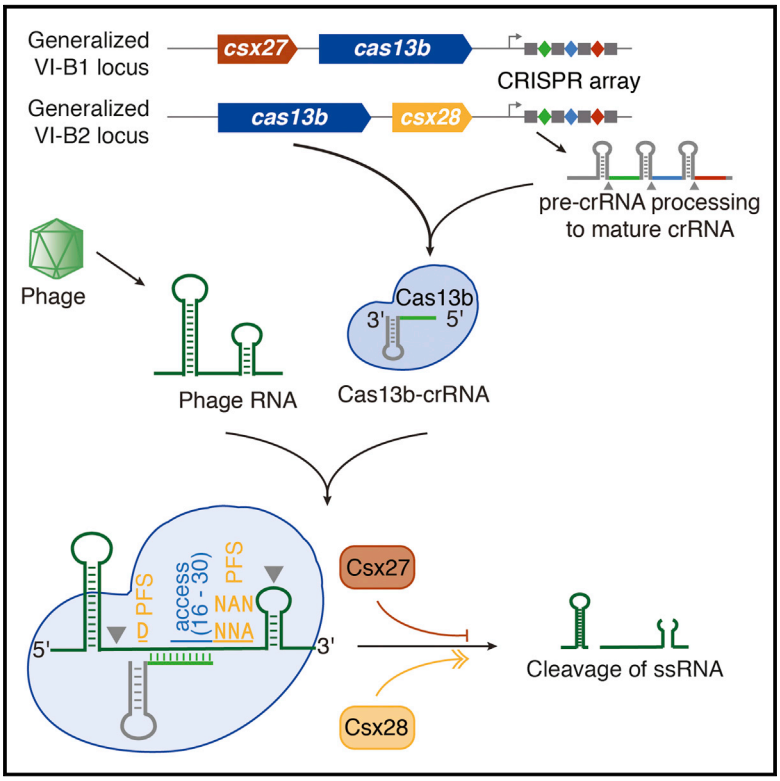
Article

Molecular Cell

Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28

Graphical Abstract



Highlights

- CRISPR-Cas13b is a class 2 type VI-B CRISPR system lacking Cas1 and Cas2
- Cas13b is a CRISPR-associated RNA-guided RNase with two crRNA variants
- Csx27 represses, whereas Csx28 enhances, Cas13b activity
- Cas13b RNA targeting is dependent on a double-sided PFS and RNA accessibility

Authors

Aaron A. Smargon, David B.T. Cox, Neena K. Pyzocha, ..., Kira S. Makarova, Eugene V. Koonin, Feng Zhang

Correspondence

zhang@broadinstitute.org

In Brief

Smargon et al. identify and characterize two class 2 type VI-B CRISPR systems lacking Cas1 and Cas2 and containing the RNA-guided RNase Cas13b, differentially regulated by Csx27 and Csx28. Through an *E. coli* essential gene screen they show that Cas13b RNA targeting is dependent on a double-sided PFS and RNA accessibility.

Smargon et al., 2017, Molecular Cell *65*, 618–630
February 16, 2017 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.molcel.2016.12.023

CellPress

Molecular Cell
Article

Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28

Aaron A. Smargon,[1,2,3,4,10] David B.T. Cox,[1,2,3,5,6,10] Neena K. Pyzocha,[1,2,3,6,10] Kaijie Zheng,[1,2,3] Ian M. Slaymaker,[1,2,3] Jonathan S. Gootenberg,[1,2,3,7] Omar A. Abudayyeh,[1,2,3,5] Patrick Essletzbichler,[1,2,3] Sergey Shmakov,[8,9] Kira S. Makarova,[9] Eugene V. Koonin,[9] and Feng Zhang[1,2,3,4,11,*]

[1]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2]McGovern Institute for Brain Research at MIT, Cambridge, MA 02139, USA
[3]Departments of Brain and Cognitive Science and Biological Engineering
[4]Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, MA 02115, USA
[6]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[8]Skolkovo Institute of Science and Technology, Skolkovo 143025, Russia
[9]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
[10]Co-first author
[11]Lead Contact
*Correspondence: zhang@broadinstitute.org
http://dx.doi.org/10.1016/j.molcel.2016.12.023

SUMMARY

CRISPR-Cas adaptive immune systems defend microbes against foreign nucleic acids via RNA-guided endonucleases. Using a computational sequence database mining approach, we identify two class 2 CRISPR-Cas systems (subtype VI-B) that lack Cas1 and Cas2 and encompass a single large effector protein, Cas13b, along with one of two previously uncharacterized associated proteins, Csx27 and Csx28. We establish that these CRISPR-Cas systems can achieve RNA interference when heterologously expressed. Through a combination of biochemical and genetic experiments, we show that Cas13b processes its own CRISPR array with short and long direct repeats, cleaves target RNA, and exhibits collateral RNase activity. Using an *E. coli* essential gene screen, we demonstrate that Cas13b has a double-sided protospacer-flanking sequence and elucidate RNA secondary structure requirements for targeting. We also find that Csx27 represses, whereas Csx28 enhances, Cas13b-mediated RNA interference. Characterization of these CRISPR systems creates opportunities to develop tools to manipulate and monitor cellular transcripts.

INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) systems are divided into two classes, class 1 systems, which utilize multiple Cas proteins and CRISPR RNA (crRNA) to form an effector complex, and the more compact class 2 systems, which employ a large, single effector with crRNA to mediate interference (Makarova et al., 2015). CRISPR-Cas systems display a wide evolutionary diversity, involving distinct protein complexes and different modes of operation, including the ability to target RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Hale et al., 2009; Jiang et al., 2016; Staals et al., 2013, 2014; Tamulaitis et al., 2014).

Computational sequence database mining for diverse CRISPR-Cas systems has been carried out by searching microbial genomic sequences for loci harboring the *cas1* gene, the most highly conserved *cas* gene involved in the adaptation phase of CRISPR immunity (Marraffini, 2015). Among other findings, this approach led to the discovery of the class 2 subtype VI-A system with its signature effector Cas13a (previously known as C2c2), which targets RNA (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Shmakov et al., 2015). Since distinct variants of functional class 1 CRISPR systems have been characterized that lack *cas1* (Makarova et al., 2015), we sought to identify class 2 CRISPR-Cas systems lacking *cas1* by modifying the computational discovery pipeline so that it is not seeded on Cas1. Here we report the characterization of a class 2 subtype, VI-B, which was discovered through this computational approach, and demonstrate that the VI-B effector, Cas13b, is an RNA-guided RNase.

RESULTS

Computational Discovery of Class 2 Subtype VI-B CRISPR Systems

We designed a computational pipeline to search specifically for putative CRISPR-Cas loci lacking Cas1 and Cas2 (Figure 1A).

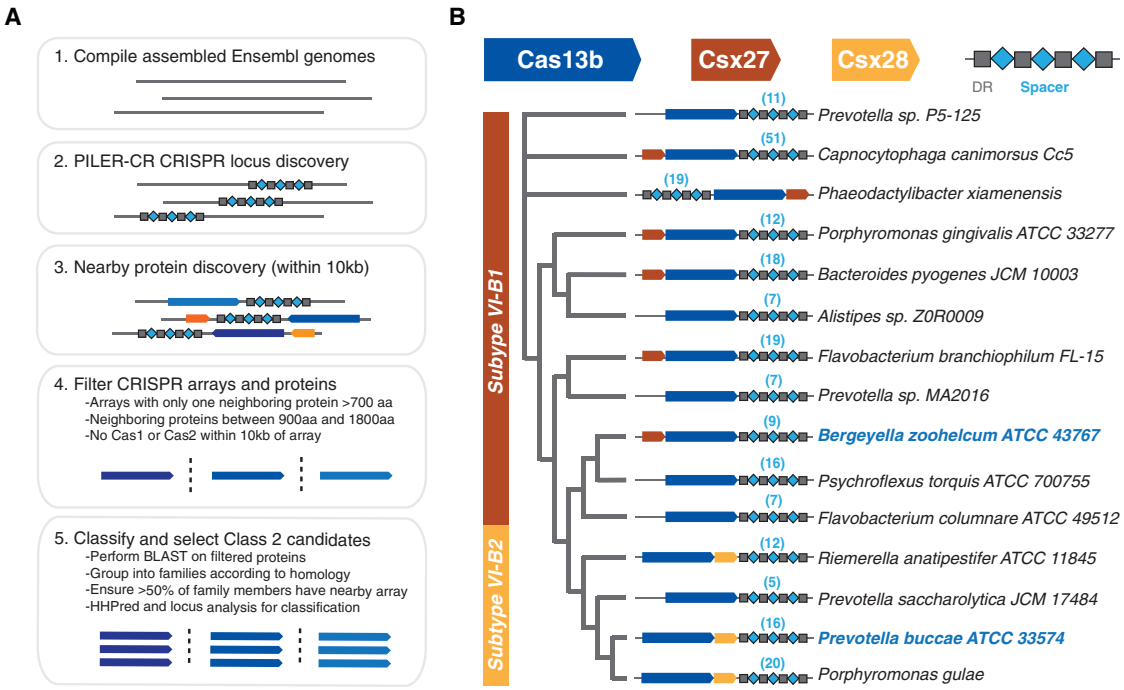


Figure 1. Discovery of Two Class 2 CRISPR-Cas Systems, Subtype VI-B1 and VI-B2, Containing Cas13b

(A) Bioinformatic pipeline to discover putative class 2 CRISPR loci lacking Cas1 and Cas2.

(B) A schematic phylogenetic tree of the subtype VI-B loci. Loci with Csx27 (brown) comprise variant VI-B1; loci with Csx28 (gold) comprise variant VI-B2. See also Figures S1–S3.

Fully assembled microbial genomes were searched for all proteins within 10 kb of CRISPR arrays (Edgar, 2007; Yates et al., 2016). The list of identified loci was further narrowed down using the following criteria: no more than one neighboring protein larger than 700 aa (to eliminate class 1 system false positives), presence of a putative single effector of size 900–1,800 aa (informed by the size distribution of previously classified class 2 effectors), and absence of *cas1* and *cas2* genes within 10 kb of the CRISPR array (STAR Methods). Candidate effectors were grouped into families according to homology (Camacho et al., 2009; Hildebrand et al., 2009; Remmert et al., 2011) and discarded if they matched previously identified CRISPR-Cas systems (Makarova et al., 2015). To focus on likely functional CRISPR loci, we limited the candidate list to families of at least ten non-redundant effectors in which the putative effector was near a CRISPR array for at least 50% of the members.

Among the candidates, we identified two genetically diverse putative class 2 CRISPR-Cas systems (105 genomic loci, 81 containing a unique entry Cas13b in the non-redundant NCBI protein database, and 71 of these 81 containing an annotated CRISPR array) represented in Gram-negative bacteria (Figure S1A). For some genera, in particular *Porphyromonas* and *Prevotella*, Cas13b proteins are encoded in several unique sequenced loci and, occasionally, in the same sequenced genome. These systems often co-occur with other CRISPR-Cas systems. Of the 81 type VI-B loci found across complete and incomplete bacterial genomes, 62 also possess at least one other CRISPR-Cas locus that includes the key adaptation endonuclease, Cas1. However, three complete genomes carrying the type VI-B locus (Flavobacterium_branchiophilum_FL_15_GCA_000253275.1, Paludibacter_propionicigenes_WB4_GCA_000183135.1, and Porphyromonas_gingivalis_AJW4_GCA_001274615.1) lack Cas1 altogether (Figure S1A).

All these loci encode a large (~1,100 aa) candidate effector protein and, in about 80% of the cases, an additional small (~200 aa) protein (Figures 1B and S1A). The putative effector proteins contain two predicted HEPN domains (Anantharaman et al., 2013) at their N and C termini (Figure S1B), similar to the domain architecture of the large effector of subtype VI-A (Cas13a) (Shmakov et al., 2015). Beyond the occurrence of two HEPN domains, however, there is no significant sequence similarity between the predicted effector and Cas13a. These systems were also identified by a generalized version of the pipeline described above as part of a comprehensive analysis of class 2 CRISPR-Cas systems and were classified into subtype VI-B, with predicted effector protein Cas13b (Shmakov et al., 2017).

