**RESEARCH ARTICLE SUMMARY**

**CRISPR EVOLUTION**

**C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector**


**INTRODUCTION:** Almost all archaea and about half of bacteria possess clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated genes (Cas) adaptive immune systems, which protect microbes against viruses and other foreign DNA. All functionally characterized CRISPR systems have been reported to target DNA, with some multicomponent type III systems also targeting RNA. The putative class 2 type VI system, which has not been functionally characterized, encompasses the single-effector protein C2c2, which contains two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains commonly associated with ribonucleases (RNases), suggesting RNA-guided RNA-targeting function.

**RATIONALE:** Existing studies have only established a role for RNA interference, in addition to DNA interference, in the multicomponent type III-A and III-B systems. We investigated the possibility of C2c2-mediated RNA interference by heterologously expressing C2c2 locus from *Leptotrichia shahii* (LshC2c2) in the model system *Escherichia coli*. The ability of LshC2c2 to protect against MS2 single-stranded RNA (ssRNA) phage infection was assessed by using every possible spacer sequence against the phage genome. We next developed protocols to reconstitute purified recombinant LshC2c2 protein and test its biochemical activity when incubated with its mature CRISPR RNA (crRNA) and target ssRNA. We systematically evaluated the parameters necessary for cleavage. Last, to demonstrate the potential utility of the LshC2c2 complex for RNA targeting in living bacterial cells, we guided it to knockdown red fluorescent protein (RFP) mRNA in vivo.

**RESULTS:** This work demonstrates the RNA-guided RNase activity of the putative type VI CRISPR-effector LshC2c2. Heterologously expressed C2c2 can protect *E. coli* from MS2 phage, and by screening against the MS2 genome, we identified a H (non-G) protospacer flanking site (PFS) following the RNA target site, which was confirmed by targeting a complementary sequence in the β-lactamase transcript followed by a degenerate nucleotide sequence. Using purified LshC2c2 protein, we demonstrate that C2c2 and crRNA are sufficient in vitro to achieve RNA-guided, PFS-dependent RNA cleavage. This cleavage preferentially occurs at uracil residues in ssRNA regions and depends on conserved catalytic residues in the two HEPN domains. Mutation of these residues yields a catalytically inactive RNA-binding protein. The secondary structure of the crRNA direct repeat (DR) stem is required for LshC2c2 activity, and mutations in the 3′ region of the DR eliminate cleavage activity. Targeting is also sensitive to multiple or consecutive mismatches in the spacer:protospacer duplex. C2c2 targeting of RFP mRNA in vivo results in reduced fluorescence. The knockdown of the RFP mRNA by C2c2 slowed *E. coli* growth, and in agreement with this finding, in vitro cleavage of the target RNA results in “collateral,” nonspecific cleavage of other RNAs present in the reaction mix.

**CONCLUSION:** LshC2c2 is a RNA-guided RNase which requires the activity of its two HEPN domains, suggesting previously unidentified mechanisms of RNA targeting and degradation by CRISPR systems. Promiscuous RNase activity of C2c2 after activation by the target slows bacterial growth and suggests that C2c2 could protect bacteria from virus spread via programmed cell death and dormancy induction. A single-effector RNA targeting system has the potential to serve as a general chassis for molecular tools for visualizing, degrading, or binding RNA in a programmable, multiplexed fashion.

The list of author affiliations is available in the full article online.

*Corresponding author. Email: zhang@broadinstitute.org (F.Z.); koonin@ncbi.nlm.nih.gov (E.V.K.)

Cite this article as O. O. Abudayyeh et al., Science 353, aaf5573 (2016). DOI: 10.1126/science.aaf5573

---

C2c2 is an RNA-guided RNase that provides protection against RNA phage. CRISPR-C2c2 from *L. shahii* can be reconstituted in *E. coli* to mediate RNA-guided interference of the RNA phage MS2. Biochemical characterization of C2c2 reveals crRNA-guided RNA cleavage facilitated by the two HEPN nuclease domains. Binding of the target RNA by C2c2-crRNA also activates a nonspecific RNase activity, which may lead to promiscuous cleavage of RNAs without complementarity to the crRNA guide sequence.
C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector

Omar O. Abudayeh,1,2,3,4 Jonathan S. Gootenberg,2,3,4,5 Silvana Konermann,2,3,4,6 Julia Young,2,3,4 Tan M. Slaimaker,2,3,4 David B. T. Cox,1,2,3,4,6 Sergey Shmakov,7,8 Kira S. Makarova,8 Ekaterina Semenova,9 Leonid Minakhin,9 Konstantin Severinov,7,9,10 Aviv Regev,3,6 Eric S. Lander,2,5,6 Eugene V. Koonin,8 Feng Zhang2,3,4,6

The clustered regularly interspaced short palindromic repeat (CRISPR) CRISPR-associated genes (Cas) adaptive immune system defends microbes against foreign genetic elements via DNA or RNA-DNA interference. We characterize the class 2 type VI CRISPR-Cas effector C2c2 and demonstrate its RNA-guided ribonuclease function. C2c2 from the bacterium *Leptotrichia shahii* provides interference against RNA phage. In vitro biochemical analysis shows that C2c2 is guided by a single CRISPR RNA and can be programmed to cleave single-stranded RNA targets carrying complementary protospacers. In bacteria, C2c2 can be programmed to knock down specific mRNAs. Cleavage is mediated by catalytic residues in the two conserved Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains, mutations of which generate catalytically inactive RNA-binding proteins. These results broaden our understanding of CRISPR-Cas systems and suggest that C2c2 can be used to develop new RNA-targeting tools.

