g1 ssRNA 3 | Fig S11A-B |
| Cas12g1m crRNA 4 | gggUUUACCGGCUCUGACACCCAUCUUGCGAAUAUAUGUGUAGAA | UUUACCGGCUCUGACACC | CAUCUUGCGAAUAUAUGUGUAGAA | Cas12g1 ssDNA 4 | Fig S6 |

**Table S6: Target substrates used in this study for *in vitro* biochemistry**

| Name | Sequence | Nucleic acid | Fig |
|---|---|---|---|
| Cas12g1 ssRNA 2 | CAACGAGGUAUUUAUCCGAUU AAUUCUCAUGUUUGACAGCUU AUCAUCGAUAAGCUUUAAUGC GGUAGUUUAUCACAGUUAAAU UGCUAACGCAGUCAGGCACCG UGUAUUUGUAGCCCGGGGACU GUUU | RNA | Fig 2E-H, Fig S8, Fig S9, Fig S10, Fig S11C, Fig S12 |
| Cas12g1 ssRNA 3 | CAACGAGGUAUUUAUCCGAUA AAUAAGAUCACUACCGGGCGU AUUUUUUGAGUUAUCGAGAUU UUCAGGAGCUAAGGAAGCUAA AAUGGAGAAAAAAAUCACUGG AUAUAUUGUAGCCCGGGGACU GUUU | RNA | Fig S11A |
| Cas12g1 sense RNA 2 | CAACGAGGUAUUUAUCCGAUU AAUUCUCAUGUUUGACAGCUU AUCAUCGAUAAGCUUUAAUGC GGUAGUUUAUCACAGUUAAAU UGCUAACGCAGUCAGGCACCG UGUAUUUGUAGCCCGGGGACU GUUU | RNA | Fig S8 |
| Cas12g1 sense DNA 2 | CAACGAGGTATTTATCCGATT AATTCTCATGTTTGACAGCTT ATCATCGATAAGCTTTAATGC GGTAGTTTATCACAGTTAAAT TGCTAACGCAGTCAGGCACCG TGTATTTGTAGCCCGGGGACT GTTT | DNA | Fig S8 |
| Cas12g1 antisense RNA 2 | AAACAGUCCCCGGGCUACAAA UACACGGUGCCUGACUGCGUU AGCAAUUUAACUGUGAUAAAC UACCGCAUUAAAGCUUAUCGA UGAUAAGCUGUCAAACAUGAG AAUUAAUCGGAUAAAUACCUC GUUG | RNA | Fig S8 |
| Cas12g1 antisense DNA 2 | AAACAGTCCCCGGGCTACAAA TACACGGTGCCTGACTGCGTT AGCAATTTAACTGTGATAAAC TACCGCATTAAAGCTTATCGA TGATAAGCTGTCAAACATGAG AATTAATCGGATAAATACCTC GTTG | DNA | Fig S8 |
| Cas12g1 ssDNA 4, dsDNA 4 | CAACGAGGTATTTATCCGATCGC TCTGGAGTGAATACCACGACGAT TTCCGGCAGTTTCTACACATATAT TCGCAAGATGTGGCGTGTTACGG TGAAAACCTGGCCTATTTTGTAG CCCGGGGACTGTTT | DNA | Fig S6 |

| | | | |
|---|---|---|---|
| Cas12i1 ssDNA2, dsDNA2 | CATGTGGACCACATTAGGCTTGGTTGTTGCTGCCGACGACGGTGTGATGCCGCAGACCATCGAAGCAATCCAGCACGCGAAAGCGGCGCAGGTACCGGTGGTGGTTGCGTAATTTCTCGAGGCGTATT | DNA | Fig 3C-F, Fig S17 |

**Table S7: Non target substrates used in this study for *in vitro* biochemistry**

| Name | Sequence | Nucleic acid | Fig |
|---|---|---|---|
| Cas12g1 ssRNA 2_RC | AAACAGUCCCCGGGCUACAAAUACACGGUGCCUGACUGCGUUAGCAAUUUAACUGUGAUAAACUACCGCAUUAAAGCUUAUCGAUGAUAAGCUGUCAAACAUGAGAAUUAAUCGGAUAAAUACCUCGUUG | RNA | Fig 2E-F, Fig S9, Fig S10, Fig S11D |
| Cas12g1 ssRNA 3_RC | AAACAGUCCCCGGGCUACAAUAUAUCCAGUGAUUUUUUUCUCCAUUUUAGCUUCCUUAGCUCCUGAAAAUCUCGAUAACUCAAAAAAUACGCCCGGUAGUGAUCUUAUUUAUCGGAUAAAUACCUCGUUG | RNA | Fig S11B |
| Cas12i1 ssDNA1, dsDNA1 | CATGTGGACCACATTAGGCTGCAAAACTGCGCATTTACGAAAACGCGAAAGTTTGCGTGGTTAATGCTGATGATGCCTTAACAATGCCGATTCGCGGTGCGGATGAACGTAATTTCTCGAGGCGTATT | DNA | Fig 3 |