CRISPR-Cas13b Loci Contain Small Accessory Proteins

The identity of the putative accessory protein correlates with the two distinct branches in the phylogenetic tree of Cas13b (Figures 1B and S1A) (Henikoff and Henikoff, 1992), indicative of the existence of two variant systems, which we denote VI-B1 (accessory protein referred to as Csx27) and VI-B2 (accessory protein referred to as Csx28). While subtype VI-B2 systems almost invariably contain *csx28*, *csx27* is less consistently represented

in VI-B1 loci. The protein sequences of Csx27 and Csx28 show no significant similarity to any previously identified Cas proteins. Both putative accessory proteins were predicted to contain one or more transmembrane segments (Figure S2A) (Möller et al., 2001). However, Csx27 of *Bergeyella zoohelcum* and Csx28 of *Prevotella buccae* tagged with RFP at either the N or the C terminus did not show membrane localization when expressed in *E. coli* (Figure S2B). In addition to the predicted hydrophobic domains, analysis of the multiple sequence alignment of Csx28 proteins indicated the presence of a divergent HEPN domain (Figure S1C).

Cas13b-Associated CRISPR Arrays Display Unique Features

In contrast to their differing putative accessory proteins, both variants of subtype VI-B systems show distinct, conserved features in the CRISPR arrays. The direct repeats in the CRISPR arrays are conserved in size, sequence, and structure, with a length of 36 nt, a poly-U stretch in the open loop region, and complementary sequences 5′-GUUG and CAAC-3′ at the ends of the repeat predicted to yield a defined secondary structure mediated by intramolecular base-pairing (Figures S3A–S3C) (Lorenz et al., 2011). Our analysis revealed 36 Cas13b spacers mapped with greater than 80% homology to unique protospacers in phage genomes. Twenty-seven of the identified Cas13b spacers targeted the coding strand of phage mRNA, while seven spacers targeted the noncoding strand and two spacers targeted regions of the phage genome without predicted transcripts. Although the composite of these imperfect mappings revealed no consensus flanking region sequence (Figure S3D) (Biswas et al., 2013), the well-conserved protospacer length of 30 nt, combined with the conserved direct repeat sequence and length, suggests that the nucleic acid targeting rules may be similar among different VI-B loci.

RNA sequencing of the total RNA from *B. zoohelcum* (subtype VI-B1) showed processing of the pre-crRNA into a 66 nt mature crRNA, with the full 30 nt 5′ spacer followed by the full 36 nt 3′ direct repeat (Figure 2A) (Heidrich et al., 2015; Li and Durbin, 2009; Shmakov et al., 2015). A longer 118 nt crRNA, distal to the 36 nt crRNAs in the CRISPR array and with a direct repeat consisting of 5′ and 3′ fragments of the 36 nt direct repeat sequence interrupted by an intervening repeat sequence, was also processed. This phenomenon was computationally predicted to occur in additional VI-B loci, such as those from *Capnocytophaga canimorsus*, *Myroides odoratimimus*, and *Riemerella anatipestifer*. Other CRISPR class 2 effectors are known to process their arrays without involvement of additional RNases (East-Seletsky et al., 2016; Zetsche et al., 2015). Similarly, we find that purified BzCas13b is capable of cleaving its CRISPR array, generating mature crRNAs with short or long direct repeats, and spacers which are not further processed beyond 30 nt, an activity which is not affected by mutation of the predicted catalytic residues of the HEPN domain (Figures 2B and S4; Table S1 and Table S2).

An *E. coli* Essential Gene Screen Reveals Targeting Rules for *BzCas13b*

To validate the expected interference activity of the VI-B system and to determine the targeting rules for the VI-B1 locus from *B. zoohelcum*, we developed an *E. coli* essential gene screen (Figure 3A). For this negative selection screen, we generated a library of 54,600 unique spacers tiled with single-nucleotide resolution over the coding region of 45 monocistronic essential genes (Baba et al., 2006; Gerdes et al., 2003), plus 60 nt into the 5′ and 3′ UTRs. We also included 1,100 randomly generated non-targeting spacers to establish baseline activity (Table S3 and Table S4). We then transformed this library with plasmids carrying *bzcas13b* (*cas13b* gene from *B. zoohelcum*) and *bzcsx27*, just *bzcas13b*, or a control empty vector. After quality-control filtering of all screened spacers, we found a statistically significant depletion of targeting spacers over non-targeting spacers, indicating that Cas13b, alone or with Csx27, can achieve nucleic acid interference (Figure 3B).

To assess the targeting rules for Cas13b, we established two spacer depletion levels: strongly depleted (top 1% of depleted spacers) and safely depleted (spacers depleted 5σ above the mean depletion of the filtered non-targeting spacers). From spacers passing the strongly depleted cutoff we derived sequence motifs qualitatively identifying a double-sided protospacer flanking sequence (PFS) (Figure 3C) (Crooks et al., 2004). Because each position in a sequence motif is assumed to be independent, we developed a more quantitative, base-dependent PFS score defined as the ratio of the number of safely depleted spacers to the number of all spacers with a given PFS, normalized across all PFS scores (Figure 3D).

The normalized PFS scores revealed a 5′ PFS of D (A, U, or G) and 3′ PFS of NAN or NNA, consistent for Cas13b with Csx27, as well as for Cas13b alone. To validate these sequence-targeting rules, we performed an orthogonal depletion screen with Cas13b alone, targeting the Kanamycin resistance gene (Figures 3E and 3F). Four classes of spacers were created: non-targeting, targeting with both 5′ and 3′ PFS rules, targeting with only the 5′ or 3′ PFS rule, and targeting with neither rule. Consistent with our findings from the *E. coli* essential gene screen, the combined 5′ and 3′ PFS spacers resulted in the highest Kanamycin sensitivity (Figures 3G and S5A; Table S5).

BzCas13b Cleaves Single-Stranded RNA and Exhibits Collateral Activity In Vitro

Based on the presence of the computationally predicted HEPN domains that function as RNases in other CRISPR-Cas systems, including VI-A and some class 1 systems (Abudayyeh et al., 2016; Kim et al., 2013; Sheppard et al., 2016; Staals et al., 2014), we anticipated that Cas13b interferes with RNA. We confirmed this by demonstrating that purified Cas13b exclusively cleaves single-stranded RNA with both direct repeat architectures (Figures 4A and S6A). We then validated the PFS targeting rules biochemically, showing that a 5′ PFS of C greatly inhibits single-stranded RNA cleavage (Figure 4B), whereas a 3′ PFS of NAN or NNA enhances this activity (Figure 4C).

Other HEPN domain-containing CRISPR-Cas RNA-targeting systems, such as Csx1 from the type III-B CRISPR-Cas systems, preferentially cleave targets containing specific single-stranded nucleotides (Sheppard et al., 2016). To determine if Cas13b exhibits such a preference, we tested an RNA substrate with a variable homopolymer loop outside of the spacer:protospacer duplex region (Figure 4D). A heteropolymer loop consisting of

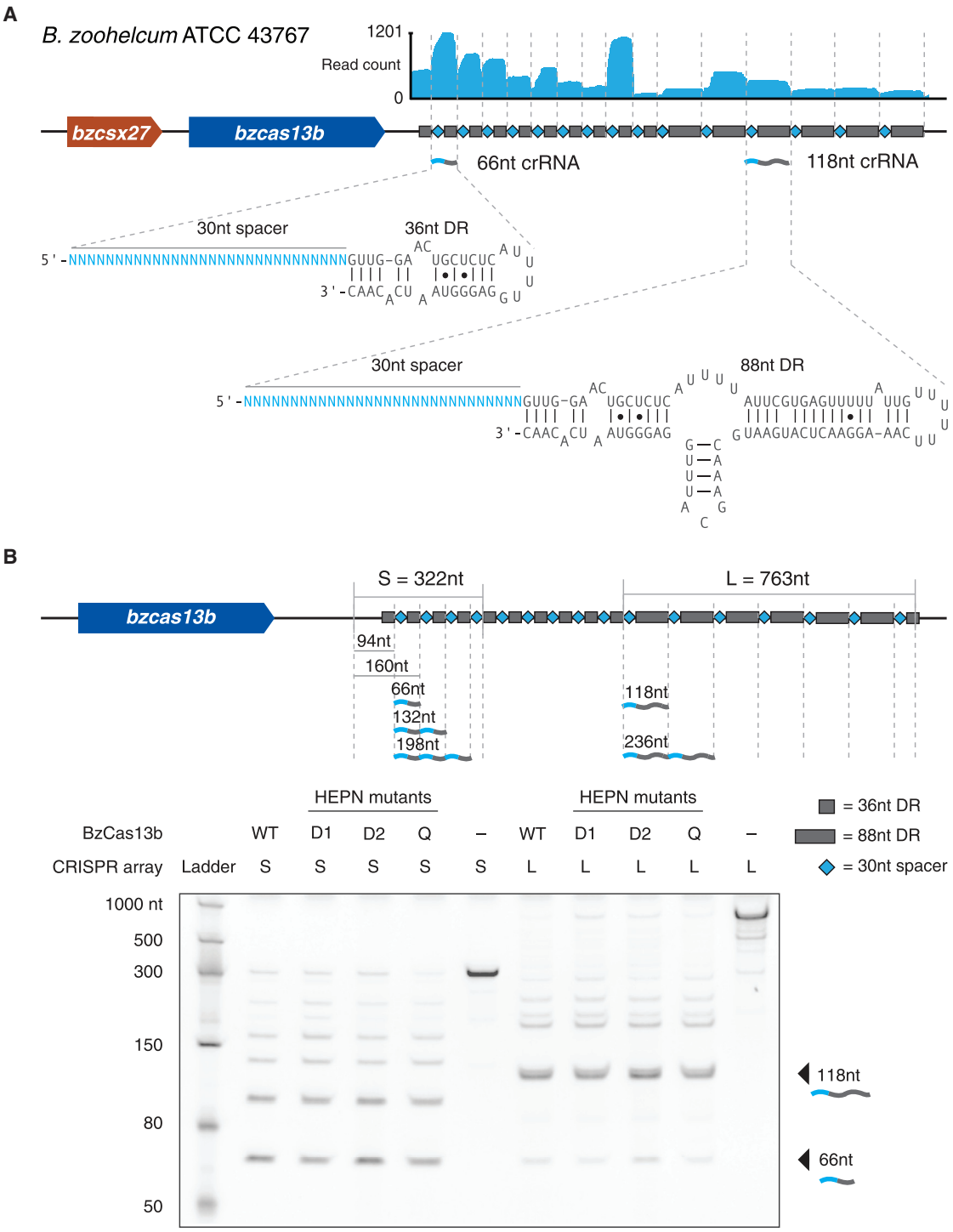
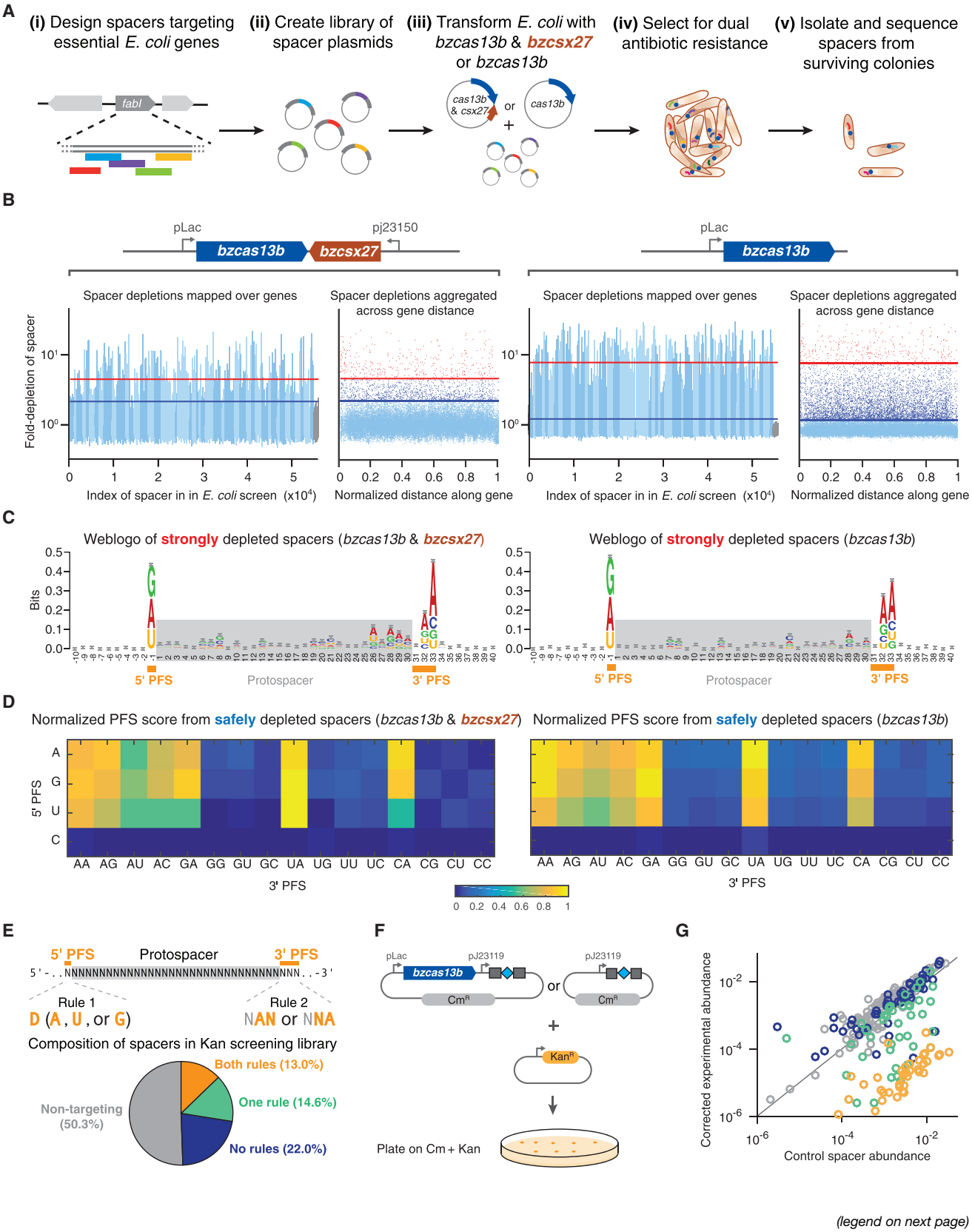