Almost all archaea and about half of bacteria possess clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated genes (CRISPR-Cas) adaptive immune systems (1, 2), which protect microbes from viruses and other invading DNA through three steps: (i) adaptation insertion of foreign nucleic acid segments (spacers) into the CRISPR array in between pairs of direct repeats (DRs); (ii) transcription and processing of the CRISPR array to produce mature CRISPR RNAs (crRNAs); and (iii) interference, by which Cas enzymes are guided by the crRNAs to target and cleave cognate sequences in the respective invader genomes (3-5). All CRISPR-Cas systems characterized to date follow these three steps, although the mechanistic implementation and proteins involved in these processes display extensive diversity.

The CRISPR-Cas systems are broadly divided into two classes on the basis of the architecture of the interference module: Class 1 systems rely on multifunctional protein complexes, whereas class 2 systems use single-effectors proteins (1). Within these two classes, types and subtypes are delineated according to the presence of distinct signature genes, protein sequence conservation, and organization of the respective genomic loci. Class 1 systems include type I, in which interference is achieved through assembly of multiple Cas proteins into the Cascade complex, and type III systems, which rely on either the Csm (type III-A/D) or Cmr (type III-B/C) effector complexes, which are distantly related to the Cascade complex (1, 6-11).

Class 2 CRISPR systems comprise type II systems, characterized by the single-component effector protein Cas9 (12-17), which contains RuvC and HNH nuclease domains, and type V systems, which use single RuvC domain-containing effectors such as Cpf1 (18), C2c1, and C2c3 (19). All functionally characterized systems, to date, have been reported to target DNA, and only the multicomponent type III-A and III-B systems additionally target RNA (7, 20-25). However, the putative class 2 type VI system is characterized by the presence of the single-effecter protein C2c2, which lacks homology to any known DNA nuclease domain but contains two higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains (19). Given that all functionally characterized HEPN domains are ribonucleases (RNases) (26), there is a possibility that C2c2 functions solely as an RNA-guided RNA-targeting CRISPR effector.

HEPN domains are also found in other Cas proteins. Csm6, a component of type III-A systems, and the homologous protein Csx1, in type III-B systems, each contain a single HEPN domain and have been biochemically characterized as single-stranded RNA (ssRNA)-specific endoribonucleases (endoRNases) (21, 27, 28). In addition, type III systems contain complexes of other Cas enzymes that bind and cleave ssRNA through acidic residues associated with RNA-recognition motif (RRM) domains. These complexes (Cas10-Csm6 in type II-A and Cmr in type III-B) carry out RNA-guided cotranscriptional cleavage of mRNA in concert with DNA target cleavage (22, 29, 30). In contrast, the roles of Csm6 and Csx1, which cleave their targets with little specificity, are less clear, although in some cases, RNA cleavage by Csm6 apparently serves as a second line of defense when DNA-targeting fails (21). Additionally, Csm6 and Csx1 have to dimerize to form a composite active site (27, 28, 31), but C2c2 contains two HEPN domains, which suggests that it functions as a monomeric endoRNase.

As is common with class 2 systems, type VI systems are simply organized. In particular, the type VI locus in *Leptotrichia shahii* contains Csa1, Csa2, C2c2, and a CRISPR array, which is expressed and processed into mature crRNAs (19). In all CRISPR-Cas systems characterized to date, Csa1 and Csa2 are exclusively involved in spacer acquisition (32-37), which suggests that C2c2 is the sole effector protein that uses a crRNA guide to achieve interference, likely targeting RNA.

**Reconstitution of the L. shahii C2c2 locus in Escherichia coli confers RNA-guided immunity**

We explored whether LshC2c2 could confer immunity to MS2 (25), a lytic ssRNA phage without DNA intermediates in its life cycle that infects *E. coli*. We constructed a low-copy plasmid carrying the entire LshC2c2 locus (pLshC2c2) so as to allow for heterologous reconstitution in *E. coli* (fig. S1A). Because expressed mature crRNAs from the LshC2c2 locus have a maximum spacer length of 28 nucleotides (nt) (fig. S1A) (19), we tiled all possible 28-nt target sites in the MS2 phage genome (fig. 1A). This resulted in a library of 3473 spacer sequences (along with 588 non-targeting guides designed to have a Levenshtein distance of 8 with respect to the MS2 and *E. coli* genomes), which we inserted between pLshC2c2 DRs. After transformation of this construct into *E. coli*, we infected cells with varying dilutions of MS2 (10^{-1}, 10^{-2}, and 10^{-3}) and analyzed surviving cells to determine the spacer sequences carried by cells that survived the infection. Cells carrying spacers that confer robust interference against MS2 are expected to proliferate faster than those that lack such sequences. After growth for 16 hours, we identified a number of spacers that were consistently enriched across three independent infection replicates in both the 10^{-1} and 10^{-3} dilution conditions, suggesting that they enabled interference against MS2. Specifically, 152 and 144 spacers showed >0.8 log{-fold} enrichment in all three replicates for the 10^{-1} and 10^{-3} dilutions.