**Table S8: Collateral substrates used in this study for *in vitro* biochemistry**

| Name | Sequence | Nucleic acid | Fig |
|---|---|---|---|
| Cas12g1 ssRNA 5_RC | AAACAGUCCCCGGGCUACAAUCUCGCCGAAACGUUUGGUGGCGGGACCAGUGACGAAGGCUUGAGCGAGGGCGUGCAAGAUUCCGAAUACCGCAAGCGACAGGCCGAUCAAUCGGAUAAAUACCUCGUUG | RNA | Fig S11, Fig 2F |
| Cas12g1 ssDNA 4_RC | AAACAGTCCCCGGGCTACAAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATA | DNA | Fig S9, 2E, 2G |

| | | | |
|---|---|---|---|
| | TATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATCGGATAAATACCTCGTTG | | |
| Cas12i1 ssDNA6_RC | AATACGCCTCGAGAAATTACAAAGTGATGCAGGCGTTTCCAGGTGCTTTCCCTAATGCGGAAGTAGTGGGTAACCCGGTGCGTACCGATGTGTTGGCGCTGCCGTTGCAGCCTAATGTGGTCCACATG | DNA | Fig 3D |

**Table S9: IDT Template oligos and primers for crRNAs used in this study for *in vitro* biochemistry**

| Name | Template Sequence | T7 fwd primer | Rev primer |
|---|---|---|---|
| Cas12g1 pre-crRNA 1 | GGTGTCAGAGCCGGTAAATCTCCGGCCCTTTCCACCACAAGTTTTGGTGACTGCGCTCCTCCAAGCGGTGTCAGAGCCGGTAAATCTCCGGCCCTTTCCACCCTGAATAAGTGATAATAAGCGGATGAATGGGGTGTCAGAGCCGGTAAATCTCCGGCCCTTTCCACCcctatagtgagtcgtatta | TAATACGACTCACTATAG | GGTGTCAGAGCCGGTAAATCTCCGGCCCTTTCCACCACAAGTTTTGGTGACTGCGCTCCTCCAAGC |
| Cas12i1 pre-crRNA 1 | GTGCCAACGATGGGCACAAAAATTAGAATTGGAGCCAATCAATTCTTGCGGAGTGCCAACGATGGGCACAAAAATTAGAATTGGAGCCAATCAATTCTTGCGGAGTGCCAACGATGGGCACAAAAATTccctatagtgagtcgtattactcgagggatccTTATTACATTT | TAATACGACTCACTATAG | GTGCCAACGATGGGCACAAAAATTAGAATTGGAGCCAATCAATTCTTGCGGA |
| Cas12i1 pre-crRNA 2 | GTGCCAACGATGGGCACAAAAATTCGCAGACCATCGAAGCAATCCAGCACGCGTGCCAACGATGGGCACAAAAATTGTTTGCGTGGTTAATGCTGATGATGCCTGTGCCAACGATGGGCACAAAAATTccctatagtgagtcgtattactcgagggatccTTATTACATTT | TAATACGACTCACTATAG | GTGCCAACGATGGGCACAAAAATTCGCAGACCATCGAAGCAATCCAGCACGC |
| Cas12g1 mcrRNA 2 | TTAATGCGGTAGTTTATCACAGTTGGTGTCAGAGCCGGTAAAccctatagtgagtcgtatta | TAATACGACTCACTATAG | |
| Cas12g1 mcrRNA 3 | CGAGATTTTCAGGAGCTAAGGAAGGGTGTCAGAGCCGGTAAAccctatagtgagtcgtatta | TAATACGACTCACTATAG | |
| Cas12g1 mcrRNA 4 | TTCTACACATATATTCGCAAGATGGGTGTCAGAGCCGGTAAAccctatagtgagtcgtatta | TAATACGACTCACTATAG | |
| Cas12g1 tracr 1 | AAAAATTTCCGGCATTTACCGGCTCGATGCTTCTCGGCTGGGTTCCTCTCCATTGGTCCCGAAGGTCCGCGGAGGTTTGCCAACCAGATGACTAAGTAAGCATCccctatagtgagtcgtatta | TAATACGACTCACTATAG | |
| Cas12g1 tracr 2 | TTTCCGGCATTTACCGGCTCGATGCTTCTCGGCTGGGTTCCTCTCCATTGGTCCCGAAGGTCCGCGGAGGTTTGCCAACCAGATGACTAAGTAAGCATCcccctatagtgagtcgtatta | TAATACGACTCACTATAG | |
| Cas12g1 tracr 3 | ATTTACCGGCTCGATGCTTCTCGGCTGGGTTCCTCTCCATTGGTCCCGAAGGTCCGCGGAGGTTTGCCAACCAGATGACTAAGTAAGCATCccctatagtgagtcgtatta | TAATACGACTCACTATAG | |