Figure 2. Cas13b from the VI-B1 Locus Processes a CRISPR Array with Two Direct Repeat Variants

(A) RNA sequencing of the native VI-B1 locus from *Bergeyella zoohelcum* ATCC 43767.

(B) Denaturing gel showing cleavage products of in vitro-synthesized short-DR-containing or long-DR-containing CRISPR arrays from the *B. zoohelcum* genome by either wild-type or HEPN mutant BzCas13b (D1, R116A/H121A; D2, R1177A/H1182A; and Q, R116A/H121A/R1177A/H1182A). The schematic shows fragment lengths of a cleaved CRISPR array. See also Figure S4 and Table S1.



(legend on next page)

alternating A then U was also tested (Figure S6B). We observed cleavage at pyrimidine residues, with a strong preference for uracil. This activity is abolished in the presence of EDTA (Figure S6C), suggesting a divalent metal ion-dependent mechanism for RNA cleavage akin to that of a similar HEPN-containing, class 2 effector protein, Cas13a (Abudayyeh et al., 2016; East-Seletsky et al., 2016).

Given that Cas13a has also been reported to cleave RNA non-specifically once activated by interaction with the target ("collateral effect") (Abudayyeh et al., 2016; East-Seletsky et al., 2016), we sought to test the ability of Cas13b to cleave a second, non-specific substrate following target cleavage. Using an in vitro assay similar to the one we previously used with Cas13a (Abudayyeh et al., 2016), we incubated Cas13b-crRNA complexes with both a target and non-target RNA substrate. We observed collateral cleavage of the non-targeted RNA, but only in the presence of the target RNA (Figure 4E).

Cas13b Shows Robust HEPN-Dependent Interference and Is Repressed by Csx27 Activity

To validate RNA interference in vivo, we assayed interference against the lytic, single-stranded RNA bacteriophage MS2, whose life cycle contains no DNA intermediates. We performed an MS2 drop plaque assay at serial dilutions of phage for both *bzcas13b* with *bzcsx27* and for *bzcas13b* alone with three spacers targeting the MS2 genome, two at the *lys-rep* interface and one in *rep*, as well as one non-targeting spacer (Figure 5A). We observed substantial reduction in plaque formation for all targeting spacers compared to the non-targeting spacer, confirming sequence-specific RNA targeting by VI-B1 systems. (Figures 5A and S5B; Table S6). Notably, the presence of *bzcsx27* weakened RNA interference by *bzcas13b* for all three targeting spacers.

To confirm the lack of DNA interference in vivo, we adapted a previously established plasmid interference assay (Zetsche et al., 2015) with a protospacer placed either in-frame at the 5′ end of the *bla* ampicillin-resistance gene (transcribed target) or upstream of the *bla* gene promoter on the opposite strand (non-transcribed target). Bacteria co-transformed with *bzcas13b* and spacer as well as the non-transcribed target plasmid survived at a rate comparable to that of co-transformation of the same target with the empty vector on dual antibiotic selection. For bacteria co-transformed with the transcribed target, the colony-forming unit rate under dual antibiotic selection was reduced by approximately two orders of magnitude in the presence of *bzcas13b*, corroborating that Cas13b exclusively targets RNA in vivo (Figure 5B).

We next tested if predicted catalytic residues in the HEPN domains were responsible for RNA cleavage by Cas13b. Three HEPN mutants were generated by replacing the conserved catalytic arginines and histidines in the two HEPN domains with alanines (R116A/H121A, termed domain 1 [D1]; R1177A/H1182A, termed domain 2 [D2]; and R116A/H121A/R1177A/H1182A, termed quadruple [Q]) (Figure S4). All mutants lacked observable cleavage activity (Figure 5C), yet retained RNA binding capacity in vitro (Figures 5D and S6E). The wild-type and all three HEPN mutant Cas13b proteins showed comparable binding affinities for a single-stranded target RNA substrate, with K_D values ranging from 27 nM to 42 nM (Figures 5D and S6E; Table S7). The K_D for off-target binding was found to be greater than 188nM (Figure S6F).

We confirmed the involvement of the HEPN domains in RNA interference in vivo, finding ~5.5 orders of magnitude decrease in resistance to MS2 phage in the quadruple HEPN mutants versus wild-type Cas13b (Figures 5E and S5B). Interestingly, quadruple mutant Cas13b with spacers 2 and 3 still showed weak phage resistance, potentially due to catalytically inactive Cas13b binding to phage genomic RNA, leading to reduced phage replication.

Computational Modeling Predicts Additional Targeting Rules Governing Cas13b

Our sequence-based targeting results from the *E. coli* essential gene screen implied the existence of additional RNA-targeting rules beyond the PFS (only ~18% of spacers were safely depleted for *bzcas13b*; from the PFS rules alone, the expected value would be ~33%). Given that RNA targets contain a variety of secondary structures, we sought to determine how RNA accessibility impacts targeting. Using the Vienna RNAplfold method (Bernhart et al., 2006), which has been successfully employed to predict RNAi efficiency (Tafer et al., 2008) (Figure 6A), we trained and tested an RNA accessibility model for spacer efficiency on our screen data and found that RNA accessibility matters the most in the protospacer region most distal to the direct repeat of the crRNA (Figures 6B and 6C).

Given the collateral activity observed in vitro, we examined our screen data for indications of non-specific RNA cleavage by Cas13b. To this end, we calculated the empirical cumulative

Figure 3. Heterologous Expression of Cas13b Mediates Knockdown of *E. coli* Essential Genes by a Double-Sided PFS

(A) Design of *E. coli* essential gene screen to determine targeting rules of nucleic acid interference.

(B) Manhattan plots of mean spacer depletions mapped over 45 genes and aggregated across normalized gene distance for either the full *B. zoohelcum* VI-B1 locus (left) or *cas13b* alone (right), with non-targeting spacers in gray, safely depleted spacers (>5σ above mean depletion of non-targeting spacers) above blue line, and strongly depleted spacers (top 1% depleted) above red line. For the full locus, 36,142 targeting spacers and 630 non-targeting spacers passed QC filter. Of the targeting, 367 are strongly depleted, and 1,672 are safely depleted. For *cas13b* alone, 35,272 targeting spacers and 633 non-targeting spacers passed QC filter. Of the targeting, 359 are strongly depleted, and 6,374 are safely depleted.

(C) Weblogo of sequence motifs of strongly depleted *B. zoohelcum* spacers.

(D) Normalized PFS score matrix, where each score is the ratio of number of safely depleted *B. zoohelcum* spacers to total number of spacers for a given PFS, scaled so that maximum PFS score is 1. The 3′ PFS letters represent the RNA bases at the second and third 3′ PFS position.

(E) Spacers targeting kanamycin to validate PFS targeting rules of 5′ PFS (D) and 3′ PFS (NAN or NNA).

(F) Schematic of kanamycin validation screen for *B. zoohelcum cas13b* in *E. coli*.

(G) Results from kanamycin validation screen; spacer abundances versus control for individual *B. zoohelcum* spacers, with abundances colored by type of spacer. See also Figure S5, Table S2, Table S3, Table S4, and Table S5.

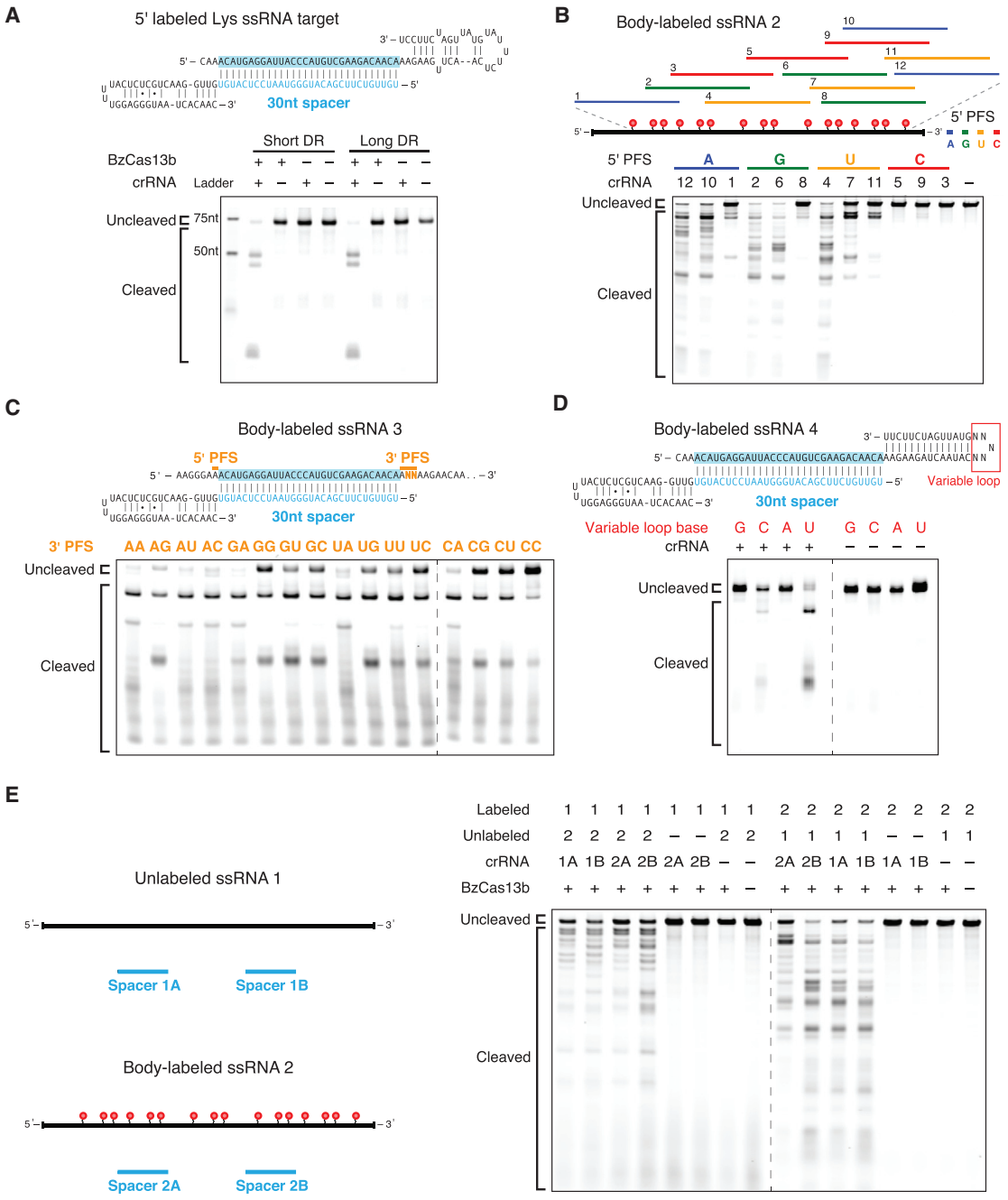


Figure 4. Cas13b Is a Programmable Single-Stranded RNase with Collateral Activity

(A) Schematic showing the RNA secondary structure of the cleavage target in complex with a targeting 30 nt spacer connected to short direct repeat (top). Denaturing gel demonstrating short direct repeat and long direct repeat crRNA-mediated ssRNA cleavage (bottom). Reactions were incubated for 10 min. The ssRNA target is 5′ labeled with IRDye 800. Three cleavage sites are observed.