*These authors contributed equally to this work.†Corresponding author. Email: zhang@broadinstitute.org (F.Z.); koonin@ncbi.nlm.nih.gov (E.V.K.)

1Department of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 2Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. 3MGovern Institute for Brain Research at MIT, Cambridge, MA 02139, USA. 4Departments of Brain and Cognitive Science and Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 5Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. 6Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 7Skolkovo Institute of Science and Technology, Skolkovo, 143025, Russia. 8National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. 9Kolosov Institute for Microbiology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA. 10Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, 123182, Russia. 11These authors contributed equally to this work. ** Corresponding author. Email: zhang@broadinstitute.org (F.Z.); koonin@ncbi.nlm.nih.gov (E.V.K.)
10−3 phage dilutions, respectively; of these two groups of top enriched spacers, 75 are shared (Fig. 1B; fig. S2, A to G; and table S1). Additionally, no nontargeting guides were found to be consistently enriched among the three 10−1, 10−3, or 10−5 phage replicates (fig. S2, D and G). We also analyzed the flanking regions of protospacers on the MS2 genome corresponding to the enriched spacers and found that spacers with a G immediately flanking the 3′ end of the protospacer were less fit relative to all other nucleotides at this position (A, U, or C), suggesting that the 3′ protospacer-flanking site (PFS) affects the efficacy of C2c2-mediated targeting (Fig. 1C and figs. S2, E and F, and S3). Although the PFS is adjacent to the protospacer target, we chose not to use the commonly used protospacer-adjoining motif (PAM) nomenclature because it has come to connote a sequence used in self versus nonself differentiation (38), which is irrelevant in a RNA-targeting system. It is worth noting that the avoidance of G by C2c2 echoes the absence of PAMs observed for other RNA-targeting CRISPR systems and effector proteins (20, 22, 24, 25, 39, 40).

That only ~5% of crRNAs are enriched may reflect other factors influencing interference activity, such as accessibility of the target site that might be affected by RNA-binding proteins or secondary structure. In agreement with this hypothesis, the enriched spacers tend to cluster into regions of strong interference, where they are closer to each other than one would expect by random chance (fig. S3, F and G).

To validate the interference of the enriched spacers, we individually cloned four top-enriched spacers into pLShC2c2 CRISPR arrays and observed a 3- to 4-log₁₀ reduction in plaque formation, which is consistent with the level of enrichment observed in the screen (Fig. 1B and fig. S4). We cloned 16 guides targeting distinct regions of the MS2 mat gene (four guides per possible single-nucleotide PFS). All 16 crRNAs mediated MS2 interference, although higher levels of resistance were observed for the C, A, and U PFS-targeting guides (Fig. 1, D and E, and fig. S5), indicating that C2c2 can be effectively re-targeted in a crRNA-dependent fashion to sites within the MS2 genome.

To further validate the observed PFS preference with an alternate approach, we designed a protospacer site in the pUC19 plasmid at the 5′ end of the phage-treated conditions and control screen (no phage) biological replicates (n = 3). The box extends from the first to third quartile, with whiskers denoting 1.5 times the interquartile range. The mean is indicated by the red horizontal bar. The 10−3 and 10−5 phage dilution distributions are significantly different from each of the control replicates [***P < 0.0001 by means of analysis of variance (ANOVA) with multiple hypothesis correction]. (C) Sequence logo generated from sequences flanking the 3′ end of protospacers corresponding to enriched spacers in the 10−3 phage dilution condition, revealing the presence of a 3′ H PFS (not G). (D) Plaque assay used to validate the functional importance of the H PFS in MS2 interference. All protospacers flanked by non-G PFSs exhibited robust phage interference. Spacers were designed to target the MS2 mat gene, and their sequences are shown above the plate images; the spacer used in the nontargeting control is not complementary to any sequence in either the E. coli or MS2 genome. Phage spots were applied as series of half-log dilutions. (E) Quantitation of MS2 plaque assay validating the H (non-G) PFS preference. Four MS2-targeting spacers were designed for each PFS. Each point on the scatter plot represents the average of three biological replicates and corresponds to a single spacer. Bars indicate the mean of four spacers for each PFS and standard error (SEM).