**Table S10: IDT template oligos and primers for ssRNA substrates used in this study for *in vitro* biochemistry**

| Name | Template Sequence | sense RNA fwd primer | sense RNA rev primer | antisense RNA fwd primer | antisense RNA rev primer |
|---|---|---|---|---|---|
| Cas12g1 ssRNA 1 | AAACAGTCCCCGGGCTACAAAGCT ACCAACTCTTTGAACCGAGGTAAC TGGCTTGGAGGAGCGCAGTCACCA AAACTTGTCCTTTCAGTTTAGCCT TAACCGGCGCATGAATCGGATAAA TACCTCGTTG | AAACAGTC CCCGGGCT ACAA | AAATGTAATAA ggatccctcga gtaatacgact cactatagggC AACGAGGTATT TATCCGAT | AAATGTAATAA ggatccctcga gtaatacgact cactatagggA AACAGTCCCCG GGCTACAA | CAACGAGGTAT TTATCCGAT |
| Cas12g1 ssRNA 2 | AAACAGTCCCCGGGCTACAAATAC ACGGTGCCTGACTGCGTTAGCAAT TTAACTGTGATAAACTACCGCATT AAAGCTTATCGATGATAAGCTGTC AAACATGAGAATTAATCGGATAAA TACCTCGTTG | AAACAGTC CCCGGGCT ACAA | AAATGTAATAA ggatccctcga gtaatacgact cactatagggC AACGAGGTATT TATCCGAT | AAATGTAATAA ggatccctcga gtaatacgact cactatagggA AACAGTCCCCG GGCTACAA | CAACGAGGTAT TTATCCGAT |
| Cas12g1 ssRNA 3 | AAACAGTCCCCGGGCTACAATATA TCCAGTGATTTTTTTCTCCATTTT AGCTTCCTTAGCTCCTGAAAATCT CGATAACTCAAAAAATACGCCCGG TAGTGATCTTATTTATCGGATAAA TACCTCGTTG | AAACAGTC CCCGGGCT ACAA | AAATGTAATAA ggatccctcga gtaatacgact cactatagggC AACGAGGTATT TATCCGAT | AAATGTAATAA ggatccctcga gtaatacgact cactatagggA AACAGTCCCCG GGCTACAA | CAACGAGGTAT TTATCCGAT |
| Cas12g1 ssRNA 4 | AAACAGTCCCCGGGCTACAAAATA GGCCAGGTTTTCACCGTAACACGC CACATCTTGCGAATATATGTGTAG AAACTGCCGGAAATCGTCGTGGTA TTCACTCCAGAGCGATCGGATAAA TACCTCGTTG | AAACAGTC CCCGGGCT ACAA | AAATGTAATAA ggatccctcga gtaatacgact cactatagggC AACGAGGTATT TATCCGAT | AAATGTAATAA ggatccctcga gtaatacgact cactatagggA AACAGTCCCCG GGCTACAA | CAACGAGGTAT TTATCCGAT |
| Cas12g1 ssRNA 5 | AAACAGTCCCCGGGCTACAATCTC GCCGAAACGTTTGGTGGCGGGACC AGTGACGAAGGCTTGAGCGAGGGC GTGCAAGATTCCGAATACCGCAAG CGACAGGCCGATCAATCGGATAAA TACCTCGTTG | AAACAGTC CCCGGGCT ACAA | AAATGTAATAA ggatccctcga gtaatacgact cactatagggC AACGAGGTATT TATCCGAT | AAATGTAATAA ggatccctcga gtaatacgact cactatagggA AACAGTCCCCG GGCTACAA | CAACGAGGTAT TTATCCGAT |

**References and Notes**


1. D. Burstein, L. B. Harrington, S. C. Strutt, A. J. Probst, K. Anantharaman, B. C. Thomas, J. A. Doudna, J. F. Banfield, New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237–241 (2017). doi:10.1038/nature21059 Medline

2. S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018). doi:10.1126/science.aar4120 Medline

3. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017). doi:10.1146/annurev-micro-090816-093830 Medline

4. R. Barrangou, P. Horvath, A decade of discovery: CRISPR functions and applications. *Nat. Microbiol.* **2**, 17092 (2017). doi:10.1038/nmicrobiol.2017.92 Medline

5. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007). doi:10.1126/science.1138140 Medline

6. D. B. T. Cox, R. J. Platt, F. Zhang, Therapeutic genome editing: Prospects and challenges. *Nat. Med.* **21**, 121–131 (2015). doi:10.1038/nm.3793 Medline

7. S. E. Klompe, S. H. Sternberg, Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems. *CRISPR J.* **1**, 141–158 (2018). doi:10.1089/crispr.2018.0012

8. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018). doi:10.1126/science.aat5011 Medline

9. K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, R. M. Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen, E. V. Koonin, An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015). doi:10.1038/nrmicro3569 Medline

10. S. Shmakov, O. O. Abudayyeh, K. S. Makarova, Y. I. Wolf, J. S. Gootenberg, E. Semenova, L. Minakhin, J. Joung, S. Konermann, K. Severinov, F. Zhang, E. V. Koonin, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell* **60**, 385–397 (2015). doi:10.1016/j.molcel.2015.10.008 Medline

11. S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, O. O. Abudayyeh, J. S. Gootenberg, K. S. Makarova, Y. I. Wolf, K. Severinov, F. Zhang, E. V. Koonin, Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017). doi:10.1038/nrmicro.2016.184 Medline

12. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586 (2012). doi:10.1073/pnas.1208507109 Medline

13. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi:10.1126/science.1225829 Medline

14. E. V. Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017). doi:10.1016/j.mib.2017.05.008 Medline

15. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015). doi:10.1016/j.cell.2015.09.038 Medline

16. J. S. Chen, E. Ma, L. B. Harrington, M. Da Costa, X. Tian, J. M. Palefsky, J. A. Doudna, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018). doi:10.1126/science.aar6245 Medline

17. O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D. B. T. Cox, S. Shmakov, K. S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E. S. Lander, E. V. Koonin, F. Zhang, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016). doi:10.1126/science.aaf5573 Medline

18. A. East-Seletsky, M. R. O'Connell, S. C. Knight, D. Burstein, J. H. D. Cate, R. Tjian, J. A. Doudna, Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270–273 (2016). doi:10.1038/nature19802 Medline

19. W. X. Yan, S. Chong, H. Zhang, K. S. Makarova, E. V. Koonin, D. R. Cheng, D. A. Scott, Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Mol. Cell* **70**, 327–339.e5 (2018). doi:10.1016/j.molcel.2018.02.028 Medline

20. L. B. Harrington, D. Burstein, J. S. Chen, D. Paez-Espino, E. Ma, I. P. Witte, J. C. Cofsky, N. C. Kyrpides, J. F. Banfield, J. A. Doudna, Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018). doi:10.1126/science.aav4294 Medline

21. J. S. Gootenberg, O. O. Abudayyeh, M. J. Kellner, J. Joung, J. J. Collins, F. Zhang, Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360**, 439–444 (2018). doi:10.1126/science.aaq0179 Medline

22. J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, F. Zhang, Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438–442 (2017). doi:10.1126/science.aam9321 Medline

23. D. B. T. Cox, J. S. Gootenberg, O. O. Abudayyeh, B. Franklin, M. J. Kellner, J. Joung, F. Zhang, RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017). doi:10.1126/science.aaq0180 Medline

24. S. Konermann, P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev, P. D. Hsu, Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665–676.e14 (2018). doi:10.1016/j.cell.2018.02.033 Medline

25. D. Dong, K. Ren, X. Qiu, J. Zheng, M. Guo, X. Guan, H. Liu, N. Li, B. Zhang, D. Yang, C. Ma, S. Wang, D. Wu, Y. Ma, S. Fan, J. Wang, N. Gao, Z. Huang, The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016). doi:10.1038/nature17944 Medline

26. T. Yamano, H. Nishimasu, B. Zetsche, H. Hirano, I. M. Slaymaker, Y. Li, I. Fedorova, T. Nakane, K. S. Makarova, E. V. Koonin, R. Ishitani, F. Zhang, O. Nureki, Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962 (2016). doi:10.1016/j.cell.2016.04.003 Medline

27. F. A. Ran, P. D. Hsu, C.-Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, F. Zhang, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013). doi:10.1016/j.cell.2013.08.021 Medline

28. J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin, D. R. Liu, Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018). doi:10.1038/nature26155 Medline

29. D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers, GenBank. *Nucleic Acids Res.* **41** (D1), D36–D42 (2013). doi:10.1093/nar/gks1195 Medline

30. K. D. Pruitt, T. Tatusova, G. R. Brown, D. R. Maglott, NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Res.* **40** (D1), D130–D135 (2012). doi:10.1093/nar/gkr1079 Medline

31. V. M. Markowitz, I.-M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, B. Jacob, J. Huang, P. Williams, M. Huntemann, I. Anderson, K. Mavromatis, N. N. Ivanova, N. C. Kyrpides, IMG: The Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40** (D1), D115–D122 (2012). doi:10.1093/nar/gkr1044 Medline

32. J. Söding, Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005). doi:10.1093/bioinformatics/bti125 Medline

33. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *CRISPR J.* **1**, 325–336 (2018). doi:10.1089/crispr.2018.0033