(B) Schematic showing three numbered protospacers for each colored 5′ PFS on a body-labeled ssRNA target (top); denaturing gel showing crRNA-guided ssRNA cleavage activity demonstrating the requirement for a D 5′ PFS (not C) (bottom). Reactions were incubated for 60 min. crRNAs correspond to protospacer numbered from the 5′ to the 3′ end of the target. Gel lane containing RNA ladder is not shown.

(C) Schematic of a body-labeled ssRNA substrate being targeted by a crRNA (top). The protospacer region is highlighted in blue, and the orange bars indicate the 5′ PFS and 3′ PFS sequences. The orange letters represent the altered sequences in the experiment. Denaturing gel showing crRNA-guided ssRNA cleavage activity after 60 min of incubation, with the 5′ PFS tested as A and the 3′ PFS tested as ANN (bottom). The orange 3′ PFS letters represent the RNA bases at the second and third 3′ PFS position within each target ssRNA. Gel lane containing RNA ladder is not shown. Dashed line indicates two separate gels shown side by side.

(legend continued on next page)

distribution functions of safely depleted spacers aggregated across all essential genes from the 5′ UTR into the gene and from the 3′ UTR into the gene (Figure 6D). Because cleavage closer to the 5′ UTR is more likely to disrupt gene function, without non-specific RNase activity we would expect an overrepresentation of spacers in the 5′ UTR and an underrepresentation in the 3′ UTR. By contrast, in the presence of collateral activity a nearly uniform distribution would be expected. From our screen data, we observed a marginal underrepresentation of spacers in the 3′ UTR compared to a uniform distribution, suggesting that collateral activity may occur in vivo.

CRISPR-Cas13b Effectors Are Differentially Regulated by Csx27 and Csx28

To determine if the established RNA targeting rules generalize across the subtype VI-B systems from diverse bacteria, we characterized the subtype VI-B2 locus from *P. buccae.* RNA sequencing of the CRISPR array revealed processing effectively identical to that of *B. zoohelcum*, excluding the long crRNA (Figure S7A). The *E. coli* essential gene screen with *pbcas13b* and *pbcsx28* or *pbcas13* alone led to the identification of a PFS matrix similar to that of *B. zoohelcum*, with certain PFSs disfavored (Figures 7A, S7B, and S7C). Similar to BzCas13b, PbCas13b was found to cleave targeted single-stranded RNA in vitro (Figure S6D). As with *bzcsx27*, the presence of *pbcsx28* did not appreciably alter the PFS. We also repeated the secondary structure analysis with *pbcas13b*, and a comparable RNAplfold model applied (Figure S7D). Strikingly, in these experiments the safely depleted spacers for *pbcas13b* alone were highly biased to the beginning of the 5′ UTR of genes, suggestive of inhibited or more spatially localized RNase activity in the absence of *pbcsx28* (Figure S7E). We further explored the apparent reduced activity of *pbcas13b* alone relative to the respective full CRISPR-Cas locus using the MS2 phage drop plaque assay and found that *pbcsx28* enhances MS2 phage interference by up to four orders of magnitude (Figures 7B and S5B). The differential ability of *csx27* to repress and *csx28* to enhance *cas13b* activity generalizes across thousands of spacers in the *E. coli* essential gene screen (Figure 7C), highlighting the distinctive regulatory modes of the two variants of subtype VI-B CRISPR-Cas systems.

To further explore the ability of the small accessory proteins to modulate Cas13b activity, we tested if Csx27 can also repress PbCas13b using the MS2 drop plaque assay. Cells co-transformed with *pbcas13b* and *bzcsx27* expression plasmids exhibited a 10^5-fold reduction in interference activity relative to *pbcas13b* expression plasmid and pUC19 empty vector, indicating that Csx27 exerts an inhibitory effect on PbCas13b (Figures 7D and S5B). The ability of Csx27 to modulate the interference activity of BzCas13b and PbCas13b suggests that it is a modular protein that can function across multiple VI-B loci.

DISCUSSION

Here we describe two RNA-targeting CRISPR class 2 systems of subtype VI-B (VI-B1 and VI-B2), containing the computationally discovered RNA-guided RNase Cas13b. Type VI-B systems show several notable similarities to the recently characterized VI-A system. The single protein effectors of both systems cleave single-stranded RNA via HEPN domains, process their CRISPR arrays independent of the HEPN domains, and exhibit collateral RNase activity. Cas13b proteins, however, show only limited sequence similarity to Cas13a, and the common ancestry of the two type VI subtypes remains uncertain. Furthermore, the type VI-B systems differ from VI-A in several other ways, including the absence of both *cas1* and *cas2*, which are involved in spacer acquisition in other CRISPR-Cas systems (Mohanraju et al., 2016). The VI-B CRISPR arrays contain multiple spacers that differ among closely related bacterial strains, suggesting that acquisition does occur, either autonomously or possibly in *trans*, by recruiting Cas1 and Cas2 encoded in other CRISPR-Cas loci from the same genome. In *trans* utilization of adaptation modules of other CRISPR-Cas systems is compatible with the finding that the great majority of type VI-B systems co-occur in the same bacterial genome as other CRISPR-Cas loci that include *cas1* and *cas2* genes; conceivably, the three VI-B-carrying genomes that lack adaptation modules have lost them recently. Additionally, VI-B systems differ from VI-A systems by the presence of the small accessory proteins Csx27 (VI-B1 systems) and Csx28 (VI-B2 systems), which exert opposing regulatory effects on Cas13b activity.

Repression of Cas13b by Csx27 in VI-B1 systems could be part of an important regulatory mechanism of phage interference. The ability of Csx27 to repress Cas13b activity may be a general property, as we found that it can also repress PbCas13b (subtype VI-B2). In the case of type VI-B2 systems, Csx28 might enhance the collateral activity of Cas13b to inactivate numerous transcripts of invading bacteriophages or to promote programmed cell death. Both Csx27 and Csx28 contain predicted long, hydrophobic α helices that might enable them to interact physically with Cas13b, but this remains to be determined. We did not find homologs of Csx27 or Csx28 encoded in any CRISPR-Cas loci other than type VI-B loci, suggesting that these proteins might function in tight association with Cas13b.

As with previously characterized class 2 CRISPR-Cas effectors, such as Cas9 and Cpf1, there is enormous potential to harness Cas13b for use as a molecular tool (Cong et al., 2013;

(D) Schematic showing the secondary structure of the body-labeled ssRNA targets used in the denaturing gel. The variable loop of the schematic (represented as N^5) is substituted with five monomers of the variable loop base in the gel (top). Denaturing gel showing cleavage bands of the homopolymer variable loop base (bottom). The targets were incubated for 30 min. Dashed line indicates seprate gel images (shown in Figure S7B). Gel lane containing RNA ladder is not shown.

(E) Denaturing gel showing BzCas13b collateral cleavage activity after 30 min of incubation, with schematic of cleavage experiment to the right. Two crRNAs (A and B) target substrate 1 (1A and 1B) or substrate 2 (2A and 2B). Gel lane containing RNA ladder is not shown. Dashed line indicates two separate gels shown side by side. See also Figures S4 and S6 and Table S1.

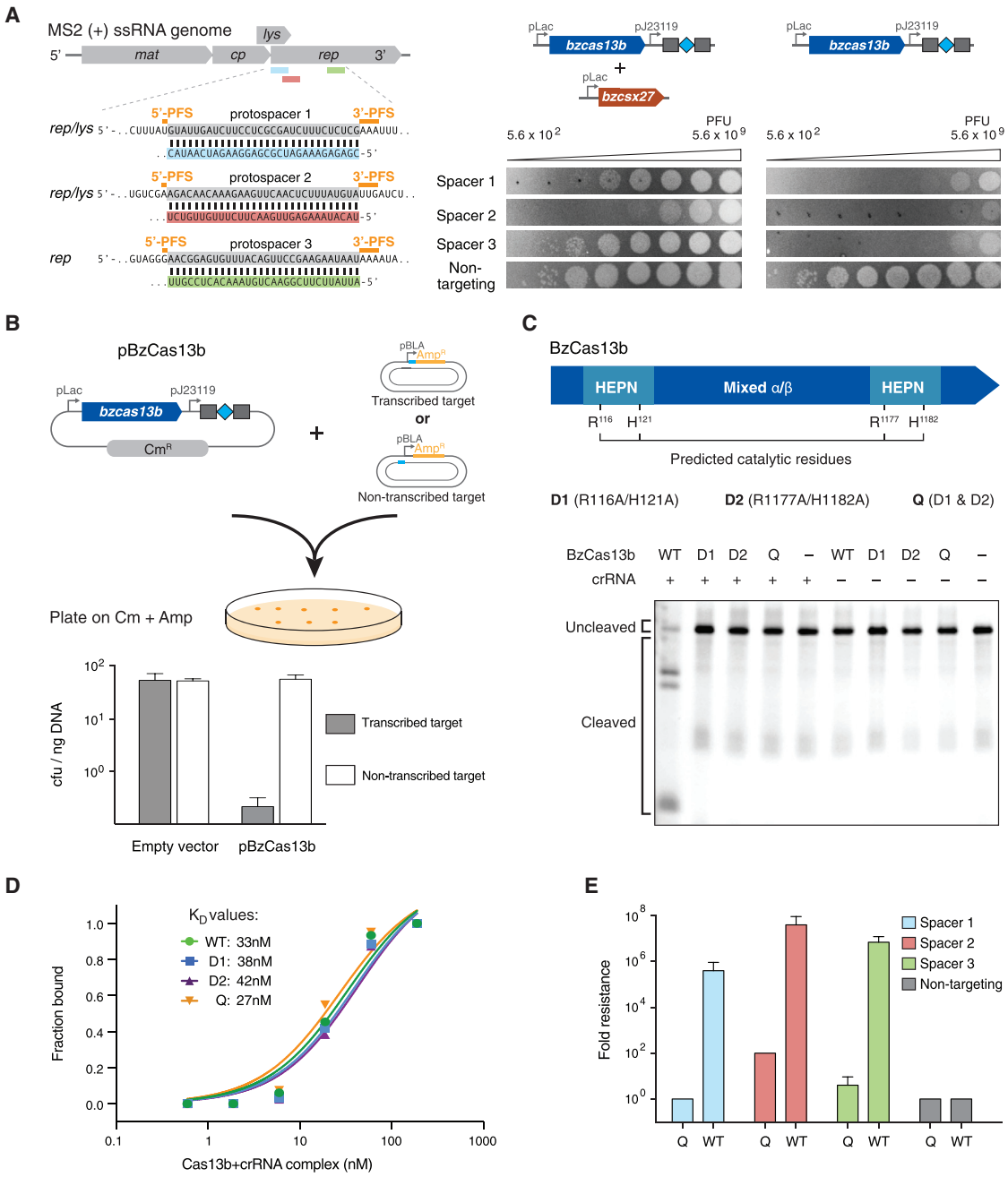


Figure 5. HEPN Domains Mediate RNA Cleavage by Cas13b, the Activity of Which Is Repressed by Csx27

(A) Protospacer design for MS2 phage drop plaque assay to test RNA interference (left); drop plaque assay for full *B. zoohelcum* VI-B1 locus (center) and *bzcas13b* (right).