C2c2 is a single-effector endoRNase mediating ssRNA cleavage with a single crRNA guide

We purified the LshC2c2 protein (fig. S7) and assayed its ability to cleave an in vitro–transcribed 173-nt ssRNA target (Fig. 2A and fig. S8) containing

---

**Fig. 1. Heterologous expression of the L. shahii C2c2 locus mediates robust interference of RNA phage in E. coli.** (A) Schematic for the MS2 bacteriophage interference screen. A library consisting of spacers targeting all possible sequences in the MS2 RNA genome was cloned into the LshC2c2 CRISPR array. Cells transformed with the MS2-targeting spacer library were then treated with phage and plated, and surviving cells were harvested. The frequency of spacers was compared with an untreated control (no phage), and enriched spacers from the phage-treated condition and control screen (no phage) biological replicates (n = 3). The box extends from the first to third quartile, with whiskers denoting 1.5 times the interquartile range. The mean is indicated by the red horizontal bar. The 10−3 and 10−5 phage dilution distributions are significantly different from each of the control replicates [***P < 0.0001 by means of analysis of variance (ANOVA) with multiple hypothesis correction]. (C) Sequence logo generated from sequences flanking the 3′ end of protospacers corresponding to enriched spacers in the 10−3 phage dilution condition, revealing the presence of a 3′ H PFS (not G). (D) Plaque assay used to validate the functional importance of the H PFS in MS2 interference. All protospacers flanked by non-G PFSs exhibited robust phage interference. Spacers were designed to target the MS2 mat gene, and their sequences are shown above the plate images; the spacer used in the nontargeting control is not complementary to any sequence in either the E. coli or MS2 genome. Phage spots were applied as series of half-log dilutions. (E) Quantitation of MS2 plaque assay validating the H (non-G) PFS preference. Four MS2-targeting spacers were designed for each PFS. Each point on the scatter plot represents the average of three biological replicates and corresponds to a single spacer. Bars indicate the mean of four spacers for each PFS and standard error (SEM).
Fig. 2. LshC2c2 and crRNA mediate RNA-guided ssRNA cleavage. (A) Schematic of the ssRNA substrate being targeted by the crRNA. The protosporer region is highlighted in blue, and the PFS is indicated by the magenta bar. (B) A denaturing gel demonstrating crRNA-mediated ssRNA cleavage by LshC2c2 after 1 hour of incubation. The ssRNA target is either 5' labeled with IRDye 800 or 3' labeled with Cy5. Cleavage requires the presence of the crRNA and is abolished by addition of EDTA. Four cleavage sites are observed. Reported band lengths are matched from RNA sequencing. (C) A denaturing gel demonstrating the requirement for an H PFS (not G) after 3 hours of incubation. Four ssRNA substrates that are identical except for the PFS (indicated by the magenta “X” in the schematic) were used for the in vitro cleavage assay (fig. S11A). Together, these results indicate that C2c2 cleaves specific ssRNA sites directed by the target complementarity encoded in the crRNA, with a H PFS preference.

C2c2 cleavage depends on local target sequence and secondary structure

Given that C2c2 did not efficiently cleave dsRNA substrates and that ssRNA can form complex secondary structures, we reasoned that cleavage by C2c2 might be affected by secondary structure of the ssRNA target. Indeed, after tiling ssRNA substrates and that ssRNA can form complex secondary structures, we reasoned that cleavage by C2c2 might be affected by secondary structure of the ssRNA target. Therefore, we designed five crRNAs complementary to each of these crRNAs. As expected, we observed less cleavage activity for G PFS target (Fig. 2C). Additionally, we designed five crRNAs for each target (20 total) across the ssRNA target 1 and evaluated cleavage activity for LshC2c2 paired with each of these crRNAs. As expected, we observed less cleavage activity for G PFS–targeting crRNAs as compared with other crRNAs tested (Fig. 2D).

We then generated a dsDNA plasmid library with protosporer 14 flanked by seven random nucleotides so as to account for any PFS preference. When incubated with LshC2c2 protein and a crRNA complementary to protosporer 14, no cleavage of the dsDNA plasmid library was observed (fig. S10C). We also did not observe cleavage when targeting a ssRNA version of ssRNA target 1 (fig. S10D). To rule out cotranscriptional DNA cleavage, which has been observed in type III CRISPR-Cas systems (22), we recapitulated the E. coli RNA polymerase cotranscriptional cleavage assay (fig. S1A) (22), expressing ssRNA target 1 from a DNA substrate. This assay involving purified LshC2c2 and crRNA targeting ssRNA target 1 did not show any DNA cleavage (fig. S1B). Together, these results indicate that C2c2 cleaves specific ssRNA sites directed by the target complementarity encoded in the crRNA, with a H PFS preference.
flanking identical 28-nt protospacers resulted in three distinct patterns of cleavage (Fig. 3A). RNA-sequencing of the cleavage products for the three targets revealed that cleavage sites mainly localized to uracil-rich regions of ssRNA or ssRNA-dsRNA junctions within the in silico-predicted cofolds of the target sequence with the crRNA (Fig. 3, B and C, and Fig. S12, A to D). To test whether the LshC2c2-crRNA complex prefers cleavage at uracils, we analyzed the cleavage efficiencies of homopolymeric RNA targets (a 28-nt protospacer extended with 120 As or Us regularly flanked by different sequences. Despite identical protospacers, different flanking sequences resulted in different cleavage patterns. Reported band lengths are matched from RNA sequencing. (B) The cleavage sites of nonhomopolymer ssRNA target 1 were mapped with RNA-sequencing of the cleavage products. The frequency of cleavage at each base is colored according to the z-score and shown on the predicted crRNA-ssRNA cofold secondary structure. Fragments used to generate the frequency analysis contained the complete 5' end. The 5' and 3' end of the ssRNA target are indicated by blue and red outlines, on the ssRNA and secondary structure, respectively. The 5' and 3' end of the spacer (outlined in yellow) is indicated by the blue and orange residues highlighted, respectively. The crRNA nucleotides are highlighted in orange. (C) Plot of the frequencies of cleavage sites for each position of ssRNA target 1 for all reads that begin at the 5' end. The protospacer is indicated by the blue highlighted region. (D) Schematic of a modified ssRNA 1 target showing sites (red) of single U-to-A flips (left). Denaturing gel showing C2c2-crRNA-mediated cleavage of each of these single nucleotide variants after 3 hours of incubation (right). Reported band lengths are matched from RNA sequencing.