(B) DNA interference assay schematic (top) and results (bottom). A target sequence is placed in-frame at the start of the transcribed *bla* gene that confers ampicillin resistance or in a non-transcribed region on the opposite strand of the same target plasmid. Target plasmids were co-transformed with *bzcas13b* plasmid or empty vectors conferring chloramphenicol resistance and plated on double selection antibiotic plates.

(C) Schematic (top) and denaturing gel (bottom) showing ssRNA cleavage activity of WT and HEPN mutant BzCas13b. The protein and targeting crRNA complexes were incubated for 10 min. Gel lane containing RNA ladder is not shown.

(D) Electrophoretic mobility shift assay (EMSA) graph showing the affinity of BzCas13b proteins and targeting crRNA complex to a 5′ end-labeled ssRNA. EMSA was performed with supplemental EDTA to reduce any cleavage activity.

(E) Quantification of MS2 phage drop plaque assay with *B. zoohelcum* wild-type and Q (R116A/H121A/R1177A/H1182A) mutant Cas13b. See also Figures S4–S6, Table S1, Table S2, Table S6, and Table S7.

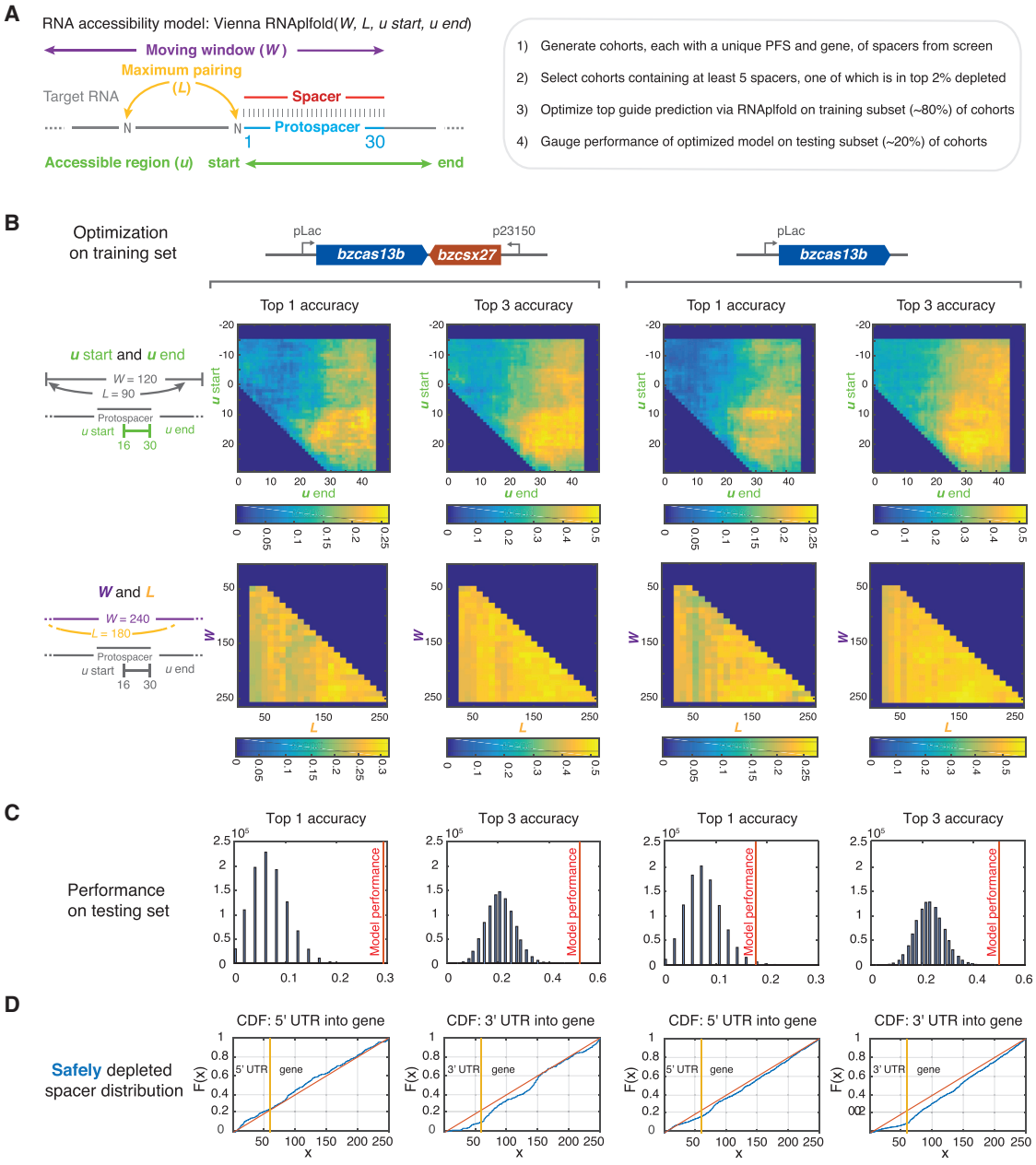


Figure 6. Efficient RNA Targeting by Cas13b Is Correlated with Local RNA Accessibility

(A) Methodology of secondary structure-mediated spacer efficiency analysis of *E. coli* essential gene screen data with Vienna RNAplfold.

(B) Optimization of *top 1 accuracy* (computationally predicted most accessible spacer matches the top experimentally depleted spacer) and *top 3 accuracy* (computationally predicted top spacer falls in top three experimentally depleted spacers) on randomly selected *B. zoohelcum* training dataset using RNAplfold, first with *u start* and *u end*, and then with *W* and *L*.

(C) Performance of optimized RNAplfold model on randomly selected *B. zoohelcum* testing dataset (48 cohorts for full *B. zoohelcum* VI-B1 locus, 56 cohorts for *bzcas13b*) against 10^6 Monte Carlo simulations: empirical p values from left to right of 3e-6, 1e-6, 8.7e-3, and 6e-6.

(D) Empirical cumulative distribution function of safely depleted *B. zoohelcum* spacers over all genes from 5′ UTR into gene and from 3′ UTR into gene. Yellow line separates UTR and gene, red line is theoretical cumulative distribution function of uniformly distributed spacers, and blue line is empirical cumulative distribution of safely depleted *B. zoohelcum* spacers. See also Table S3 and Table S4.

Mali et al., 2013; Wright et al., 2016). A holistic understanding of the factors that affect target selection is essential to the success of any such tools, particularly those that target RNA, where secondary structure will likely impact activity. We therefore developed an *E. coli* essential gene screen to explore the targeting rules of Cas13b more fully. This *E. coli* screen offers several

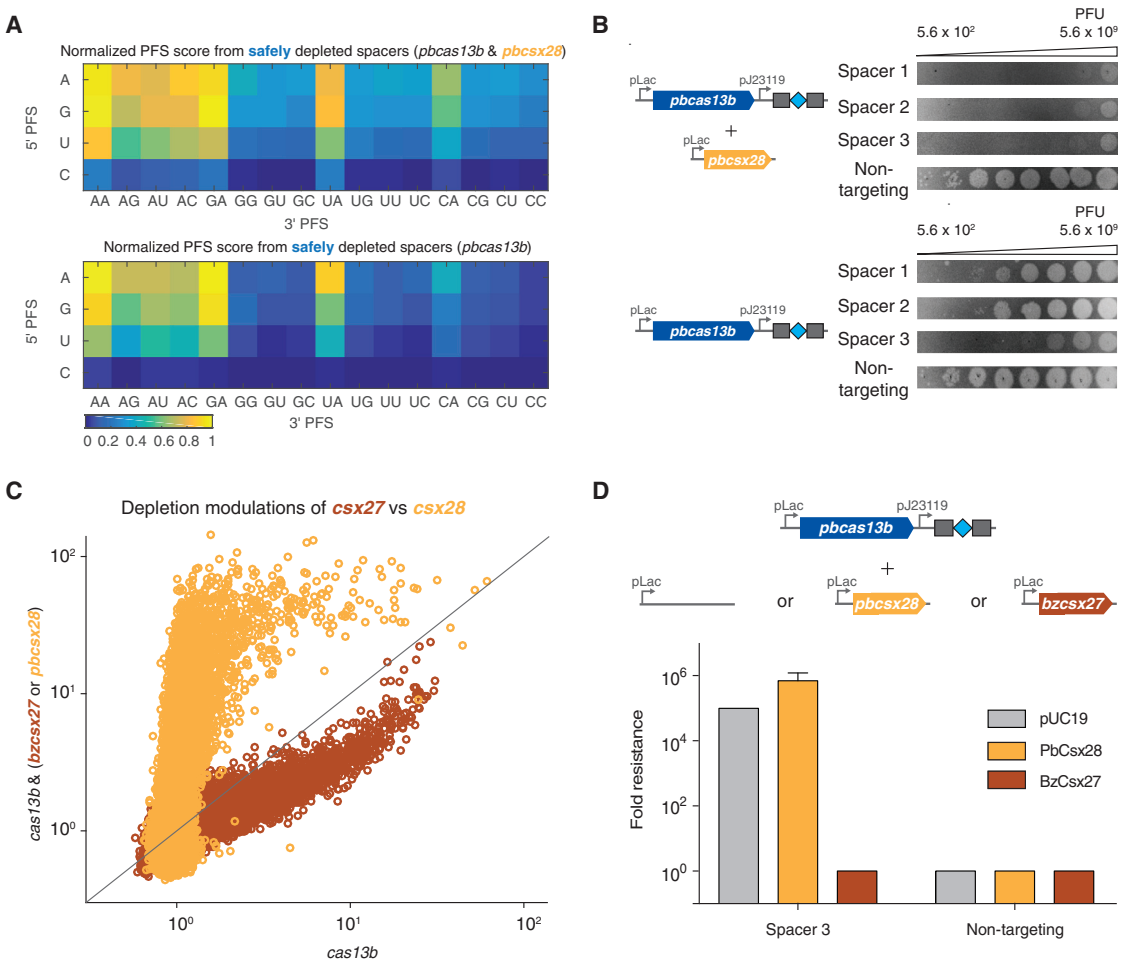


Figure 7. Class 2 Type VI-B Systems Are Differentially Regulated across Two Loci by Csx27 and Csx28

(A) Normalized PFS matrix for *P. buccae* VI-B2 locus (top) and *pbcas13b* (bottom). The 3′ PFS letters represent the RNA bases at the second and third 3′ PFS position.

(B) MS2 drop plaque assay for full *P. buccae* VI-B2 locus (top) and *pbcas13b* (bottom).

(C) Spacer depletions of *bzcas13b* with and without *bzcsx27* (brown), as compared to *pbcas13b* with and without *pbcsx28* (gold).

(D) Fold resistance to MS2 infection for cells co-transformed with *pbcas13b* and the indicated *csx* expression plasmid. See also Figures S5 and S7, Table S2, Table S3, Table S4, and Table S6.

advantages by increasing the number of guides testable in a single experiment to explore how diverse spacer and flanking sequences may affect Cas13b activity. This screen revealed a double-sided PFS in VI-B systems, which may give insight into Cas13b protein-RNA interactions and could help improve specificity by expanding sequence targeting constraints (Ran et al., 2015).

The characterization of Cas13b and other RNA-targeting CRISPR systems raises the prospect of a suite of precise and robust in vivo RNA manipulation tools for studying a wide range of biological processes (Abil and Zhao, 2015; Filipovska and Rackham, 2011; Mackay et al., 2011). The ability of Cas13b to process its own CRISPR array could be extended to multiplex transcriptome engineering. In addition, the VI-B functional long direct repeats could be altered to incorporate stem loops akin to the Cas9-SAM system (Konermann et al., 2015). Like Cas9 and Cpf1, Cas13a and Cas13b may be utilized for complementary applications in science and technology.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - *B. zoohelcum*
  - *E. coli*
  - One Shot Stbl3 *E. coli*
  - NEB 10-Beta Competent *E. coli* (High Efficiency)
  - MegaX DH10B T1R Electrocompetent Cells

- One Shot BL21(DE3)pLysE Chemically Competent *E. coli*
- METHOD DETAILS
  - Computational Sequence Analysis
  - Nucleic Acid Preparation
  - BzCas13b Protein Purification
  - BzCas13b HEPN Mutant Protein Purification
  - PbCas13b Protein Purification
  - Nuclease Assay
  - EMSA Assay
  - RFP-Tagged Protein Fluorescent Imaging
  - Bacterial RNA-Sequencing
  - *E. coli* Essential Gene Screen Experiment
  - *E. coli* Essential Gene Screen Analysis
  - Kanamycin Validation Screen Experiment
  - Kanamycin Validation Screen Analysis
  - MS2 Phage Drop Plaque Assay
  - DNA Interference Assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - MS2 Interference Assay-HEPN Mutants
  - DNA Interference Assay
  - *E. coli* Essential Gene Screen
  - K_D Calculations
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article at http://dx.doi.org/10.1016/j.molcel.2016.12.023.

AUTHOR CONTRIBUTIONS

A.A.S., D.B.T.C, and N.K.P. are co-first authors. A.A.S., D.B.T.C., and N.K.P., under the guidance of F.Z., designed and performed all experiments and analyzed data. A.A.S., with help from K.Z. and N.K.P., led the design and implementation of the CRISPR discovery pipeline. A.A.S. also performed screening experiments and computational analysis and modeling of Cas13b targeting. D.B.T.C., with help from J.S.G., O.A.A., and P.E., led the microbiological assays for characterizing in vivo function of Cas13b, Csx27, and Csx28. N.K.P., with help from A.A.S. and I.M.S., led the biochemical characterizations of Cas13b activity. S.S., K.S.M., and E.V.K. provided input on annotation, classification, and naming of Cas13b, Csx27, and Csx28. A.A.S., D.B.T.C., N.K.P., and F.Z. wrote the manuscript with input from all authors.

ACKNOWLEDGMENTS

We would like to thank R. Belliveau for overall research support, R. Macrae for critical reading of the manuscript, and the entire Zhang laboratory for support and advice. D.B.T.C. is supported by award number T32GM007753 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the NIH. I.M.S. is supported by the Simons Center for the Social Brain. J.S.G. is supported by a D.O.E. Computational Science Graduate Fellowship. O.A.A. is supported by a Paul and Daisy Soros Fellowship, a Friends of the McGovern Institute Fellowship, and the Poitras Center for Affective Disorders. K.S.M., E.V.K., and S.S. are supported by the intramural program of the U.S. Department of Health and Human services (to the National Library of Medicine). S.S. is also supported by a Skoltech-MIT Next Generation Program grant (to K. Severinov and F.Z.). F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049); NSF; the New York Stem Cell Foundation; the Allen Distinguished Investigator Program, through The Paul G. Allen Frontiers Group; the Simons and Vallee Foundations; the Howard Hughes Medical Institute; the Skoltech-MIT Next Generation Program; James and Patricia Poitras and the Poitras Center for Affective Disorders; Robert Metcalfe; and David Cheng. F.Z. is a New York Stem Cell Foundation-Robertson Investigator. A patent application has been filed related to this work, and the authors plan to make the reagents widely available to the academic community through Addgene and to provide protocols and software tools via the Zhang lab website (http://www.genome-engineering.org) and GitHub (https://www.github.com/fengzhanglab).

Received: November 10, 2016
Revised: December 6, 2016
Accepted: December 23, 2016
Published: January 5, 2017

REFERENCES

Abil, Z., and Zhao, H. (2015). Engineering reprogrammable RNA-binding proteins for study and manipulation of the transcriptome. Mol. Biosyst. *11*, 2658–2665.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Anantharaman, V., Makarova, K.S., Burroughs, A.M., Koonin, E.V., and Aravind, L. (2013). Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. Biol. Direct *8*, 15.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006 0008.

Bernhart, S.H., Hofacker, I.L., and Stadler, P.F. (2006). Local RNA base pairing probabilities in large sequences. Bioinformatics *22*, 614–615.

Biswas, A., Gagnon, J.N., Brouns, S.J., Fineran, P.C., and Brown, C.M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. RNA Biol. *10*, 817–827.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

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

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

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

Edgar, R.C. (2007). PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics *8*, 18.

Filipovska, A., and Rackham, O. (2011). Designer RNA-binding proteins: new tools for manipulating the transcriptome. RNA Biol. *8*, 978–983.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides, N.C., Anderson, I., Gelfand, M.S., et al. (2003). Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J. Bacteriol. *185*, 5673–5684.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945–956.

Heidrich, N., Dugar, G., Vogel, J., and Sharma, C.M. (2015). Investigating CRISPR RNA biogenesis and function using RNA-seq. Methods Mol. Biol. *1311*, 1–21.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA *89*, 10915–10919.

Hildebrand, A., Remmert, M., Biegert, A., and Söding, J. (2009). Fast and accurate automatic structure prediction with HHpred. Proteins *77* (*Suppl 9*), 128–132.

Jiang, W., Samai, P., and Marraffini, L.A. (2016). Degradation of phage transcripts by CRISPR-associated rnases enables type III CRISPR-Cas immunity. Cell *164*, 710–721.

Kim, Y.K., Kim, Y.G., and Oh, B.H. (2013). Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of Pyrococcus furiosus. Proteins *81*, 261–270.

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

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

Mackay, J.P., Font, J., and Segal, D.J. (2011). The prospects for designer single-stranded RNA-binding proteins. Nat. Struct. Mol. Biol. *18*, 256–261.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. *13*, 722–736.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science *353*, aad5147.

Möller, S., Croning, M.D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics *17*, 646–653.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191.

Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175.

Sheppard, N.F., Glover, C.V., 3rd, Terns, R.M., and Terns, M.P. (2016). The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease. RNA *22*, 216–224.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol. Cell *60*, 385–397.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol. http://dx.doi.org/10.1038/nrmicro.2016.184.

Staals, R.H., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., et al. (2013). Structure and activity of the RNA-targeting type III-B CRISPR-Cas complex of *Thermus thermophilus*. Mol. Cell *52*, 135–145.

Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., et al. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. Mol. Cell *56*, 518–530.

Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J., and Hofacker, I.L. (2008). The impact of target site accessibility on the design of effective siRNAs. Nat. Biotechnol. *26*, 578–583.

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A.O., Dickman, M.J., Horvath, P., and Siksnys, V. (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. Mol. Cell *56*, 506–517.

Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell *164*, 29–44.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44* (D1), D710–D716.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
BzCas13b	This study	Table S2
BzCas13b mutants (D1, D2, Q)	This study	Table S2
PbCas13b	This study	Table S2
SUPERase in RNase inhibitor	Thermo Fisher Scientific	AM2696
TURBO DNase	Life Technologies	AM2238
SUMO protease	Thermo Fisher Scientific	12588018
Critical Commercial Assays		
MiSeq Reagent Kit v3 (150-cycle)	Illumina	MS-102-3001
NextSeq 500/550 High Output v2 kit (150 cycles)	Illumina	FC-404-2002
HiScribe T7 High Yield RNA Synthesis kit	New England Biolabs	E2040S
HiScribe T7 Quick High Yield RNA Synthesis kit	New England Biolabs	E2050S
5′ oligonucleotide kit	VectorLabs	MB-9001
Deposited Data		
EMSA raw data	This study	Table S7
Next-generation sequencing for bacterial RNA sequencing, *E. coli* essential gene screen, kanamycin validation screen	This study	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA358111
Experimental Models: Cell Lines		
Experimental Models: Organisms/Strains		
B. zoohelcum	ATCC	43767
E. coli	ATCC	15597
E. coli bacteriophage MS2	ATCC	15597-B1
One Shot Stbl3 *E. coli*	Thermo Fisher Scientific	C737303
NEB 10-beta Competent *E. coli* (High Efficiency)	New England BioLabs	C3019H
MegaX DH10B T1R electrocompetent cells	Thermo Fisher Scientific	C640003
One Shot BL21(DE3)pLysE chemically competent *E. coli*	Invitrogen	C656503
Recombinant DNA		
pMAX-GFP	Lonza	Not commercially available, except as part of a nucleofection kit: http://bio.lonza.com/fileadmin/groups/FAQs/public/Technology_Flyer.pdf
6x His/Twin Strep SUMO, a pET-based vector	Gift from Ilya Finkelstein	N/A
Plasmids generated in this study	This study	Table S2
Sequence-Based Reagents		
List of spacers for *E. coli* essential gene screen	This study	Table S4
List of spacers for kanamycin validation screen	This study	Table S5
List of spacers for MS2 interference and pBLA assays	This study	Table S6
ssRNA targets	This study	Table S1
DNA primers	This study	Table S1
crRNAs	This study	Table S1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
PILER-CR	Edgar, 2007	http://drive5.com/pilercr/
BLASTP	Camacho et al., 2009	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
HHpred	Remmert et al., 2011	https://toolkit.tuebingen.mpg.de/hhpred
BLOSUM62	Henikoff and Henikoff, 1992	
Vienna RNAfold	Anantharaman et al., 2013; Lorenz et al., 2011	http://rna.tbi.univie.ac.at/
CRISPRTarget	Biswas et al., 2013	http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html
Burrows-Wheeler Aligner	Li and Durbin, 2009	http://bio-bwa.sourceforge.net
The Galaxy Project	Center for Comparative Genomics and Bioinformatics at Penn State, and Department of Biology and at Johns Hopkins University	https://usegalaxy.org
Vienna RNAplfold	Bernhart et al., 2006	http://rna.tbi.univie.ac.at/
ImageJ	Wayne Rasband, NIH	https://imagej.nih.gov/ij/
GraphPad Prism version 7	GraphPad Software, La Jolla California USA	https://www.graphpad.com/scientific-software/prism/
MATLAB	MathWorks, Natick, Massachusetts, United States	https://www.mathworks.com/products/matlab.html

CONTACT FOR RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Feng Zhang (zhang@broadinstitute.org). The authors plan to make the reagents widely available to the academic community through Addgene subject to a MTA and to provide protocols and software tools via the Zhang lab website (www.genome-engineering.org) and GitHub (https://www.github.com/fengzhanglab).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

B. zoohelcum

B. zoohelcum ATCC 43767 was grown in ATCC medium 44 (Brain Heart Infusion broth) at 37°C at 250 rpm overnight.

E. coli

E. coli was grown in LB at 37°C at 250 rpm overnight.

One Shot Stbl3 *E. coli*

E. coli was grown in LB at 37°C at 250 rpm overnight.

NEB 10-Beta Competent *E. coli* (High Efficiency)

NEB 10-beta Competent *E. coli* was transformed on LB agar at 37°C overnight.

MegaX DH10B T1R Electrocompetent Cells

NEB 10-beta Competent *E. coli* was transformed on LB agar at 37°C overnight.

One Shot BL21(DE3)pLysE Chemically Competent *E. coli*

The BzCas13b expression construct (Table S2) was transformed into One Shot BL21(DE3)pLysE (Invitrogen) cells. 25 mL of 6hr growing culture were inoculated into 2 l of Terrific Broth 4 growth media (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K2HPO, 2.2 g/L KH2PO4, Sigma). Cells were then grown at 37°C to a cell density of 0.6 OD600, and then SUMO-BzCas13b expression was induced by supplementing with IPTG to a final concentration of 500 μM. Induced culture was grown for 16-18 hr before harvesting cell paste, which was stored at −80°C until subsequent purification. For each BzCas13b mutant, 1 L of Terrific Broth was used to

generate cell paste and all other reagents were scaled down accordingly. Protein purification was performed using the same protocol as wild-type Cas13b. PbCas13b was cloned into the same pET based vector and purified using a similar protocol as BzCas13b with the following differences: cells were grown at 21°C for 18 hr.