Cleavage was occurring at uracil residues, we mutated single-uracil residues in ssRNA 1 that showed cleavage in the RNA-sequencing (Fig. 3B) to adenines. This experiment showed that by mutating each uracil residue, we could modulate the presence of a single cleavage band, which is consistent with LshC2c2 cleaving at uracil residues in ssRNA regions (Fig. 3D).

The HEPN domains of C2c2 mediate RNA-guided ssRNA-cleavage
Bioinformatic analysis of C2c2 has suggested that the HEPN domains are likely to be responsible for the observed catalytic activity (19). Each of the two HEPN domains of C2c2 contains a dyad of conserved arginine and histidine residues (Fig. 4A), which is in agreement with the catalytic mechanism of the HEPN endoRNAse (26–28). We mutated each of these putative catalytic residues separately to alanine (R597A, H602A, R1278A, H1283A) in the LshC2c2 locus plasmids and assayed for MS2 interference. None of the four mutant plasmids were able to protect E. coli from phage infection (Fig. 4B and fig. S13). (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, R597A indicates that arginine at position 597 was replaced by alanine.)

We purified the four single-point mutant proteins and assayed their ability to cleave 5' end-labeled ssRNA target 1 (Fig. 4C). In agreement with our in vivo results, all four mutations abolished cleavage activity. The inability of either of the two wild-type HEPN domains to compensate for inactivation of the other implies cooperation between the two domains. These results agree with observations that several bacterial and eukaryotic single-HEPN proteins function as dimers (27, 28, 41).
Catalytically inactive variants of Cas9 retain target DNA binding, allowing for the creation of programmable DNA-binding proteins (22, 23). Electrophoretic mobility shift assays (EMSA) on both the wild-type (Fig. 4D) and R1278A mutant LshC2c2 (Fig. 4E) in complex with crRNA showed the wild-type LshC2c2 complex binding strongly [dissociation constant \( (K_d) \approx 46 \text{nM} \)] (Fig. S14A) and specifically to 5’ end-labeled ssRNA target 10 but not to the 5’ end-labeled nontarget ssRNA (the reverse complement of ssRNA target 10). The R1278A mutant C2c2 complex showed even stronger \( (K_d) \approx 7 \text{nM} \) (Fig. S14B) specific binding, indicating that this HEPN mutation results in a catalytically inactive, RNA-programmable, RNA-binding protein. The LshC2c2 protein or crRNA alone showed reduced levels of target affinity, as expected (fig. S14, C to E). Additionally, no specific binding of LshC2c2-crRNA complex to ssDNA was observed (fig. S15).

These results demonstrate that C2c2 cleaves RNA via a catalytic mechanism distinct from other known CRISPR-associated RNases. In particular, the type III Csm and Cmr multiprotein complexes rely on acidic residues of RRM domains for catalysis, whereas C2c2 achieves RNA cleavage through the conserved basic residues of its two HEPN domains.

**Sequence and structural requirements of C2c2 crRNA**

Similar to the type V-A (Cpf1) systems (18), the LshC2c2 crRNA contains a single stem loop in the DR, suggesting that the secondary structure of the crRNA could facilitate interaction with LshC2c2. We thus investigated the length requirements of the spacer sequence for ssRNA cleavage and found that LshC2c2 requires spacers of at least 22 nt length to efficiently cleave ssRNA target 1 (fig. S16A). The stem-loop structure of the crRNA is also critical for ssRNA cleavage because DR truncations that disturbed the stem loop abrogated target cleavage (fig. S16B). Thus, a DR longer than 24 nt is required to maintain the stem loop necessary for LshC2c2 to mediate ssRNA cleavage.

Single-base-pair inversions in the stem that preserved the stem structure did not affect the activity of the LshC2c2 complex. In contrast, inverting all four G-C pairs in the stem eliminated the cleavage, despite maintaining the duplex structure (fig. S17A). Other perturbations, such as those that introduced kinks and reduced or increased base-pairing in the stem, also eliminated or drastically suppressed cleavage. This suggests that the crRNA stem length is important for complex formation and activity (fig. S17A). We also found that loop deletions eliminated cleavage, whereas insertions and substitutions mostly maintained some level of cleavage activity (fig. S17B). In contrast, nearly all substitutions or deletions in the region 3’ to the DR prevented cleavage by LshC2c2 (fig. S18). Together, these results demonstrate that LshC2c2 recognizes structural characteristics of its cognate crRNA but is amenable to loop insertions and most tested base substitutions outside of the 3’ DR region.