METHOD DETAILS

Computational Sequence Analysis

From complete compiled Ensembl Release 27 genomes (Yates et al., 2016), CRISPR repeats were identified using PILER-CR (Edgar, 2007). Proteins within 10kb of identified CRISPR arrays were clustered into loci, with loci rejected if more than one protein of size 700 amino acids or larger or if either Cas1 or Cas2 were present. For candidate Class 2 effectors, only proteins in these remaining loci of size 900aa to 1800aa were selected. These candidate effectors were subjected to the BLASTP (Camacho et al., 2009) search against the NCBI non-redundant (NR) protein sequence database with an E-value cutoff of 1e-7. All discovered proteins were then grouped into putative families via a nearest-neighbor grouping with the same E-value cutoff. Only putative families with at least ten candidate effectors and more than 50% of candidate effectors within 10kb of CRISPR arrays were considered. HHpred (Remmert et al., 2011) and existing CRISPR locus classification rules (Makarova et al., 2015) were used to classify each family, leaving Cas13b as the only unclassified family. Additional Cas13b proteins in the family were found through a nearest-neighbor search of previously discovered Csx27/Csx28 against the NCBI non-redundant (NR) protein sequence database with an E-value cutoff of 1e-7, and then by searching in genomes within 1kb of any newly discovered Csx27/Csx28. Within this Cas13b family, truncated or suspected partially sequenced effectors were discarded, leaving 105 loci, and 81 with a unique protein accession number in the NCBI non-redundant (NR) protein sequence database. Multiple sequence alignments on these 81 proteins (as well as the accessory Csx27 and Csx28 proteins) were performed using BLOSUM62 (Henikoff and Henikoff, 1992) to identify the HEPN domains and to sort the loci into phylogenetic trees. Loci represented in the tree of 81 non-redundant proteins were selected first for annotated Csx27/Csx28 within 1kb of Cas13b, and next for annotated CRISPR array within 10kb of Cas13b. Vienna RNAfold (Lorenz et al., 2011) was used to predict the secondary structure of each direct repeat, whose transcriptional orientation was chosen as identical to that of Cas13b in its locus. CRISPRTarget (Biswas et al., 2013) was used to search the spacers in each locus against NCBI phage and plasmid genomes. Weblogos were generated for all unique direct repeats and protospacer flanking sequences (Crooks et al., 2004). TMHMM Server v. 2.0 (Möller et al., 2001) was used to predict the transmembrane helices in Csx27 and Csx28.

Nucleic Acid Preparation

For in vitro synthesis of RNA, a T7 DNA fragment must be generated. To create T7 DNA fragments for crRNAs, top and bottom strand DNA oligos were synthesized by IDT. The top DNA oligo consisted of the T7 promoter, followed by the bases GGG to promote transcription, the 30-nt target and then direct repeat. Oligos were annealed together using annealing buffer (30 mM HEPES pH 7.4, 100 mM potassium acetate, and 2 mM magnesium acetate). Annealing was performed by incubating the mixture for 1 min at 95°C followed by a −1°C/minute ramp down to 23°C. To create ssRNA targets, short targets (Trunc2, 3, 4) were synthesized as top and bottom strand oligos containing the T7 promoter. For long ssRNA targets (E1, E2, S and L CRISPR Arrays), DNA primers (Table S1) with a T7 handle on the forward primer were ordered and the DNA fragment was amplified using PCR. T7 DNA constructs for RNA generation without body labeling were incubated with T7 polymerase overnight (10-14 hr) at 30°C using the HiScribe T7 Quick High Yield RNA Synthesis kit (New England Biolabs). Body-labeled constructs were incubated with Cyanine 5-UTP (Perkin Elmer) and incubated with T7 polymerase overnight at 30°C using the HiScribe T7 High Yield RNA Synthesis kit (New England Biolabs). For a complete list of crRNAs and target ssRNAs used in this study see Table S1. 5′ end labeling was accomplished using the 5′ oligonucleotide kit (VectorLabs) and with a maleimide-IR800 probe (LI-COR Biosciences). 3′ end labeling was performed using a 3′ oligonucleotide labeling kit (Roche) and Cyanine 5-ddUTP (Perkin Elmer). RNAs were purified using RNA Clean and Concentrator columns^{TM}-5 (Zymo Research). Body-labeled dsRNA substrates were prepared by T7 DNA fragments for the bottom and top RNA strand. After synthesis, 1.3-fold excess of non-labeled bottom strand ssRNA was added and re-annealed to ensure the top strand would be annealed to a bottom strand by incubating the mixture for 1 min at 95°C followed by a −1°C/minute ramp down to 23°C.

BzCas13b Protein Purification

The mammalian codon-optimized gene for Cas13b (*B. zoohelcum*) was synthesized (GenScript) and inserted into a bacterial expression vector (6x His/Twin Strep SUMO, a pET based vector received as a gift from Ilya Finkelstein) after cleaving the plasmid with the BamHI and NotI restriction enzymes and cloning in the gene using Gibson Assembly Master Mix (New England Biolabs). The BzCas13b expression construct (Table S2) was transformed into One Shot BL21(DE3)pLysE (Invitrogen) cells. 25 mL of 6hr growing culture were inoculated into 2 l of Terrific Broth 4 growth media (12 g/L tryptone, 24 g/L yeast extract, 9.4 g/L K2HPO, 2.2 g/L KH2PO4, Sigma). Cells were then grown at 37°C to a cell density of 0.6 OD600, and then SUMO-BzCas13b expression was induced by supplementing with IPTG to a final concentration of 500 μM. Induced culture was grown for 16-18 hr before harvesting cell paste, which was stored at −80°C until subsequent purification. Frozen cell paste was crushed and resuspended via stirring at 4°C in 500 mL of Lysis Buffer (50mM NaH_2PO_4 pH 7.8, 400mM NaCl) supplemented with protease inhibitors (cOmplete, EDTA-free, Roche

Diagnostics Corporation) and 1250U of benzonase (Invitrogen). The resuspended cell paste was lysed by a LM20 microfluidizer at 18,000 psi (Microfluidics). Lysate was cleared by centrifugation at 10,000 g for 1 hr. Filtered lysate was incubated with StrepTactin Sepharose High Performance (GE Healthcare Life Sciences) at 4°C for 1 hr with gentle agitation, and then applied to an Econo-column chromatography column (Bio-Rad Laboratories). Resin was washed with Lysis Buffer for 10 column volumes. One column volume of fresh Lysis Buffer was added to the column and mixed with 10 units of SUMO protease (Invitrogen) and incubated overnight. The eluate was removed from the column, SUMO cleavage was confirmed by SDS-PAGE and BlueFast protein staining (Eton Bioscience), and the sample was concentrated via Centrifugal Filter Unit to 2 mL. Concentrated sample was loaded onto a HiTrap Heparin HP column (GE Healthcare Life Sciences) via FPLC (AKTA Pure, GE Healthcare Life Sciences) and eluted over a gradient with an elution buffer with salt concentration of 1.2 M. The resulting fractions were tested for presence of BzCas13b protein by SDS-PAGE; fractions containing BzCas13b were pooled, and concentrated via Centrifugal Filter Unit to 1 mL. Concentrated sample was loaded a gel filtration column (HiLoad 16/600 Superdex 200, GE Healthcare Life Sciences) via FPLC (AKTA Pure, GE Healthcare Life Sciences) with buffer 500 mM NaCl, 50 mM Tris-HCl pH 7.5, 1 mM DTT.

BzCas13b HEPN Mutant Protein Purification

Alanine mutants (Table S2) at each of the HEPN catalytic residues were generated using the Q5 site-directed mutagenesis kit (New England Biolabs) and transformed into One Shot BL21(DE3)pLysE cells (Invitrogen). For each mutant, 1 L of Terrific Broth was used to generate cell paste and all other reagents were scaled down accordingly. Protein purification was performed using the same protocol as wild-type Cas13b.

PbCas13b Protein Purification

PbCas13b (*Prevotella buccae*) was cloned into the same pET based vector and purified using a similar protocol as BzCas13b with the following differences: cells were grown at 21°C for 18 hr. Frozen cell paste was resuspended into 500 mM NaCl, 50 mM HEPES 7.5 and 2 mM DTT prior to breaking cells in the microfluidizer. The Superdex 200 column was run in 500 mM NaCl, 10 mM HEPES 7.0, and 2 mM DTT.

Nuclease Assay

Nuclease assays were performed with equimolar amounts of end-labeled or body-labeled ssRNA target, purified protein, and crRNA, for targeted ssRNA cleavage. For CRISPR array cleavage, protein was supplied in a four times molar excess of the CRISPR array. Reactions were incubated in nuclease assay buffer (10 mM TrisHCl pH 7.5, 50 mM NaCl, 0.5 mM $MgCl_2$, 20U SUPERase In (ThermoFisher Scientific), 0.1% BSA). Reactions were allowed to proceed at 37°C for times specified in the figure legends. After incubation, samples were then quenched with 0.8U of Proteinase K (New England Biolabs) for 15 min at 25°C. The reactions were mixed with equal parts of RNA loading dye (New England Biolabs) and denatured at 95°C for 5 min and then cooled on ice for 2 min. Samples were analyzed by denaturing gel electrophoresis on 10% PAGE TBE-Urea (Invitrogen) run at 45°C. Gels were imaged using an Odyssey scanner (LI-COR Biosciences).

EMSA Assay

For the Electrophoretic Mobility Shift Assay (EMSA), binding experiments were performed with a series of half-log complex dilutions (crRNA and BzCas13b) from 0.594 to 594 nM. Binding assays were performed in nuclease assay buffer (without $MgCl_2$) supplemented with 10 mM EDTA to prevent cutting, 5% glycerol, and 5μg/mL heparin in order to avoid non-specific interactions of the complex with target RNA. Protein was supplied at two times the molar amount of crRNA. Protein and crRNA were pre-incubated at 37°C for 15 min, after which the 5′-labeled target was added. Reactions were then incubated at 37°C for 10 min and then resolved on 6% PAGE TBE gels (Invitrogen) at 4°C (using 0.5X TBE buffer). Gels were imaged using an Odyssey scanner (LI-COR Biosciences). Gel shift of the RNA targets was quantified from an EMSA gel using ImageJ (Wayne Rasband, NIH) and plotted in GraphPad Prism version 7 (GraphPad Software, La Jolla California USA). Line regression was performed in Prism 7 using nonlinear fit with one-site binding hyperbola. K_D values are calculated by GraphPad Prism based on regression analysis of data (Table S7).

RFP-Tagged Protein Fluorescent Imaging

One Shot Stbl3 Chemically Competent *E. coli* were transformed with plasmids containing RFP (negative control) or RFP fused to the N- or C terminus of Csx27 of *B. zoohelcum* or Csx28 of *P. buccae* (Table S2). Clones were cultured up in 5mL of antibiotic LB overnight, then spun down at 5000 g and resuspended in PBS with 1% methanol-free formaldehyde. After 30 min fixation, cells were washed once with PBS and then diluted 1:2 in PBS. 5uL of sample was pipetted onto a silane-coated slide, which was covered with a coverslip. Fluorescent imaging was performed in a 63x objective microscope with oil immersion.

Bacterial RNA-Sequencing

RNA was isolated and prepared for sequencing using a modification of a previously described protocol (Heidrich et al., 2015; Shmakov et al., 2015). RNA was isolated from 5 mL of stationary phase of bacterial cultures by resuspending pelleted cells in 1mL of TRIzol (ThermoFisher Scientific) and then homogenizing with 300 uL zirconia/silica beads (BioSpec Products) in a BeadBeater (BioSpec

Products) for 7 1 min cycles. 200 uL of chloroform was added to the homogenized sample and then samples were centrifuged for 15 min. (12000xg, 4°C). The aqueous phase was then used for input into the Direct-Zol RNA miniprep kit (Zymo). Purified RNA was DNase treated with TURBO DNase (Life Technologies) and 3′ dephosphorylated/5′ phosphorylated with T4 Polynucleotide Kinase (New England Biolabs). rRNA was eliminated using the bacterial Ribo-Zero rRNA removal kit (Illumina). Next, RNA was treated with RNA 5′ polyphosphatase (Epicenter Bio) to convert 5′-triphosphates to 5′-monophosphates for adaptor ligation. Samples were then polyA tailed with *E. coli* Poly(A) polymerase (New England Biolabs), and a 5′ RNA Illumina sequencing adaptor ligated to cellular RNA using T4 RNA Ligase 1 (ssRNA ligase) (New England Biolabs). RNA was reverse transcribed using AffinityScript cDNA synthesis kit (Agilent Technologies) and an oligo-dT primer. cDNA was amplified with Herculase II polymerase (Agilent Technologies) and barcoded primers. The prepared cDNA libraries were sequenced on a MiSeq (Illumina).

For RNA sequencing of native *B. zoohelcum* ATCC 43767, we repeated the experiment with a modified protocol, omitting RNA 5′ polyphosphatase prior to 5′ adaptor ligation, to promote enrichment of processed transcripts originating from the CRISPR array. For heterologous *P. buccae* ATCC 33574 RNA sequencing in *E. coli*, we cloned the locus into pACYC184 (Table S1). Reads from each sample were identified on the basis of their associated barcode and aligned to the appropriate RefSeq reference genome using BWA (Li and Durbin, 2009). Paired-end alignments were used to extract entire transcript sequences using Galaxy (https://usegalaxy.org), and these sequences were analyzed using Geneious 8.1.8.

E. coli Essential Gene Screen Experiment

The intersection of two *E. coli* DH10B strain essential gene studies (Baba et al., 2006; Gerdes et al., 2003) was taken, and further pared down to 45 genes by only selecting genes exclusive to their respective operons (Table S3). Over these 45 genes 54,600 spacers were designed to tile at single resolution across the coding region, as well as to extend 60 nt into the 5′ UTR and 3′UTR. In addition, 1100 non-targeting, pseudorandomly generated spacers with no precise match to the *E. coli* DH10B strain genome were added to the library as a non-targeting negative control. The library of spacers (Table S4) was cloned into a *B. zoohelcum* or *P. buccae* direct repeat-spacer-direct repeat backbone containing a chloramphenicol resistance gene using Golden Gate Assembly (NEB) with 100 cycles, and then transformed over five 22.7cm x 22.7cm chloramphenicol LB Agar plates. Libraries of transformants were scraped from plates and DNA was extracted using the Macherey-Nagel Nucleobond Xtra Midiprep Kit (Macherey-Nagel). 50ng of library plasmid and equimolar gene plasmid containing an ampicillin resistance gene (*bzcas13b, bzcas13b & bzcsx27, pbcas13b, pbcas13b & pbcsx28*, empty vector pBR322) (Table S2) were transformed into MegaX DH10B T1R Electrocomp Cells (ThermoFisher) according to manufacturer's protocol, with four separate 22.7cm x 22.7cm carbenicillin-chloramphenicol LB Agar plates per bioreplicate, and three bioreplicates per condition (twelve transformations total per condition). Eleven hours post-transformation, libraries of transformants were scraped from plates and DNA extracted using the Macherey-Nagel Nucleobond Xtra Maxiprep Kit (Macherey-Nagel).

E. coli Essential Gene Screen Analysis

Prepared DNA libraries were sequenced on a NextSeq (Illumina), with reads mapped to the input library of spacers. Spacer depletions were calculated as the read abundance of a spacer in the empty vector condition divided by read abundance in each gene plasmid condition. Mean depletions over three bioreplicates were calculated. We imposed a two-step quality-control filter on the data: a maximum coefficient of variation of 0.2 for depletion over three bioreplicates, and a minimum spacer read abundance of $1/3N$ in each bioreplicate, where N = 55,700. Weblogos of the strongly depleted (top 1% depleted) spacers were generated (Crooks et al., 2004), and from each identified PFS, heatmaps of the ratio of safely depleted (> 5σ above mean depletion of non-targeting spacers) spacers to all spacers in the screen were generated. For spatial analysis via empirical cumulative distribution functions, safely depleted spacers were aggregated across the first or last 250 nt of genes.

For secondary structure analysis, we utilized the RNA accessibility model from Vienna RNAplfold (Bernhart et al., 2006). RNAplfold calculates through a moving average of RNA folds the probability that a region u of RNA is unpaired given its *cis* sequence context in a four-parameter model, where W is the moving average window length in nucleotides, L is the maximum permissible pairing distance between nucleotides in the window, and u_{start} and u_{end} are the start and end of the region u, respectively. To apply this model to our data, we separated spacers from our *E. coli* essential gene screen into training/testing cohorts of five or more, each represented by a unique permissible PFS and gene and containing at least one spacer in the top 2% of depleted spacers from the screen (to enhance predictive signal). We then randomly divided these cohorts into a training set (~80%) and a testing set (~20%). For optimizing a secondary structure-mediated model of efficient spacer design we selected as objective functions *top 1* or *top 3 accuracy*, the percent of cohorts for which the top spacer is accurately predicted or falls in the top 3 depleted spacers in a cohort, respectively. We optimized the two objective functions on the training dataset, first by fixing W and L while varying u_{start} and u_{end}, then by fixing u_{start} and u_{end} and varying W and L (Figure 4B). In the case of *bzcas13b* with *bzcsx27*, as well as that of *bzcas13b* alone, the optimized parameters were found to be approximately $W = 240$, $L = 180$, $u_{start} = 16$, and $u_{end} = 30$. We gauged the performance of this RNAplfold model relative to 10^6 Monte Carlo simulations performed on the testing dataset and found empirical P-values of less than 1e-2 for *top 1 accuracy*, and less than 1e-5 for *top 3 accuracy.* Similar predictive power applied to *pbcas13b* with *pbcsx28*, as well as to *pbcas13b* alone.

Kanamycin Validation Screen Experiment

A total of 160 kanamycin-targeting spacers was selected, 42 of which contain both PFS rules, 47 of which contain one rule, and 71 of which contain no rules, to which 162 non-targeting control spacers were added (Table S5). The library of spacers was cloned into

either a *bzcas13b* and *B. zoohelcum* direct repeat-spacer-direct repeat backbone or simply a *B. zoohelcum* direct repeat-spacer-direct repeat backbone containing a chloramphenicol resistance gene using Golden Gate Assembly (NEB) with 100 cycles, and then transformed over one 22.7cm x 22.7cm carbenicillin LB Agar plate. The two cloned library plasmids were then re-transformed with over a 22.7cm x 22.7cm chloramphenicol LB Agar plate or a 22.7cm x 22.7cm kanamycin-chloramphenicol LB Agar plate. Libraries of transformants were scraped from plates and DNA extracted using the QIAGEN Plasmid Plus Maxi Kit (QIAGEN). 100 ng of library DNA and 100 ng of pMAX-GFP (Lonza), containing a kanamycin resistance gene were added to 50 uL of chemically competent 10-beta cells (NEB) and transformed according to the manufacturer's protocol.

Kanamycin Validation Screen Analysis

Prepared DNA libraries were sequenced on a NextSeq (Illumina), with reads mapped to the input library of spacers. For normalizing the abundance of spacers of two separate clonings, the corrected experimental read abundance of a given spacer was calculated as the read abundance of that spacer in the *bzcas13b* plasmid (kanamycin-chloramphenicol transformation) multiplied by the ratio of the read abundance ratio of that spacer in the non-*bzcas13b* plasmid (chloramphenicol-only transformation) to the read abundance ratio of that spacer in the *bzcas13b* plasmid (chloramphenicol-only transformation).

MS2 Phage Drop Plaque Assay

Individual spacers for bacteriophage MS2 interference were ordered as complementary oligonucleotides containing overhangs allowing for directional cloning in between two direct repeat sequences in vectors containing *cas13b* (Table S2 and Table S6). 10 uM of each complementary oligo were annealed in 10X PNK Buffer (NEB), supplemented with 10mM ATP and 5 units of T4PNK (NEB). Oligos were incubated at 37°C for 30 min., followed by heating to 95°C for 5 min. and then annealed by cooling to 4°C. Annealed oligos were then diluted 1:100 and incubated with 25 ng of Eco31I digested *cas13b* vector in the presence of Rapid Ligation Buffer and T7 DNA ligase (Enzymatics). Individual plasmids were prepared using the QIAprep Spin Miniprep Kit (QIAGEN), sequence confirmed and then transformed into C3000 (ATCC 15597) cells made competent using the Mix & Go *E. coli* Transformation Kit (Zymo). In the case of experiments using *csx27* or *csx28*, C3000 cells harboring *csx* plasmids were made competent and then transformed with *cas13b* direct repeat-spacer-direct repeat plasmids. Following transformation, individual clones were picked and grown overnight at 37°C in LB containing the appropriate antibiotics. The following morning, cultures were diluted 1:100 and grown to an OD_{600} of 2.0 by shaking at 37°C with 5% C02 at 250 rpm, then mixed with 4mL of antibiotic containing Top Agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride, 5 g/L agar) and poured on to LB-antibiotic base plates. 10-fold serial-dilutions of MS2 phage (ATCC 15597-B1) were made in LB and then spotted onto hardened top agar with a multi-channel pipette. Plaque formation was assessed after overnight incubation of the spotted plates at 37°C. For assessing interference levels in Figures 5E and 7D, samples were blinded using a key and the lowest dilution of phage at which plaque formation occurred was compared to a pACYC condition by eye, where the lowest dilution of MS2 that formed plaques on pACYC was set to 1. The lowest dilution of phage used for Figure 5E was $1.05*10^8$ pfu.

DNA Interference Assay

A 34-nt target sequence consisting of a 30-nt protospacer and a permissive PFS (5′-G, 3′-AAA) was cloned into pUC19 in two locations (Table S2 and Table S6). For the transcribed target, the target sequence was cloned into the coding strand of the *bla* gene, in frame immediately after the start codon, with the G of the start codon serving as the 5′ PFS. For the non-transcribed target the identical target sequence (protospacer and PFS) were cloned into the AatII site of pUC19, so that the protospacer appears on the non-transcribed strand with respect to the pBla and pLac promoters. To determine interference, 25 ng of the ampicillin resistant target plasmid and 25 ng of the chloramphenicol resistant *bzcas13b* or empty vector (pACYC) were added to 5 uL of NovaBlue GigaSingle cells (Novagen). The cells were incubated for 30 min on ice, heatshocked for 30 s at 42°C and incubated on ice for 2 min. Then, 95 uL of SOC was added to cells and they were incubated with shaking at 37°C for 90 min, before plating the entire outgrowth (100 uL) on plates containing both chloramphenicol and ampicillin.

QUANTIFICATION AND STATISTICAL ANALYSIS

MS2 Interference Assay-HEPN Mutants

Three bioreplicates of the MS2 interference assay were performed for the fold resistance quantification in Figures 5E and 7D. For assessing interference levels in Figures 5E and 7D, samples were blinded using a key and the highest dilution of phage at which plaque formation occurred was compared to a vector only condition by eye, where the highest dilution of MS2 that formed plaques on pACYC was set to 1. The error bars are the standard deviation of the fold-resistance for each condition.

DNA Interference Assay

Three bioreplicates of the DNA interference assay were performed for the colony forming unit quantification. The mean values were taken from the mean of number of colony forming units from a standard colony forming unit count, and the standard deviation values accordingly from the same standard count.

E. coli Essential Gene Screen

Spacer depletions from the screen were calculated as the read abundance of a spacer in the empty vector condition divided by read abundance in each gene plasmid condition. Mean depletions over three bioreplicates were calculated. We imposed a two-step quality-control filter on the data: a maximum coefficient of variation of 0.2 for depletion over three bioreplicates, and a minimum spacer read abundance of 1/3*N* in each bioreplicate, where N = 55,700. This reduced the number of guides represented from N to approximately 30,000-40,000.

For secondary structure analysis, we utilized the RNA accessibility model from Vienna RNAplfold (Bernhart et al., 2006). To apply this model to our data, we separated spacers from our *E. coli* essential gene screen into training/testing cohorts of five or more, each represented by a unique permissible PFS and gene and containing at least one spacer in the top 2% of depleted spacers from the screen (to enhance predictive signal). We then randomly divided these cohorts into a training set (~80%) and a testing set (~20%), with the size of a testing set ranging from approximately n = 40 to n = 60, depending on the screen. For optimizing a secondary structure-mediated model of efficient spacer design we selected as objective functions *top 1* or *top 3 accuracy*, the percent of cohorts for which the top spacer is accurately predicted or falls in the top 3 depleted spacers in a cohort, respectively. We gauged the performance of this RNAplfold model relative to 10^6 Monte Carlo simulations performed on the testing dataset and found empirical *P*-values of less than 1e-2 for *top 1 accuracy*, and less than 1e-5 for *top 3 accuracy.* Similar predictive power applied to *pbcas13b* with *pbcsx28*, as well as to *pbcas13b* alone.

K_D Calculations

Gel shift of the RNA targets was quantified from an EMSA gel using ImageJ (Wayne Rasband, NIH) and plotted in GraphPad Prism version 7 (GraphPad Software, La Jolla California USA). Line regression was performed in Prism 7 using nonlinear fit with one-site binding hyperbola. K_D values are calculated by GraphPad Prism based on analysis of regression data (Table S7).

DATA AND SOFTWARE AVAILABILITY

Data Resources

Data have been deposited in the following resources:

Next-Generation Sequencing for bacterial RNA-sequencing, *E. coli* essential gene screen, kanamycin validation screen: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA358111