---

**Fig. 4. The two HEPN domains of C2c2 are necessary for crRNA-guided ssRNA cleavage but not for binding.** (A) Schematic of the LshC2c2 locus and the domain organization of the LshC2c2 protein, showing conserved residues in HEPN domains (dark blue). (B) Denaturing gel showing conserved residues of the HEPN motif, indicated as catalytic residues in (A), are necessary for crRNA-guided ssRNA target 1 cleavage after 3 hours of incubation. Reported band lengths are matched from RNA sequencing. (C) Electrophoretic mobility shift assay (EMSA) evaluating affinity of the wild-type LshC2c2-crRNA complex against on-target and nontarget ssRNA substrate. The nontarget ssRNA substrate is the reverse-complement of the targeted ssRNA 10. EDTA is supplemented to reaction condition in order to reduce any cleavage activity. (D) Electrophoretic mobility shift assay with LshC2c2(R1278A)-crRNA complex against on-target and nontarget ssRNA [same substrate sequences as in (D)].

---

**Sequence and structural requirements of C2c2 crRNA**

Similar to the type V-A (Cpf1) systems (18), the LshC2c2 crRNA contains a single stem loop in the DR, suggesting that the secondary structure of the crRNA could facilitate interaction with LshC2c2. We thus investigated the length requirements of the spacer sequence for ssRNA cleavage and found that LshC2c2 requires spacers of at least 22 nt length to efficiently cleave ssRNA target 1 (fig. S16A). The stem-loop structure of the crRNA is also critical for ssRNA cleavage because DR truncations that disturbed the stem loop abrogated target cleavage (fig. S16B). Thus, a DR longer than 24 nt is required to maintain the stem loop necessary for LshC2c2 to mediate ssRNA cleavage.

Single-base-pair inversions in the stem that preserved the stem structure did not affect the activity of the LshC2c2 complex. In contrast, inverting all four G-C pairs in the stem eliminated the cleavage, despite maintaining the duplex structure (fig. S17A). Other perturbations, such as those that introduced kinks and reduced or increased base-pairing in the stem, also eliminated or drastically suppressed cleavage. This suggests that the crRNA stem length is important for complex formation and activity (fig. S17A). We also found that loop deletions eliminated cleavage, whereas insertions and substitutions mostly maintained some level of cleavage activity (fig. S17B). In contrast, nearly all substitutions or deletions in the region 3’ to the DR prevented cleavage by LshC2c2 (fig. S18). Together, these results demonstrate that LshC2c2 recognizes structural characteristics of its cognate crRNA but is amenable to loop insertions and most tested base substitutions outside of the 3’ DR region.
These results have implications for the future application of C2c2-based tools that require guide engineering for recruitment of effectors or modulation of activity (42–44).

C2c2 cleavage is sensitive to double mismatches in the crRNA-target duplex

We tested the sensitivity of the LshC2c2 system to single mismatches between the crRNA guide and target RNA by mutating single bases across the spacer to the respective complementary bases (for example, A to U). We then quantitated plaque formation with these mismatched spacers in the center of the crRNA-target duplex. We tested the sensitivity of the LshC2c2 system to single mismatches by mutating single bases across the spacer to the respective complementary bases (for example, A to U). We then quantitated plaque formation with these mismatched spacers in the center of the crRNA-target duplex.

C2c2 can be reprogrammed to mediate specific mRNA knockdown in vivo

Given the ability of C2c2 to cleave target ssRNA in a crRNA sequence–specific manner, we tested whether LshC2c2 could be reprogrammed to degrade selected nonphage ssRNA targets, and particularly mRNAs, in vivo. We cotransformed E. coli with a plasmid encoding LshC2c2 and a crRNA targeting the mRNA of red fluorescent protein (RFP) as well as a compatible plasmid expressing RFP (Fig. 5A). For optical density (OD)–matched samples, we observed an ~20 to 92% decrease in RFP-positive cells for crRNAs containing consecutive triple mismatches in the seed region (fig. S21B).
targeting protospacers flanked by C, A, or U PFSs (Fig. 5, B and C). As a control, we tested crRNAs containing reverse complements (targeting the dsDNA plasmid) of the top performing RFP mRNA-targeting spacers. As expected, we observed no decrease in RFP fluorescence by these crRNAs (Fig. 5B). We also confirmed that mutation of the catalytic arginine residues in either HEPN domain to alanine precluded RFP knockdown (Fig. S22). Thus, C2c2 is capable of general retargeting to arbitrary ssRNA substrates, governed exclusively by predictable nucleic-acid interactions.

When we examined the growth of cells carrying the RFP-targeting spacer with the greatest level of RFP knockdown, we noted that the growth rate of these bacteria was substantially reduced (Fig. 5A, spacer 7). We investigated whether the effect on growth was mediated by the RFP mRNA-targeting activity of LshC2c2 by introducing an inducible-RFP plasmid and an RFP-targeting dsDNA plasmid) of the top performing RFP mRNA-targeting spacers. As expected, we observed no decrease in RFP fluorescence by these crRNAs (Fig. 5B). We also confirmed that mutation of the catalytic arginine residues in either HEPN domain to alanine precluded RFP knockdown (Fig. S22). Thus, C2c2 is capable of general retargeting to arbitrary ssRNA substrates, governed exclusively by predictable nucleic-acid interactions.

We investigated the collateral cleavage activity of LshC2c2-crRNA in vitro cleavage reactions that included, in addition to LshC2c2 protein, crRNA and its target ssRNA, collateral RNA, which has no complementarity to the crRNA. (Fig. S23B). The analysis of the depleted and enriched PFS sequences identified a H PFS only in the assay with the transcribed sequence but no discernible motif in the nontranscribed sequence (Fig. S24, B and C).

These results suggest a HEPN-dependent mechanism by which C2c2 in a complex with crRNA is activated upon binding to target RNA and subsequently cleaves nonspecifically other available ssRNA targets. Such promiscuous RNA cleavage could cause cellular toxicity, resulting in the observed growth rate inhibition. These findings imply that in addition to their likely role in direct suppression of RNA viruses, type VI CRISPR-Cas systems could function as mediators of a distinct variety of PCD or dormancy induction that is specifically triggered by cognate invader genomes (Fig. 7). Under this scenario, dormancy would slow the infection and supply additional time for adaptive immunity. Such a mechanism agrees with the previously proposed coupling of adaptive immunity and PCD during the CRISPR-Cas defensive response (47).

### Conclusions

The class 2 type VI effector protein C2c2 is a RNA-guided RNase that can be efficiently programmed to degrade any ssRNA by specifying a 28-nt sequence on the crRNA (Fig. S10). C2c2 cleaves RNA through conserved basic residues within its two HEPN domains, in contrast to the catalytic mechanisms of other known RNases found in CRISPR-Cas systems (25, 48). Alanine substitution of any of the four predicted HEPN domain catalytic residues converted C2c2 into an inactive programmable RNA-binding protein (dC2c2, analogous to dCas9). Many different spacer sequences work well in our assays, although further screening will likely define properties and rules governing optimal function.

These results suggest a broad range of biotechnology applications and research questions (49–51). For example, the ability of dC2c2 to bind to specified sequences could be used to (i) bring effector modules to specific transcripts in order to modulate their function or translation, which could be used for large-scale screening, construction of synthetic regulatory circuits, and other purposes; (ii) fluorescently tag specific RNAs in order to visualize their trafficking and/or localization; (iii) alter RNA localization through domains with affinity for specific subcellular compartments; and (iv) capture specific transcripts (through direct pull-down of dC2c2) in order to enrich for proximal molecular partners, including RNAs and proteins.

Active C2c2 also has many potential applications, such as targeting a specific transcript for destruction, as performed here with RFP.

---

**Fig. 6. crRNA-guided ssRNA cleavage activates nonspecific RNase activity of LshC2c2.** (A) Schematic of the biochemical assay used to detect crRNA-binding–activated nonspecific RNase activity on non-crRNA–targeted collateral RNA molecules. The reaction consists of C2c2 protein, unlabeled crRNA, unlabeled target ssRNA, and a second ssRNA with 3′ fluorescent labeling and is incubated for 3 hours. C2c2-crRNA mediates cleavage of the unlabeled target ssRNA as well as the 3′ end–labeled collateral RNA, which has no complementarity to the crRNA. (B) Denaturing gel showing nonspecific RNase activity against nontargeted ssRNA substrates in the presence of target RNA after 3 hours of incubation. The nontargeted ssRNA substrate is not cleaved in the absence of the crRNA-targeted ssRNA substrate.
In addition, C2c2, once primed by the cognate target, can cleave other (noncomplementary) RNA molecules in vitro and inhibit cell growth in vivo. Biologically, this promiscuous RNase activity might reflect a PCD/dormancy-based protection mechanism of the type VI CRISPR-Cas systems (Fig. 7). Technologically, it might be used to trigger PCD or dormancy in specific cells, such as cancer cells expressing a particular transcript, neurons of a given class, or cells infected by a specific pathogen.

Further experimental study is required to elucidate the mechanisms by which the C2c2 system acquires spacers and the classes of pathogens against which it protects bacteria. The presence of the conserved CRISPR adaptation module consisting of typical Cas1 and Cas2 proteins in the LshC2c2 locus suggests that it is capable of spacer acquisition. Although C2c2 systems lack reverse transcriptases, which mediate acquisition of RNA spacers in some type III systems (52), it is possible that additional host or viral factors could support RNA spacer acquisition. Additionally or alternatively, type VI systems could acquire DNA spacers similar to other CRISPR-Cas variants but then target transcripts of the respective DNA genomes, eliciting PCD and abortive infection (Fig. 7).

The CRISPR-C2c2 system represents a distinct evolutionary path among class 2 CRISPR-Cas systems. It is likely that other, broadly analogous class 2 RNA-targeting immune systems exist, and further characterization of the diverse members of class 2 systems will provide a deeper understanding of bacterial immunity and provide a rich starting point for the development of programmable molecular tools for in vivo RNA manipulation.

**Materials and methods**

Expanded materials and methods, including computational analysis, can be found in supplementary materials and methods.

**Bacterial phage interference**

The C2c2 CRISPR locus was amplified from DNA from *Leptotrichia shahii* DSM 19757 (ATCC, Manassas, VA) and cloned for heterologous expression in *E. coli*. For screens, a library of all possible spacers targeting the MS2 genome were cloned into the spacer array; for individual spacers, single specific spacers were cloned into the array. Interference screens were performed in liquid culture and plated; surviving colonies were harvested for DNA and spacer representation was determined by next-generation sequencing. Individual spacers were tested by spotting on top agar.

**β-lactamase and transcribed/non-transcribed PFS preference screens**

Sequences with randomized nucleotides adjacent to protospacer 1 were cloned into pUC19 in corresponding regions. Libraries were screened by co-transformation with LshC2c2 locus plasmid or pACYC184 plasmid control, harvesting of the surviving colonies, and next-generation sequencing of the resulting regions.

**RFP-targeting assay**

Cells containing an RFP expressing plasmid were transformed with an LshC2c2 locus plasmid with corresponding spacers, grown overnight, and analyzed for RFP fluorescence by flow cytometry. The growth effects of LshC2c2 activity were quantified by titrating inducible RFP levels with dilutions of anhydrotetracycline inducer and then measuring OD600.

**in vitro nuclease and electrophoretic mobility shift assays**

LshC2c2 protein and HEPN mutants were purified for use in vitro reactions; RNA were synthesized via in vitro transcription. For nuclease assays, protein was co-incubated with crRNA and either 3′ or 5′-labeled targets and analyzed via denaturing gel electrophoresis and imaging or by next-generation sequencing. For electrophoretic mobility shift assays, protein and nucleic
acid were co-incubated and then resolved by gel electrophoresis and imaging.

REFERENCES AND NOTES

ACKNOWLEDGMENTS
We thank P. Boutz, J. Doench, P. Sharp, and B. Zetsche for helpful discussions and insights; R. Beliveau for overall research support; J. Francis and D. O’Connell for generous MS2Qc instrument access; D. Daniels and C. Garvie for providing bacterial incubation space for protein purification; R. Macrae for critical reading of the manuscript; and the entire Zhang laboratory for support and advice. We thank N. Ranu for generously providing plasmid pRRP and D. Daniels for providing 6-His-NTA, O.D.A is supported by a Paul and Daisy Soros Fellowship, a Friends of the McGovern Institute Fellowship, and the Poitras Center for Addictive Disorders. J.S.G. is supported by a U.S. Department of Energy (DOE) Computational Science Graduate Fellowship. S.S. is supported by the graduate program of Skoltech Data-Intensive Biomedicine and Biotechnology Center for Research, Education, and Innovation. I.M.S. is supported by the Simons Center for the Social Brain. B.D.T.C. is supported by award 2T32GM007753 from the National Institute of General Medical Sciences. K.S.M., E.V.K., and, in part, S.S. are supported by the intramural program of the U.S. Department of Health and Human Services (to the National Library of Medicine). K.S. is supported by NIH grant GM10407. Russian Science Foundation grant 14-14-00988, and Skatchek. F.Z. is a New York Stem Cell Foundation Robertson Investigator. F.Z. is supported by the NIH through the National Institute of Mental Health (1DP1-MH100706-01 and 1R01-HD001049). NSF, the New York Stem Cell, Simons, Paul G. Allen Family, and Vallejo Foundations; and J. Poitras, R. Metcalf, and D. Cheng, A.R is a member of the scientific advisory board for Synos Pharmaceuticals and Thermo Fisher, D.O.A., J.S.G., J.I., E.V.K., S.K.E., L.S.M., L.M., E.S., K.S., S.S., and F.Z. are inventors on provisional patent application 62/161,675 applied for by the Broad Institute. MIT, Harvard, NIH, Skatchek, and Rutgers that covers the C2c2 proteins described in this paper. Deep sequencing data are available at Sequence Read Archive under BioProject accession no. PRJNA238880. The authors plan to make the reagents widely available to the academic community through Addgene subject to a materials transfer agreement and to provide software tools via the Zhang laboratory website (www.genome-engineering.org) and GitHub (https://github.com/fengshanglab).

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/353/6295/aafl5573/suppl/DC1
Materials and Methods
Tables S1 to S5
Figs. S1 to S24
References (S1–S5)
2 March 2016; accepted 25 May 2016
Published online: 2 June 2016
10.1126/science.aafl5573

Downloaded from http://science.sciencemag.org/ on August 14, 2016
Editor's Summary

C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector
Science 353 (6299), doi: 10.1126/science.aaf5573 originally published online June 2, 2016

This copy is for your personal, non-commercial use only.

Article Tools
Visit the online version of this article to access the personalization and article tools:
http://science.sciencemag.org/content/353/6299/aaf5573

Permissions
Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl