Characterization and application of type VI-B RNA-targeting CRISPR systems

by

David Benjamin Turitz Cox

B.S. Biology
Stanford University

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the
Massachusetts Institute of Technology

June 2018

Signature of Author: Signature redacted
Department of Biology
March 12th, 2018

Certified by: Signature redacted
Feng Zhang
Associate Professor of Biological Engineering and Brain and Cognitive Sciences
Thesis Supervisor

Accepted by: Signature redacted
Amy Keating
Professor of Biology
Chairman, Committee for Graduate Students

Characterization and application of type VI-B RNA-targeting CRISPR systems

by

David Benjamin Turitz Cox

Submitted to the Department of Biology on March 12th, 2018 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the Massachusetts Institute of Technology

Abstract:

The ability to modify nucleic acids is critical for establishing the role of genetic and transcribed elements in mediating biological phenotypes. Manipulating endogenous DNA sequences in eukaryotic genomes has been greatly aided by the advent of genome editing technologies that utilize programmable nucleases. DNA nucleases derived from class 2 CRISPR systems, which provide adaptive immunity in prokaryotes through cleavage of nucleic acids using a single, multi-domain, RNA-guided endonuclease, have been particularly useful in this regard because they enable targeting of new sites through simple Watson-Crick base pairing rules. Recent computational studies have uncovered the existence of predicted RNA-targeting class 2 CRISPR systems, suggesting that the power of genome editing techniques might be extended to the level of transcripts.

In this thesis, I present work describing the discovery and characterization of a new RNA-targeting class 2 CRISPR system: type VI-B. Using a combination of biochemistry and bacterial genetics, we demonstrated that the predicted nuclease of the VI-B system, Cas13b, is an RNA-guided RNase, whose activity can be modulated by the *csx* genes that often appear in genetic proximity to *cas13b*.

Next, we characterized the behavior of Cas13b and the related enzymes Cas13a and Cas13c in mammalian cells, identifying orthologs of Cas13a and Cas13b with specific RNA interference activity in mammalian cells.

Finally, we showed that catalytically inactive versions of a Cas13b ortholog can direct adenosine-to-inosine deaminase activity to transcripts in human cells when fused to the catalytic domain of ADAR2. Using structure-guided mutagenesis, we created a high-specificity version of this system that can be utilized in research or potentially therapeutic contexts. The description of a Cas13b ortholog that can be used to knockdown or recruit RNA-modifying domains to transcripts in mammalian cells suggests the utility of this technology to interrogate and modify transcript function in diverse contexts.

Thesis Supervisor: Feng Zhang
Title: Associate Professor, Departments of Brain and Cognitive Sciences and Biological Engineering, MIT.

Acknowledgments

Deciding to join the lab of my thesis advisor, Feng Zhang, was one of the best decisions I've made in life. Scientifically, Feng has taught me how to think clearly about problems and what it takes to bring a scientific vision to life. More importantly, I've learned from Feng how to be both a strong and kind person; that excellence does not have to come at the expense of others, or oneself. His deep passion for doing great science will always be an inspiration to me. I can't think of someone else that I would rather have had to advise me during this critical period of my academic and personal development.

The Zhang lab is composed of an extremely talented group of individuals. I am grateful to all of them for providing a stimulating academic environment and friendship during graduate school. I would like to specifically thank my co-authors with whom much of this work was done: Neena Pyzocha, Aaron Smargon, Omar Abudayyeh and Jonathan Gootenberg. Erin Blackwell, Rich Belliveau and Rhiannon MacCrae have been immeasurably helpful in keeping the lab running smoothly and being a source of support for me during challenging times—thank you.

I am grateful to my committee for their support and advice throughout my graduate career: David Sabatini, David Bartel and David Liu. It is truly remarkable to be able to assemble three of the top scientists in the world to evaluate your work and give you thoughtful advice, and I am indebted to them for their time and effort spent on my young career.

I am also a part of other academic programs in the Boston area: HST and the MD-PhD program of Harvard Medical School. I would like to thank them for their continued support.

My close friends from college and childhood have been a tremendous source of support for me during graduate school: to Christian, Jake, Jonathan, Chou, Pradeep, Rohan, Girmay, Brett and Aaron, it's all love. I've made a lot of new friends during my time in Boston/Cambridge, there are too many to name here, but I want to say thank you for your friendship. I'd also like to thank my pick-up basketball friends, whose competitive spirit is always refreshing to be around.

Finally, I'd like to thank my family. To my parents, Carl and Gilda, thank you for raising me and loving me always. To my brothers Evan and Matt, thank you for making me who I am. And to my newest (and some soon-to-be) family members, Laura, Maddy, Ag, and Oliver, I'm so happy you guys are a part of my life. Also, to my future in-laws, Candi and Michael Men, thank you for welcoming me into your family. Last, I would like to thank my fiancé, Clara. There are some things too precious to put on paper. How much I love you is one of those things. I can't wait to start our life together!

Table of Contents

Introduction

Understanding the relationship between genes and phenotypes is a long-standing goal of biological research. Establishing these relationships requires characterizing phenotypes empirically and manipulating genetic substrates directly to establish causality. Although the former has been the subject of study for centuries, the latter has lagged behind. Efforts to link gene sequences to phenotypes have been recently aided by the advent of next-generation sequencing technologies (*1-3*), leading to the creation of large gene-variant datasets, from which phenotypic relationships can be hypothesized. Increased sequencing power has been particularly valuable for human genetics, where mapping of the human genome (*4, 5*) in combination with sequencing data from patients has established the role of genetic variation in human health: of the 20,000 genes in the human genome, approximately 3,800 have a mutation that is currently linked to disease (*6*).

However, despite increased powers of observation, the ability to make genetic perturbations to establish the causality and mechanism of potential disease-causing mutations, or to treat disease, has grown more slowly. Additionally, with the appreciation of the diverse roles that RNA can play in mediating disease (*7-9*), there is a growing need for the ability to make perturbations directly to transcripts. Below, I review the development of technologies that allow for targeted, programmable perturbations of nucleic acids in complex eukaryotic genomes, beginning with methods to make changes to DNA sequence, followed by a description of CRISPR-Cas systems that have improved these techniques, and conclude with a description of RNA-targeting CRISPR-Cas systems that hold similar promise for manipulating transcripts.

Development of genome editing techniques for eukaryotes

Genome editing is broadly defined as the creation of targeted sequence changes at defined positions in cellular genomic DNA. Eukaryotic genome editing derives from seminal studies in yeast, where it was found that introducing an exogenous DNA vector

containing homology to a genomic locus led to the incorporation of vector DNA into the native locus (*10-13*). This phenomenon required Rad52 (*13*), suggesting that this process was possibly mitotic homology-directed repair (HDR), a process normally utilized to repair DNA double-strand breaks (DSBs) during DNA replication. This led to the 'gene-targeting' method, where a targeted genetic change could be introduced into by creating a targeting vector with the desired mutation flanked by homologous sequences to the genomic locus (*11*).

Gene-targeting was extended to mammalian pluripotent cells in a series of landmark studies by initiated by Mario Cappechi, leading to the creation of animal models with targeted genetic mutations, enhancing our understanding of mammalian gene function, and raising the possibility of manipulating endogenous genetic sequences for therapy (*14-18*). Despite the utility of gene-targeting techniques, the frequencies of successful gene-targeting events in mammalian cells by introduction of a vector encoding the desired change were low (1 in 10^5 or less); and thus the use of creative selection strategies were often required to identify clones carrying the desired modification (*17, 19*). Although selection strategies were refined and improved, it was clear that a more efficient method for targeted genetic manipulation was needed to increase frequencies for more rapid generation of animal models and to explore the possibility of this approach for genetic therapy.

Studies of DSB repair pointed towards a strategy to improve the rate of HDR and thus gene-targeting, eventually leading to the next generation of genome editing technologies. In yeast, it was observed that insertion of the target site of the naturally occurring endonucleases HO and Isce-I into ectopic locations stimulated HDR at target sites (*20-23*). Similarly, studies of the mobile P element of Drosophila, which is thought to create DSBs during mobilization, showed that introduction of synthetic DNA molecules with homology to the P-element excision site resulted in their incorporation (*24*). Jasin and colleagues also provided evidence that DSBs were recombinogenic in mammalian cells,

by introducing yeast nuclease target sites into extrachromosomal substrates or endogenous loci and demonstrating that cleavage of these elements enhanced recombination rates (*25-27*).

By demonstrating that DSBs could be used to stimulate HDR at defined locations in eukaryotic genomes, these studies implied that creating DSBs near desired sites of gene-targeting might improve the efficiency of modification. Creating targeted DSBs in mammalian cells would also provide an additional functionality to gene-editing technologies. DSBs in mammalian cells can also be resolved by non-homologous mechanisms, especially in the absence of a homologous template, often through the Non-homologous end joining (NHEJ) pathway (*28-30*) (Figure 1). NHEJ resolution of DSBs typically produces a heterogeneous mixture of insertions or deletions (indels) upon repeated activation by a nuclease (*31*), leading to inactivating frameshift mutations if targeted to protein coding sequences. Thus, the ability to create targeted DSBs would provide two advantages for genetic perturbation of endogenous genes: introduction of precise mutations at increased frequencies via HDR, and imprecise, likely disabling genetic perturbations via NHEJ (Figure 1). Testing whether these predicted benefits of targeted DSB formation were true necessitated a technology that could introduce DSBs at arbitrary, user-defined locations in complex mammalian genomes.

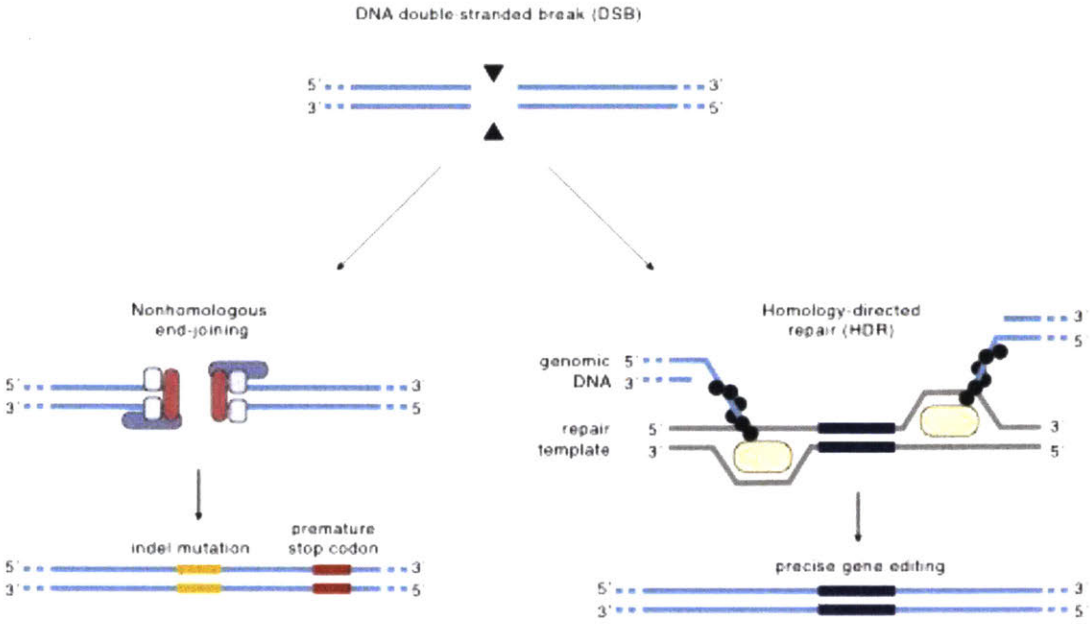

Figure 1: Schematic of the major DNA DSB repair pathways in mammalian cells, adapted from Hsu *et. al.* (*32*).

The first successful approach to this problem was to combine the sequence-agnostic nuclease domain of the *Fok* I type II restriction endonuclease with modular, sequence-specific C_2H_2 zinc finger DNA binding domains (Figure 2A) (*33-35*). C_2H_2 zinc fingers (ZFs) are a class of DNA binding domains found in certain eukaryotic transcription factors (*36, 37*). Structural studies of naturally occurring tandem ZFs suggested that each 30 amino-acid finger coordinates one Zn^+ atom using two cysteine and histidine residues, adopts a β β α conformation and contacts 3 base-pairs of DNA via major groove interactions (*38*). Importantly, the fingers appeared to bind each 3 base-pair target site independently using side-chain interactions from the α helix. The modular structure of ZFs suggested that individual fingers could be arbitrarily combined in a new order to specify target sequences. Initial evidence supporting the possibility of designer DNA

nucleases was from *in vitro* studies by Chandrasegaran, Berg and colleagues, demonstrating that fusion of the *Fok* I nuclease domain to synthetic ZFs resulted in DSBs controlled by predicted ZF specificities (*33, 39*). These proteins, termed zinc finger nucleases (ZFNs) were shown to operate most efficiently with two ZFNs being bound adjacent to a target sequence, likely due to the requirement for *Fok* I dimerization for cleavage (*35, 40*). ZFNs were soon shown to promote genome editing in *Xenopus laevis* oocytes (*35*), *in vivo* in *D. melanogaster* (*31*) and in human cells (*41*) . The studies of ZFN activity *in vivo* confirmed the predicted outcomes of targeted DSB creation at chromosomal loci: targeted mutagenesis in the absence of a homologous template (*31*), and increased rates of HDR at the DSB site in the presence of a gene-targeting vector (*35, 41*). These seminal studies established the field of genome editing by increasing the rates and ease of endogenous genetic modification, enabling selection-free genetic perturbations. Furthermore, by creating a synthetic protein with separable nuclease and DNA binding domains, it was possible to replace the nuclease activity with other effector domains, such as transcriptional effector domains (*42*) and chromatin modifying enzymes (*43*). This array of tools added new modalities for testing and manipulating endogenous gene function.

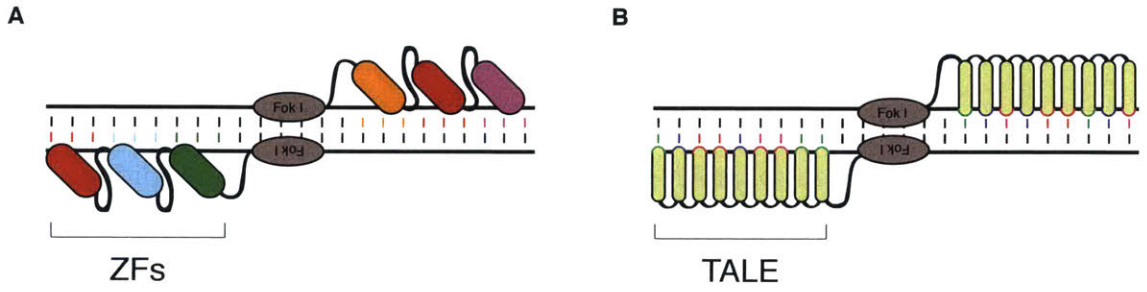

Figure 2: Schematic representation of ZF (A) and TALE (B) based nucleases for genome editing. By attaching tandem repeats of ZF or TALE modules that bind to specific DNA sequences to the sequence agnostic *Fok* I DNA nuclease domain, depicted in brown, site-specific DNA cleavage can be achieved. *Fok* I operates most effectively as a dimer, requiring two ZF/TALE-*Fok* I molecules to be bound near a target DNA sequence for effective cleavage.

Optimizations to initial ZFN designs improved genome editing capabilities. Construction of longer ZF arrays allowed for recognition of target sites long enough to specify unique target sites in complex mammalian genomes (*44, 45*) and creation of *Fok* I obligate heterodimers minimized the risk of off-target mutagenic events (*46, 47*). However, targeting ZFs to new sequences remained a challenge. The modular structure of ZFs suggested that evolving individual fingers to recognize each of the 64 possible 3-base combinations and combining them in an appropriate order could create ZFs with arbitrary specificity. Although conceptually appealing, evolution of individual finger specificity proved challenging and some ZF domains were found not to be strictly modular, making design of new ZFNs a non-intuitive, labor-intensive task (*48, 49*), a restriction that has limited their scalability. Nevertheless, ZFNs established the power of genome editing techniques for mammalian cells, showing the first possibility of its potential for therapeutics and genetics.

The challenge of designing ZFs with altered specificity highlighted the need for genome editing technologies with better modularity and a predictable nucleic-acid recognition code. Transcription activator like effectors (TALEs) derived from the bacterial plant pathogen *Xanthomonas sp.* provided these qualities and improved early editing technologies significantly. TALEs are proteins naturally secreted by *Xanthomonas* that modulate host gene expression to promote successful colonization (*50-53*). The DNA-binding domains of TALEs are composed of an array of 33 or 34 amino-acid long peptide repeats, each of which recognizes a single-base pair using two hypervariable amino acids, known as the repeat-variable di-residues (RVDs) (*54*), leading to a simple correspondence between amino acid sequence and base-recognition (*55, 56*). Likely due to their natural existence as long arrays of repeat monomers, TALEs proved to be more modular than ZFs and capable of recognizing target sites that could specify unique sequences in complex

genomes. TALEs were combined with the *Fok* I nuclease domain to make TALE nucleases (TALENs), which were shown to mediate genome editing outcomes with rates comparable to ZFNs in eukaryotic cells (Figure 2B) (*57-63*). The relative simplicity of identifying TALEs with new target specificities lead to significant excitement about their utility in the academic community.

TALEs still presented a technical challenge, however: the highly repetitive nature of the monomeric repeat domains made molecular cloning of new constructs challenging. New cloning methods partially addressed this issue (*64, 65*), but genome editing technologies that could be more easily be programmed to target new sequences were still needed. The development of RNA-guided genome editing by reconstituting the function of CRISPR-Cas systems in heterologous contexts overcame this challenge, leading to widespread use of genome editing technologies.

CRISPR-Cas systems: RNA-guided endonucleases that mediate adaptive immunity in prokaryotes

Clustered regularly interspaced short palindromic repeat (CRISPR) loci function as adaptive immune systems in prokaryotes (*66-73*). The study of CRISPR biology began unexpectedly in 1987 when Nakata and colleagues described a series of five highly homologous, 29-nt repeat sequences (direct-repeats) separated by four 32-nt non-homologous sequences (spacers) in their study sequencing the *iap* gene from *Escherichia coli*, the first example of a CRISPR array that would eventually be shown to mediate RNA-guided immunity (*74*). Additional sequencing studies uncovered genetic elements with similar structures in other bacteria and archaea (*70-72*), eventually leading to their recognition as a distinct family of prokaryotic repeat elements and a unified name for such systems: CRISPR (*75*). It was soon appreciated that CRISPR arrays did not exist as individual elements, but were often associated with adjacent ORFs predicted to encode

products that interacted with nucleic acids, termed CRISPR-associated (Cas) genes, which together form CRISPR systems (Figure 3) (*75*).

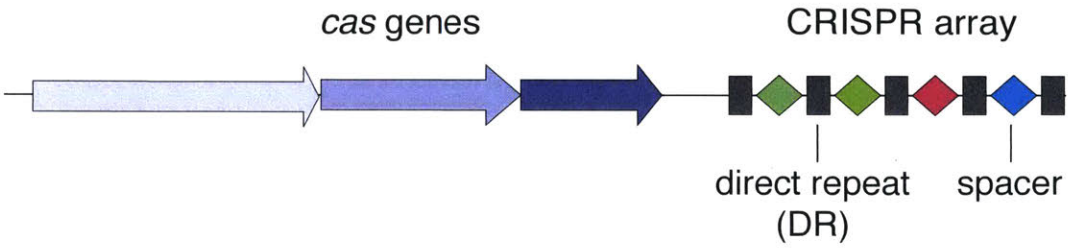

Figure 3: Schematic representation of CRISPR loci, which are composed of *cas* genes in proximity to a CRISPR array containing constant direct repeat sequences separated by hypervariable spacer sequences.

The function of these loci remained mysterious until bioinformatics studies lead to the discovery that spacers within the CRISPR array were homologous to genetic elements of extrachromosomal origin (*e.g.* bacteriophages and conjugative plasmids), raising the possibility of CRISPR as a genetic defense system (*68, 69, 73*). In 2007, Barrangou and colleagues provided direct experimental evidence that CRISPR loci encoded genetic immunity to mobile genetic elements (*66*). Using a naturally occurring CRISPR locus from *Streptococcus thermophilus*, it was shown that: (*i*) CRISPR loci integrated new spacers with homology to infecting bacteriophages in response to infection, (*ii*) spacer sequence content dictated bacteriophage resistance and (*iii*) these phenomena genetically required neighboring *cas* genes. Detailed biochemical work soon provided greater molecular detail on the mechanism by which spacer sequences mediated immunity. Using a naturally occurring CRISPR system from *E. coli*, Van der Oost and colleagues showed that the transcribed CRISPR array was processed into smaller units containing individual spacers and partial DRs by Cas proteins *in vitro* (*67*). Furthermore, once processed, the

individual spacer containing mature CRISPR RNAs (crRNAs) remained bound to the Cas enzyme complex. The Cas enzyme complex that bound the crRNAs was genetically necessary for interference, supporting a model of RNA-guided immunity. In this work, the authors also created synthetic CRISPR arrays, to re-direct immunity against the dsDNA lambda (λ) phage. CRISPR arrays predicted to generate crRNAs that hybridized to both the template and non-template strand of the virus mediated immunity, showing for the first time that CRISPR activity could be programmed, and likely acted on DNA.

It is now understood that adaptive immunity at CRISPR loci occurs in a three-step process (Figure 4). First, invading nucleic acids are inserted into the CRISPR array in a process called adaptation (*66*). Second, the CRISPR array is transcribed and processed into crRNAs containing single spacers in a process called crRNA biogenesis (*67, 76, 77*). Finally, interference against foreign genetic elements is mediated through the action of Cas nucleases, which in complex with crRNAs find and cleave their targets through Watson-Crick basepairing of the Cas ribonucleoprotein complex with target sequences (*67, 78-80*).

Although adaptation, crRNA biogenesis and interference are common to all CRISPR systems, the specific mechanisms by which individual loci execute crRNA biogenesis and interference can vary significantly. These mechanistic differences have been used to designate the class and type of the CRISPR system in question.

CRISPR systems are broadly divided into two classes based on the number of *cas* genes that mediate interference and can be further sub-divided into types and subtypes on the basis of signature *cas* gene content, sequence homology, and locus architecture (*81*). Class 1 systems utilize multiple *cas* gene products assembled in complexes to degrade target substrates, whereas class 2 systems utilize a single, large, multi-domain protein to achieve interference (Figure 4) (*78, 82, 83*). The simplicity of class 2 systems, where the specificity of a single nuclease can be reprogrammed through changing a short sequence of RNA, has facilitated their development into genome editing technologies.

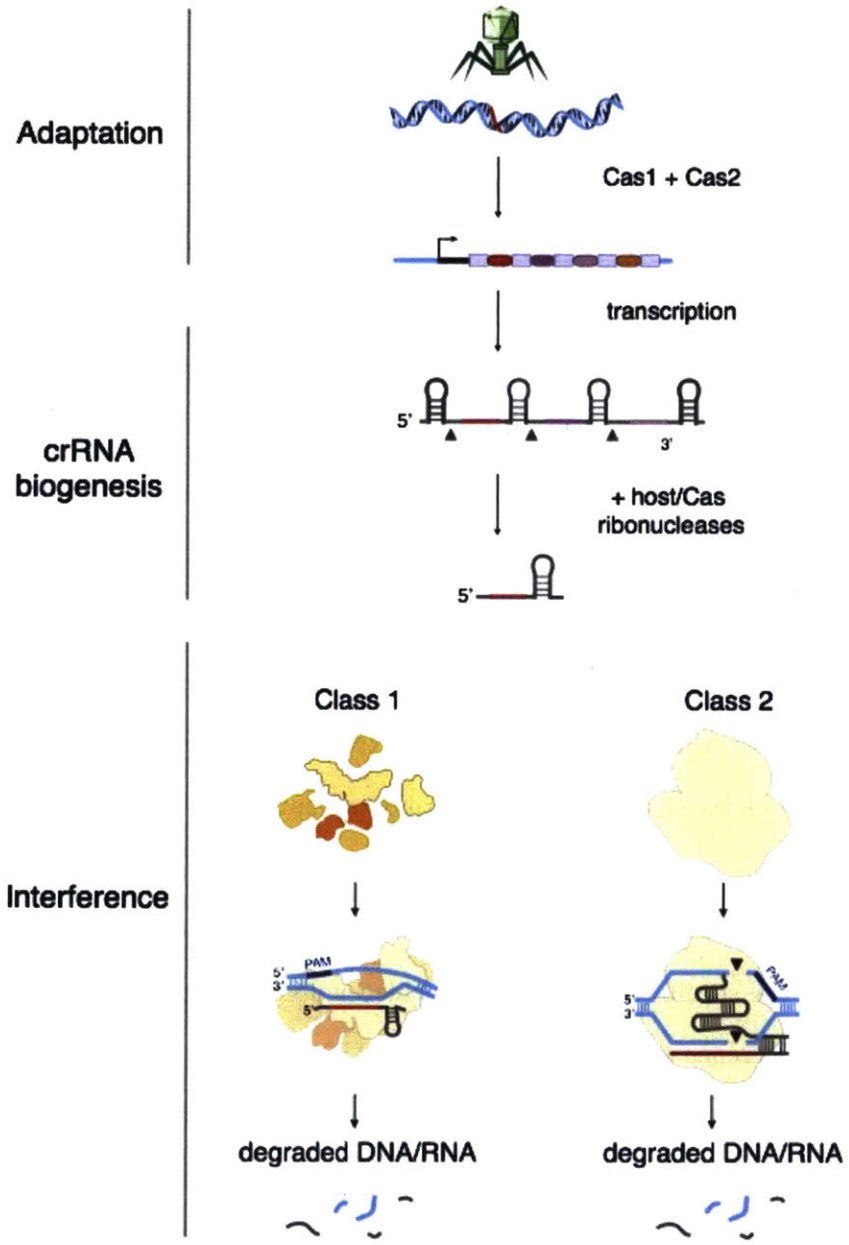

Figure 4: Mechanisms of nucleic-acid interference by CRISPR systems. Cas1 and Cas2, sometimes in combination with other Cas proteins mediate insertion of DNA derived from mobile genetic elements into the CRISPR array, in a process called adaptation. The CRISPR array is transcribed as a long primary transcript and then processed into individual spacer-containing units (crRNAs) by Cas ribonucleases that occasionally are

aided by host enzymes. crRNAs direct Cas nuclease activity through a complementarity-dependent mechanism. Class 1 CRISPR systems utilize multi-protein, Cas nuclease containing complexes for target cleavage, whereas class 2 systems utilize a Cas single nuclease for nucleic-acid interference. Adapted from Hsu *et. al.* (*32*).

Genome editing using class 2 CRISPR systems

Class 2 DNA-targeting CRISPR systems are advantageous compared to other genome editing tools because they provide a method of targeting DSBs to endogenous loci through simple Watson-Crick base pairing rules. The first effector domain from a class 2 CRISPR system to be harnessed for genome editing in mammalian cells was Cas9 from type II CRISPR loci (Figure 5) (*84, 85*). Cas9 endonuclease activity is dependent on two non-coding RNAs: the crRNA, which directs cleavage specificity and a trans-activating CRISPR RNA (tracrRNA) that promotes crRNA maturation and is also required for cleavage (Figure 5A) (*86*). Studies showed that *in vitro* (*87, 88*) and in prokaryotic cells (*78, 83, 86*), crRNA, tracrRNA and Cas9 are necessary for cleavage of complementary dsDNA targets, suggesting that reconstitution of Cas9 activity in a heterologous system could be achieved by expressing Cas9, the tracrRNA and a targeting crRNA. Notably, it was also shown that the crRNA and tracrRNA could be fused into a single RNA molecule that mediated *in vitro* cleavage, potentially reducing the complexity of the system further (*87*).

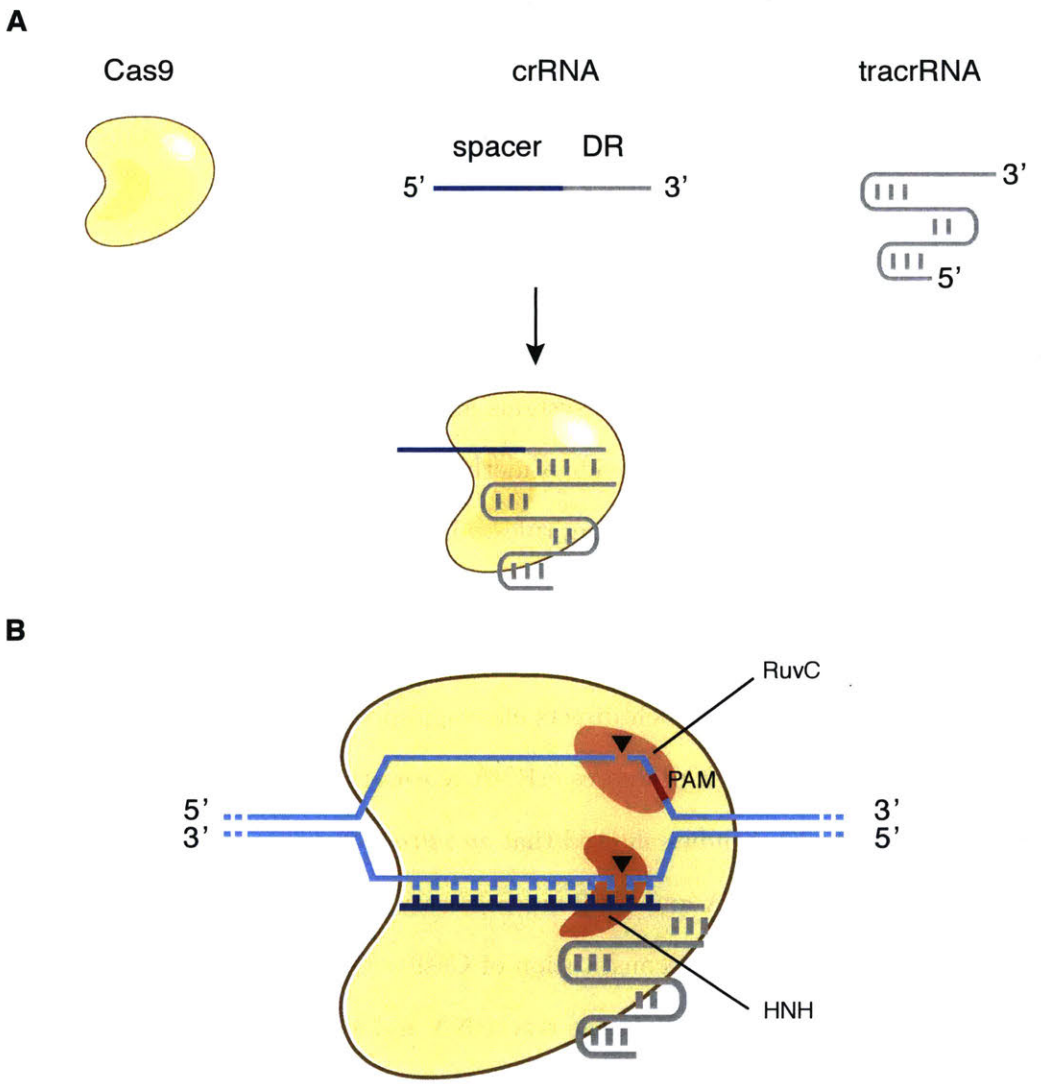

Figure 5: (A) Schematic representation of Cas9, crRNA and tracrRNA that together form a ribonucleoprotein complex with DNA cleavage activity. **(B)** The HNH and RuvC nuclease domains of Cas9 cleave the complementary and non-complementary DNA strands of a crRNA-specified target site adjacent to a compatible PAM, creating a DSB.

Recognition of a target sequence by Cas9 also requires a compatible protospacer-adjacent motif (PAM) (*68, 89-91*), a short nucleotide sequence immediately adjacent to

the target that Cas9 recognizes through protein-DNA contacts (*92*). PAM recognition initiates unwinding of the nearby dsDNA by Cas9, allowing for crRNA hybridization (*93*). Once the crRNA is hybridized to the target sequence, Cas9 uses two separate DNA nuclease domains, HNH and RuvC, to cleave the complementary and non-complementary DNA strands, respectively, 3 nucleotides from the PAM, creating a DSB within the crRNA:target heteroduplex (Figure 5B) (*87, 88*).

The targeting mechanism of Cas9 suggested this system could be utilized for editing complex mammalian genomes: the PAM of Cas9 is short (~3-5 bp) (*90, 91*), allowing for flexible target site selection, and the spacer sequence of the mature crRNA is sufficiently long (~20 nt) (*86*) to specify unique target sites. DNA endonuclease activity of Cas9 from *Streptococcus pyogenes* SF370 was reconstituted in mammalian cells in 2013, defining a new class of RNA-guided genome editing tools (*84, 85*).

Creating DSBs with Cas9 through Watson-Crick basepairing rules allowed for simple, more predictable targeting of genomic sequences and the widespread adoption of Cas9 as a genome editing tool. The flexible nature of Cas9 targeting has led to the development of new genetic-screening methodologies for human cells (*94-96*), the development of catalytically inactive variants that can be combined with other effector domains (*97-99*) and paved the way for additional RNA-guided genome editing technologies based on novel class 2 CRISPR effectors.

Extending genome editing techniques to RNA

Initial studies of class 2 CRISPR systems suggested that these loci exclusively encoded DNA nucleases, an idea that has been challenged by recent computational studies. By extending the types of features used to search for new CRISPR loci, novel class 2 systems predicted to encode single-effector, RNA-guided, RNA-targeting Cas nucleases were discovered (Figure 6) (*100*). The RNA-targeting ability of these loci is predicted due to the absence of DNA targeting domains and the presence of conserved Higher

Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) motifs, which mediate RNA degradation in a variety of contexts (*101*).

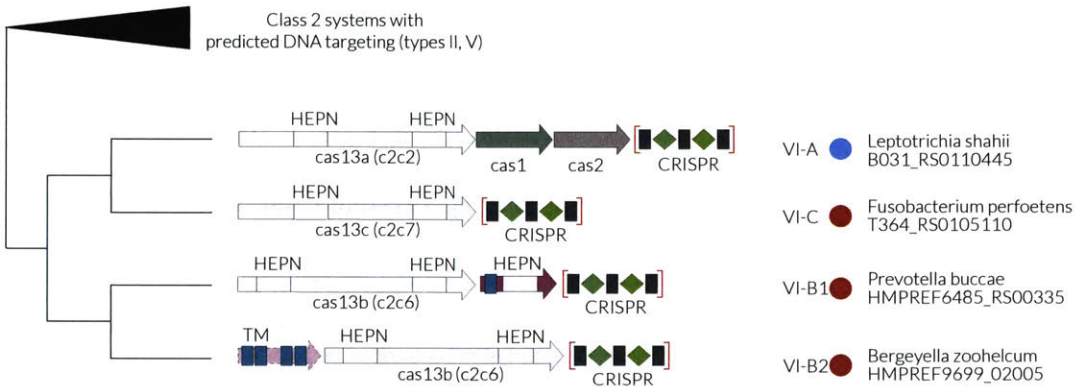

Figure 6: The phylogenetic relationship of class 2 CRISPR systems. All class 2 systems encoding predicted DNA nucleases are collapsed.

The discovery of putative class 2 RNA-targeting CRISPR systems opens the possibility of extending the simplicity of RNA-guided genome editing to transcripts. Transcriptome editing would ideally apply the functionalities of genome editing to RNA, with simple methodologies to both knockdown and modify the function of transcripts. Powerful technologies for suppressing the function of transcripts in eukaryotic cells exist, including RNA interference (RNAi) (*102, 103*) and anti-sense oligonucleotides (ASOs) (*104, 105*), which trigger enzymatic degradation by endogenous enzymes (RNAi and ASOs) or sterically occlude functional RNA motifs (ASOs) to prevent RNA function (*106*). Technologies that can modify the function of transcripts have been more challenging to develop. These techniques generally mimic the concept of TALE and ZF-based effectors for DNA applications: fusing tandem repeats of a modular RNA-binding protein domain that recognizes RNA bases through protein contacts, to RNA-modifying domains of known function (*107*). The RNA-binding domain commonly used for this application is the pumilio repeat (PUF) domain, which structural studies suggest may recognize single RNA

bases as part of a tandem array (*108, 109*). Similar to ZFs, developing PUFs with arbitrary specificity has proven challenging, possibly due to unanticipated complexity of the binding code (*110, 111*), limiting the prospects of recruiting RNA-modifying domains to endogenous transcripts.

Development of class 2 RNA-targeting CRISPR systems for transcriptome engineering could enable a simple method for modifying transcript function through recruitment of RNA-modifying enzymes fused to Cas effectors by Watson-Crick basepairing rules.

Class 2 CRISPR systems containing HEPN domains have been designated as type VI and further divided into subtypes A, B and C (Figure 6) (*112*). The RNA-targeting ability of type VI-A CRISPR systems has been recently confirmed and shown to be catalytically mediated by conserved residues in the HEPN domain of the Cas13a protein encoded by these loci (*113*). Cas13a cleaves ssRNA targets in a crRNA-dependent manner that requires complementarity with the target, suggesting the possibility of CRISPR-based transcriptome engineering.

At the outset of this thesis, the existence and function of type VI-B and VI-C CRISPR systems were unknown. The work presented in this thesis seeks to characterize novel type VI CRISPR systems and test their utility for transcriptome editing applications in mammalian cells. There are three specific questions addressed in the work presented:

(*i*) **Do other type VI CRISPR systems exist?**

By computationally mining publicly available microbial genomes using a novel CRISPR discovery pipeline, we identified type VI-B CRISPR loci, which encode a putative RNA-guided RNAse, Cas13b.

(*ii*) **How do type VI-B CRISPR loci function?**

Using *in vitro* assays and reconstitution of native VI-B loci in heterologous prokaryotic expression systems, we provide evidence for a model in which a ribonucleoprotein complex composed of Cas13b and a crRNA are sufficient to cleave

ssRNA targets. We show genetically that the phenomenon of RNA interference by VI-B systems *in vivo* can be repressed or enhanced by the expression of *csx27* or *csx28*, respectively, small effector proteins that naturally co-occur with Cas13b in type VI-B loci.

(*iii*) **Can nucleases from type VI CRISPR systems be used for transcriptome editing applications in mammalian cells?**

To test for the ability of type VI nucleases (Cas13 a/b/c) to suppress transcript function, we reconstituted their nuclease activity in mammalian cells. To address whether Cas13 nucleases could be used to modify transcript function, we created catalytically inactive variants of Cas13b (dCas13b) from *Prevotella sp. P5-125* fused to the catalytic domain of adenosine deaminase acting on RNA (ADAR) enzymes, showing that dCas13b could specifically direct the enzymatic activity of the ADAR deaminase domains to target transcripts.

The first two questions are addressed in the 2nd chapter of this thesis, and the final question in the 3rd chapter.

Note: A portion of the description of CRISPR-Cas systems in this introduction was adapted form my preliminary exam proposal.

References

1. P. Nyren, A. Lundin, Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical biochemistry* **151**, 504 (Dec, 1985).
2. E. D. Hyman, A new method of sequencing DNA. *Analytical biochemistry* **174**, 423 (Nov 1, 1988).
3. J. Shendure *et al.*, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728 (Sep 9, 2005).
4. E. S. Lander *et al.*, Initial sequencing and analysis of the human genome. *Nature* **409**, 860 (Feb 15, 2001).
5. J. C. Venter *et al.*, The sequence of the human genome. *Science* **291**, 1304 (Feb 16, 2001).
6. OMIM Gene Map Statistics. (2018).
7. T. A. Cooper, L. Wan, G. Dreyfuss, RNA and disease. *Cell* **136**, 777 (Feb 20, 2009).
8. D. S. Cabianca *et al.*, A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819 (May 11, 2012).
9. A. Jain, R. D. Vale, RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243 (Jun 8, 2017).
10. A. Hinnen, J. B. Hicks, G. R. Fink, Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 1929 (Apr, 1978).
11. R. J. Rothstein, One-step gene disruption in yeast. *Methods in enzymology* **101**, 202 (1983).
12. J. W. Szostak, R. Wu, Insertion of a genetic marker into the ribosomal DNA of yeast. *Plasmid* **2**, 536 (Oct, 1979).
13. T. L. Orr-Weaver, J. W. Szostak, R. J. Rothstein, Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 6354 (Oct, 1981).
14. K. Folger, K. Thomas, M. R. Capecchi, Analysis of homologous recombination in cultured mammalian cells. *Cold Spring Harbor symposia on quantitative biology* **49**, 123 (1984).

15. K. R. Thomas, K. R. Folger, M. R. Capecchi, High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419 (Feb 14, 1986).
16. K. R. Thomas, M. R. Capecchi, Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503 (Nov 6, 1987).
17. M. R. Capecchi, Altering the genome by homologous recombination. *Science* **244**, 1288 (Jun 16, 1989).
18. O. Smithies, R. G. Gregg, S. S. Boggs, M. A. Koralewski, R. S. Kucherlapati, Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230 (Sep 19-25, 1985).
19. S. Thompson, A. R. Clarke, A. M. Pow, M. L. Hooper, D. W. Melton, Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* **56**, 313 (Jan 27, 1989).
20. N. Rudin, E. Sugarman, J. E. Haber, Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics* **122**, 519 (Jul, 1989).
21. J. A. Nickoloff, E. Y. Chen, F. Heffron, A 24-base-pair DNA sequence from the MAT locus stimulates intergenic recombination in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 7831 (Oct, 1986).
22. A. Ray, I. Siddiqi, A. L. Kolodkin, F. W. Stahl, Intra-chromosomal gene conversion induced by a DNA double-strand break in Saccharomyces cerevisiae. *Journal of molecular biology* **201**, 247 (May 20, 1988).
23. A. Plessis, A. Perrin, J. E. Haber, B. Dujon, Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451 (Mar, 1992).
24. N. Nassif, J. Penney, S. Pal, W. R. Engels, G. B. Gloor, Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and cellular biology* **14**, 1613 (Mar, 1994).
25. P. Rouet, F. Smih, M. Jasin, Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and cellular biology* **14**, 8096 (Dec, 1994).
26. P. Rouet, F. Smih, M. Jasin, Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 6064 (Jun 21, 1994).

27. F. Smih, P. Rouet, P. J. Romanienko, M. Jasin, Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic acids research* **23**, 5012 (Dec 25, 1995).
28. P. A. Jeggo, DNA breakage and repair. *Advances in genetics* **38**, 185 (1998).
29. J. H. Wilson, P. B. Berget, J. M. Pipas, Somatic cells efficiently join unrelated DNA segments end-to-end. *Molecular and cellular biology* **2**, 1258 (Oct, 1982).
30. J. K. Moore, J. E. Haber, Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and cellular biology* **16**, 2164 (May, 1996).
31. M. Bibikova, M. Golic, K. G. Golic, D. Carroll, Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169 (Jul, 2002).
32. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262 (Jun 5, 2014).
33. Y. G. Kim, J. Cha, S. Chandrasegaran, Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1156 (Feb 6, 1996).
34. L. Li, L. P. Wu, S. Chandrasegaran, Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4275 (May 15, 1992).
35. M. Bibikova *et al.*, Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and cellular biology* **21**, 289 (Jan, 2001).
36. G. P. Diakun, L. Fairall, A. Klug, EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature* **324**, 698 (Dec 18-31, 1986).
37. J. M. Berg, Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 99 (Jan, 1988).
38. N. P. Pavletich, C. O. Pabo, Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809 (May 10, 1991).
39. J. R. Desjarlais, J. M. Berg, Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 7345 (Aug 15, 1992).

40. J. Smith *et al.*, Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic acids research* **28**, 3361 (Sep 1, 2000).
41. M. H. Porteus, D. Baltimore, Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (May 2, 2003).
42. Y. Choo, I. Sanchez-Garcia, A. Klug, In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* **372**, 642 (Dec 15, 1994).
43. A. G. Rivenbark *et al.*, Epigenetic reprogramming of cancer cells via targeted DNA methylation. *Epigenetics* **7**, 350 (Apr, 2012).
44. Q. Liu, D. J. Segal, J. B. Ghiara, C. F. Barbas, 3rd, Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5525 (May 27, 1997).
45. R. R. Beerli, C. F. Barbas, 3rd, Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology* **20**, 135 (Feb, 2002).
46. J. C. Miller *et al.*, An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature biotechnology* **25**, 778 (Jul, 2007).
47. M. Szczepek *et al.*, Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature biotechnology* **25**, 786 (Jul, 2007).
48. C. L. Ramirez *et al.*, Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature methods* **5**, 374 (May, 2008).
49. J. C. Miller, C. O. Pabo, Rearrangement of side-chains in a Zif268 mutant highlights the complexities of zinc finger-DNA recognition. *Journal of molecular biology* **313**, 309 (Oct 19, 2001).
50. K. Gu *et al.*, R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* **435**, 1122 (Jun 23, 2005).
51. S. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* **318**, 648 (Oct 26, 2007).
52. J. Boch, U. Bonas, Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annual review of phytopathology* **48**, 419 (2010).
53. A. J. Bogdanove, S. Schornack, T. Lahaye, TAL effectors: finding plant genes for disease and defense. *Current opinion in plant biology* **13**, 394 (Aug, 2010).

54. C. M. Hopkins, F. F. White, S. H. Choi, A. Guo, J. E. Leach, Identification of a family of avirulence genes from Xanthomonas oryzae pv. oryzae. *Molecular plant-microbe interactions : MPMI* **5**, 451 (Nov-Dec, 1992).
55. J. Boch *et al.*, Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509 (Dec 11, 2009).
56. M. J. Moscou, A. J. Bogdanove, A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (Dec 11, 2009).
57. M. Christian *et al.*, Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757 (Oct, 2010).
58. T. Li *et al.*, TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic acids research* **39**, 359 (Jan, 2011).
59. T. Li *et al.*, Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic acids research* **39**, 6315 (Aug, 2011).
60. J. C. Miller *et al.*, A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29**, 143 (Feb, 2011).
61. P. Huang *et al.*, Heritable gene targeting in zebrafish using customized TALENs. *Nature biotechnology* **29**, 699 (Aug 5, 2011).
62. L. Tesson *et al.*, Knockout rats generated by embryo microinjection of TALENs. *Nature biotechnology* **29**, 695 (Aug 5, 2011).
63. D. Hockemeyer *et al.*, Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology* **29**, 731 (Jul 7, 2011).
64. T. Cermak *et al.*, Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research* **39**, e82 (Jul, 2011).
65. F. Zhang *et al.*, Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* **29**, 149 (Feb, 2011).
66. R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709 (Mar 23, 2007).
67. S. J. Brouns *et al.*, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960 (Aug 15, 2008).
68. A. Bolotin, B. Quinquis, A. Sorokin, S. D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551 (Aug, 2005).

69. F. J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**, 174 (Feb, 2005).
70. F. J. Mojica, C. Diez-Villasenor, E. Soria, G. Juez, Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology* **36**, 244 (Apr, 2000).
71. F. J. Mojica, C. Ferrer, G. Juez, F. Rodriguez-Valera, Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Molecular microbiology* **17**, 85 (Jul, 1995).
72. F. J. Mojica, G. Juez, F. Rodriguez-Valera, Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Molecular microbiology* **9**, 613 (Aug, 1993).
73. C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653 (Mar, 2005).
74. Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, A. Nakata, Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology* **169**, 5429 (Dec, 1987).
75. R. Jansen, J. D. Embden, W. Gaastra, L. M. Schouls, Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **43**, 1565 (Mar, 2002).
76. T. H. Tang *et al.*, Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7536 (May 28, 2002).
77. J. Carte, R. Wang, H. Li, R. M. Terns, M. P. Terns, Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* **22**, 3489 (Dec 15, 2008).
78. J. E. Garneau *et al.*, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67 (Nov 4, 2010).
79. L. A. Marraffini, E. J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843 (Dec 19, 2008).
80. B. Wiedenheft *et al.*, RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings*

of the National Academy of Sciences of the United States of America **108**, 10092 (Jun 21, 2011).

81. K. S. Makarova *et al.*, An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology* **13**, 722 (Nov, 2015).
82. K. S. Makarova *et al.*, An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology*, (Sep 28, 2015).
83. R. Sapranauskas *et al.*, The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research* **39**, 9275 (Nov, 2011).
84. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819 (Feb 15, 2013).
85. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823 (Feb 15, 2013).
86. E. Deltcheva *et al.*, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602 (Mar 31, 2011).
87. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816 (Aug 17, 2012).
88. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2579 (Sep 25, 2012).
89. H. Deveau *et al.*, Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology* **190**, 1390 (Feb, 2008).
90. P. Horvath *et al.*, Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of bacteriology* **190**, 1401 (Feb, 2008).
91. F. J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733 (Mar, 2009).
92. H. Nishimasu *et al.*, Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935 (Feb 27, 2014).
93. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62 (Mar 6, 2014).
94. O. Shalem *et al.*, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84 (Jan 3, 2014).
95. T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80 (Jan 3, 2014).

96. S. Konermann *et al.*, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583 (Jan 29, 2015).
97. S. Konermann *et al.*, Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472 (Aug 22, 2013).
98. P. Mali *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833 (Sep, 2013).
99. I. B. Hilton *et al.*, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**, 510 (May, 2015).
100. S. Shmakov *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell* **60**, 385 (Nov 5, 2015).
101. V. Anantharaman, K. S. Makarova, A. M. Burroughs, E. V. Koonin, L. Aravind, Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biology direct* **8**, 15 (2013).
102. A. Fire *et al.*, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806 (Feb 19, 1998).
103. S. M. Elbashir *et al.*, Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494 (May 24, 2001).
104. P. C. Zamecnik, M. L. Stephenson, Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 280 (Jan, 1978).
105. M. L. Stephenson, P. C. Zamecnik, Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 285 (Jan, 1978).
106. A. Khvorova, J. K. Watts, The chemical evolution of oligonucleotide therapies of clinical utility. *Nature biotechnology* **35**, 238 (Mar, 2017).
107. J. P. Mackay, J. Font, D. J. Segal, The prospects for designer single-stranded RNA-binding proteins. *Nature structural & molecular biology* **18**, 256 (Mar, 2011).
108. X. Wang, J. McLachlan, P. D. Zamore, T. M. Hall, Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501 (Aug 23, 2002).
109. T. A. Edwards, S. E. Pyle, R. P. Wharton, A. K. Aggarwal, Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell* **105**, 281 (Apr 20, 2001).

110. M. T. Miller, J. J. Higgin, T. M. Hall, Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nature structural & molecular biology* **15**, 397 (Apr, 2008).

111. Y. Wang, L. Opperman, M. Wickens, T. M. Hall, Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20186 (Dec 1, 2009).

112. A. S. Sergey Shmakov, David Scott, David Cox, Neena Pyzocha, Winston Yan, Omar O. Abudayyeh, Jonathan S. Gootenberg, Kira S. Makarova, Yuri I. Wolf, Konstantin Severinov, Feng Zhang, Eugene V. Koonin2, Diversity and evolution of Class 2 CRISPR-Cas systems *Nature Reviews Microbiology*, (2016).

113. O. O. Abudayyeh *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (Aug 05, 2016).

Chapter 2

Adapted from the following article, which appears in *Molecular Cell*:

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, Massachusetts 02142, USA

[2] McGovern Institute for Brain Research at MIT
Cambridge, Massachusetts 02139, USA

[3] Departments of Brain and Cognitive Science and Biological Engineering
Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4] Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[5] Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School,
Boston, Massachusetts 02115, USA

[6]Department of Biology, Massachusetts Institute of Technology, Cambridge,
Massachusetts 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 should be addressed to zhang@broadinstitute.org.

Abstract

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 or 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 (*c*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats and *C*RISPR-*a*ssociated 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; Staals et al., 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). Fully assembled microbial genomes were searched for all proteins within 10kb 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 700aa (to eliminate Class 1 system false positives), presence of a putative single effector of size 900aa to 1800aa (informed by the size distribution of previously classified Class 2 effectors), and absence of *cas1* and *cas2* genes within 10kb of the CRISPR array (Method Details). Candidate effectors were grouped into families according to homology (Camacho et al., 2009; Hildebrand et al., 2009; Remmert et al., 2012), 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 10 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 (~1100aa) candidate effector protein and, in about 80% of the cases, an additional small (~200aa) 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).

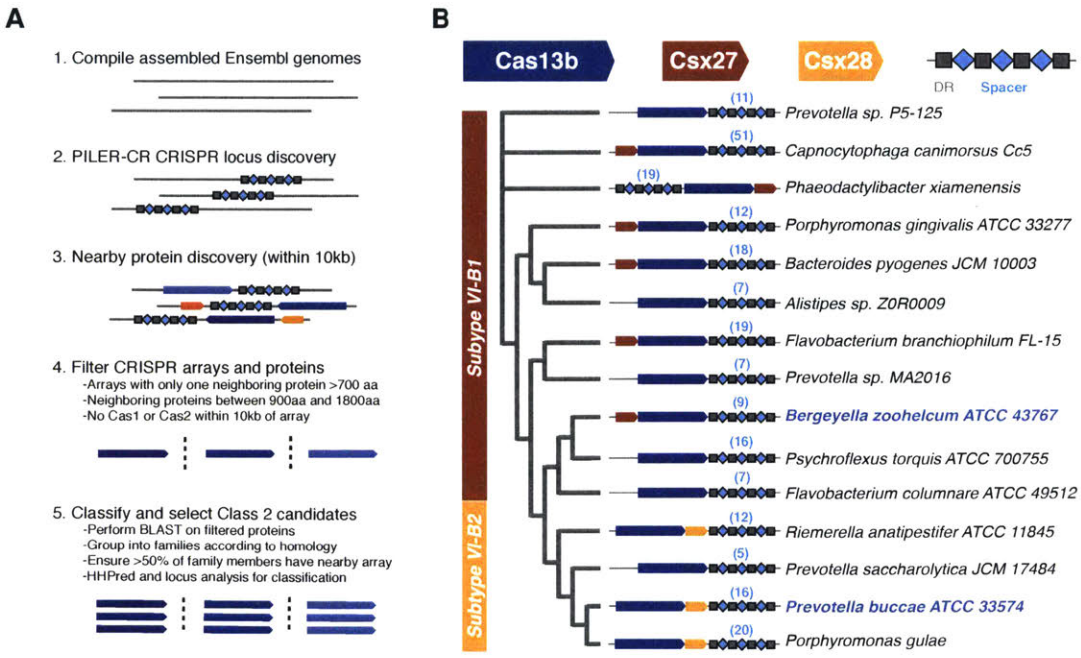

Figure 1 | Discovery of two Class 2 CRISPR-Cas systems, subtype VI-B1 and VI-B2, containing Cas13b. **(A)** Bioinformatic pipeline to discover putative Class 2 CRISPR loci lacking Cas1 and Cas2. **(B)** A schematic phylogenetic tree of the subtype VI-B loci. Loci with Csx27 (brown) comprise variant VI-B1; loci with Csx28 (gold) comprise variant VI-B2. See also Figures S1, S2, and S3.

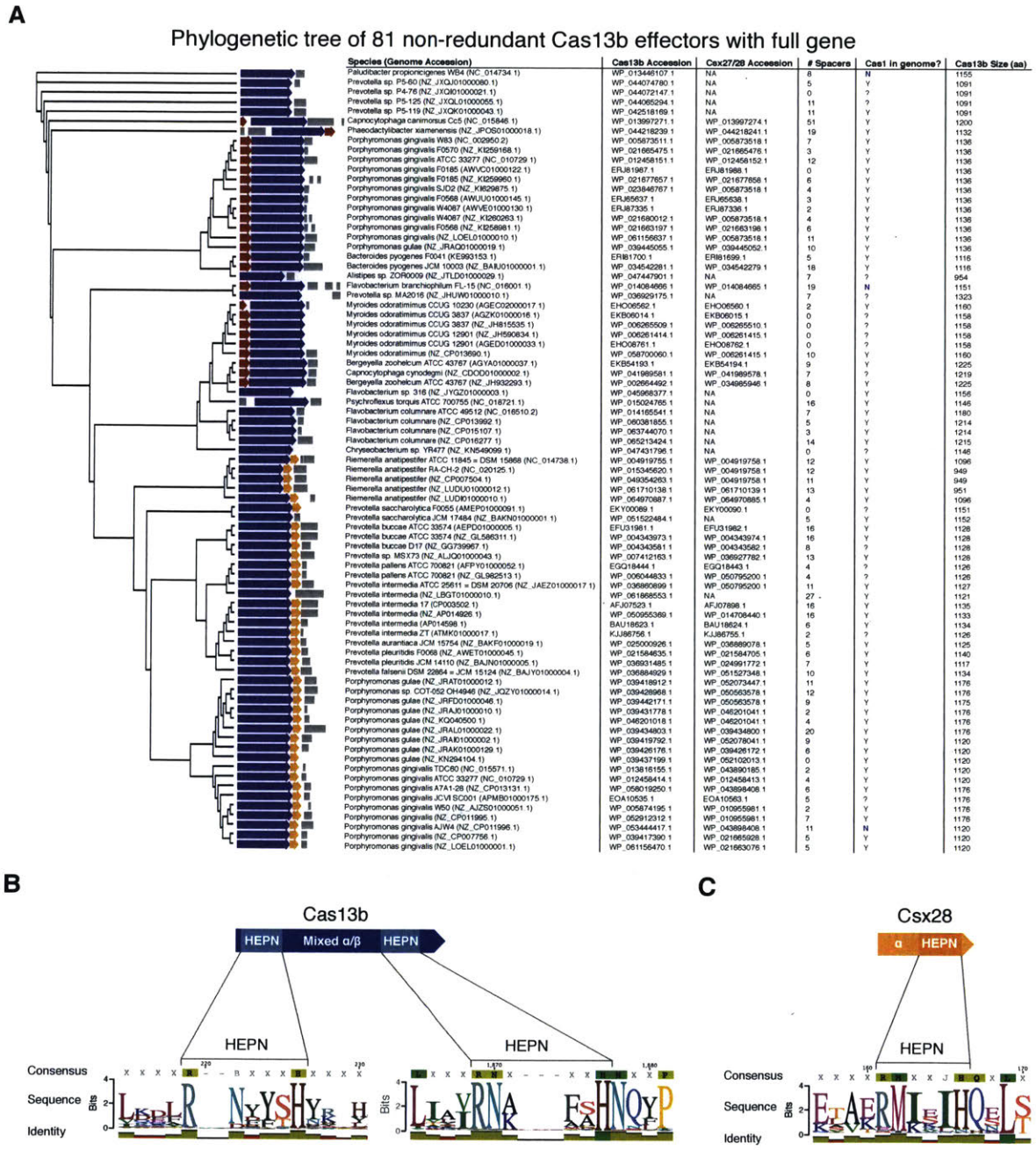

Figure S1 | Phylogenetic tree of Cas13b bifurcates into two variants of subtype VI-B CRISPR loci. Related to Figure 1. (A) A phylogenetic tree (alignment generated by BLOSUM62) of non-redundant Cas13b effectors, with the full type VI-B locus depicted in every instance. Accession numbers for genome, Cas13b (blue), and Csx27 (brown)/Csx28 (gold) are included, as well as number of nearby spacers

detected by PILER-CR, the presence of Cas1 in the sequenced genome, and the size of Cas13b. **(B)** Two HEPN sequences identified via multiple sequence alignment (BLOSUM62) of putative non-redundant Cas13b proteins. **(C)** Divergent HEPN sequence identified via multiple sequence alignment (BLOSUM62) of putative non-redundant Csx28 proteins.

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*, c*sx27* 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) (Moller et al., 2001). However, Csx27 of *Bergeyella zoohelcum* and Csx28 of *Prevotella buccae* tagged with RFP at either the N- or 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).

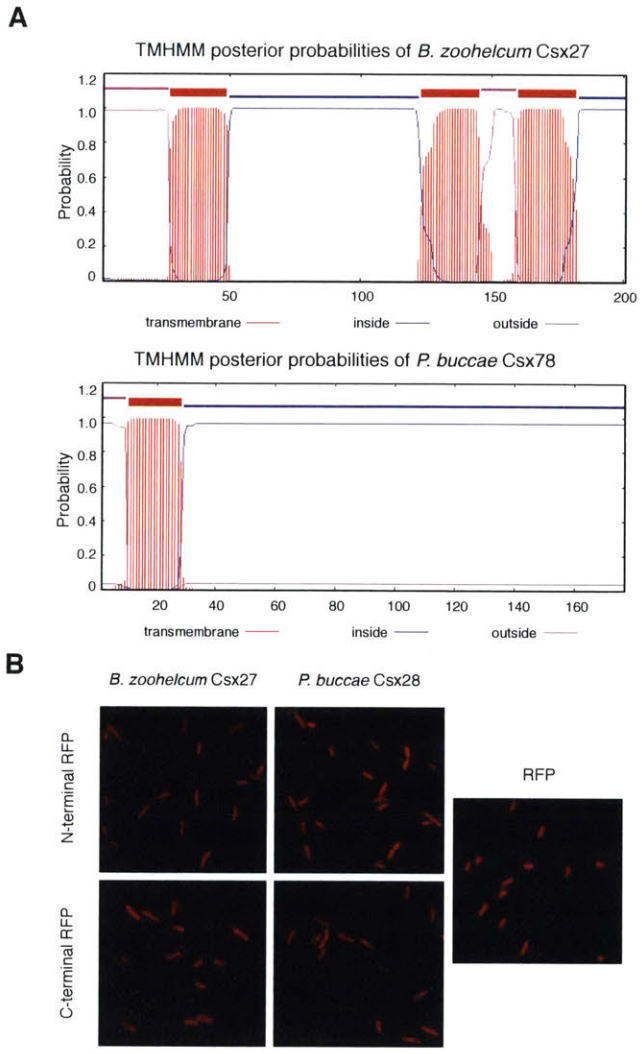

Figure S2 | Predicted transmembrane domains of Csx27 and Csx28 not validated experimentally. Related to Figure 1. (A) Transmembrane domain prediction in Csx27 of *B. zoohelcum* and Csx28 of *P. buccae* using TMHMM v2. **(B)** N- and C-terminally fused RFP imaging of Csx27 of *B. zoohelcum* and Csx28 of *P. buccae*.

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, S3B, and 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; Tables S1 and S2).

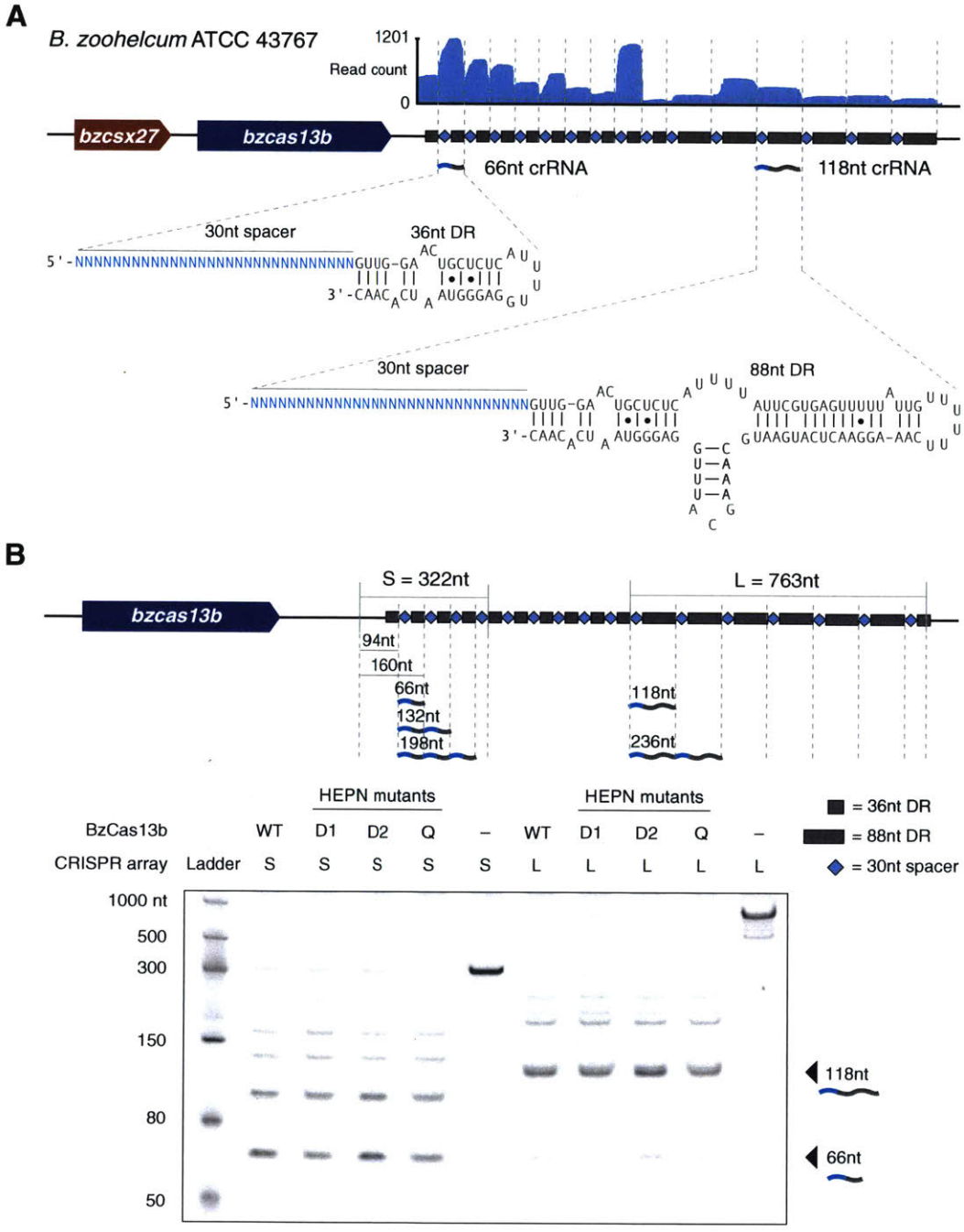

Figure 2 | Cas13b from the VI-B1 locus processes a CRISPR array with two direct repeat variants. **(A)** RNA-Sequencing of the native VI-B1 locus from *Bergeyella zoohelcum* ATCC 43767. **(B)** Denaturing gel showing cleavage products of in vitro

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

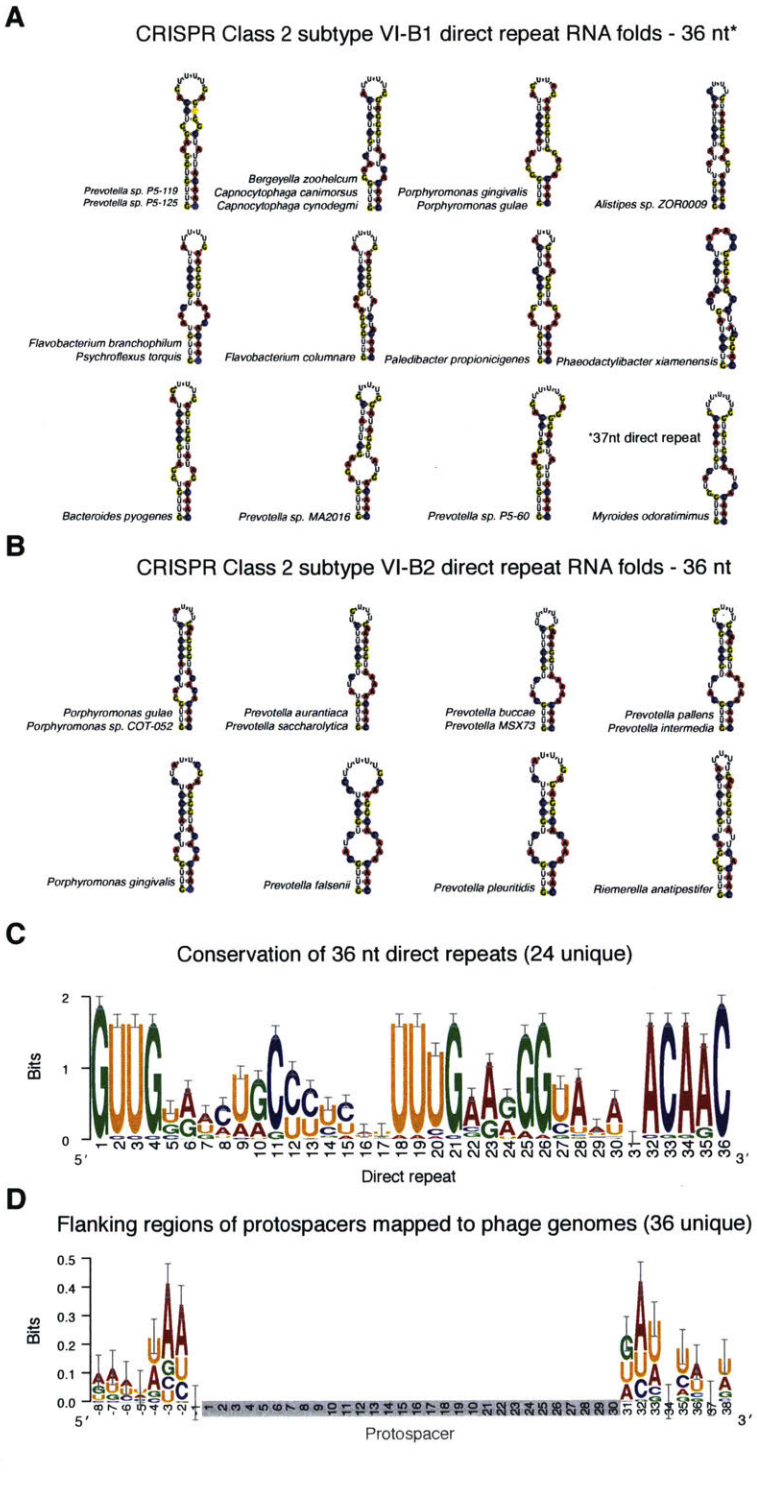
A
CRISPR Class 2 subtype VI-B1 direct repeat RNA folds - 36 nt*
Prevotella sp. P5-119
Prevotella sp. P5-125
Bergeyella zoohelcum
Capnocytophaga canimorsus
Capnocytophaga cynodegmi
Porphyromonas gingivalis
Porphyromonas gulae
Alistipes sp. ZOR0009
Flavobacterium branchophilum
Psychroflexus torquis
Flavobacterium columnare
Paludibacter propionicigenes
Phaeodactylibacter xiamenensis
Bacteroides pyogenes
Prevotella sp. MA2016
Prevotella sp. P5-60
*37nt direct repeat
Myroides odoratimimus
B
CRISPR Class 2 subtype VI-B2 direct repeat RNA folds - 36 nt
Porphyromonas gulae
Porphyromonas sp. COT-052
Prevotella aurantiaca
Prevotella saccharolytica
Prevotella buccae
Prevotella MSX73
Prevotella pallens
Prevotella intermedia
Porphyromonas gingivalis
Prevotella falsenii
Prevotella pleuritidis
Riemerella anatipestifer
C
Conservation of 36 nt direct repeats (24 unique)
Bits
Direct repeat
D
Flanking regions of protospacers mapped to phage genomes (36 unique)
Bits
Protospacer

Figure S3 | Predicted sequence and secondary structure of type VI-B direct repeats; predicted protospacer flanking sequences. Related to Figure 1. (A) Predicted secondary structure folds of structurally unique CRISPR class 2 type VI-B1 direct repeats (Vienna RNAfold). **(B)** Predicted secondary structure folds of structurally unique CRISPR Class 2 type VI-B2 direct repeats. **(C)** Weblogo of all unique VI-B direct repeat sequences of length 36 nt, taken as the same transcriptional orientation as Cas13b. **(D)** Weblogo of all unique VI-B protospacer flanking sequences from CRISPRTarget mapping of protospacers to phage databases.

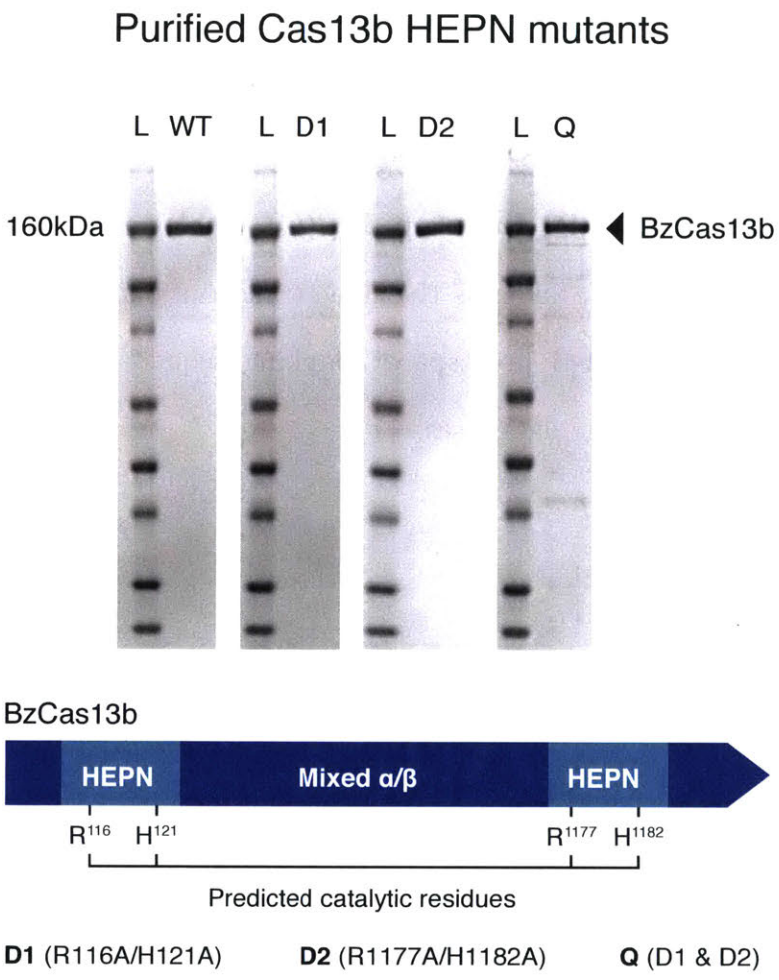

Figure S4 | Protein gels of purified WT BzCas13b and three mutant BzCas13b proteins. Related to Figures 2, 4, and 5. Denaturing protein gels of *B. zoohelcum* wildtype, D1 (R116A/H121A mutant), D2 (R1177A/H1182A) mutant, and Q (R116A/H121A/R1177A/H1182A) mutant Cas13b.

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 1100 randomly generated non-targeting spacers to establish baseline activity (Tables S3 and 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).

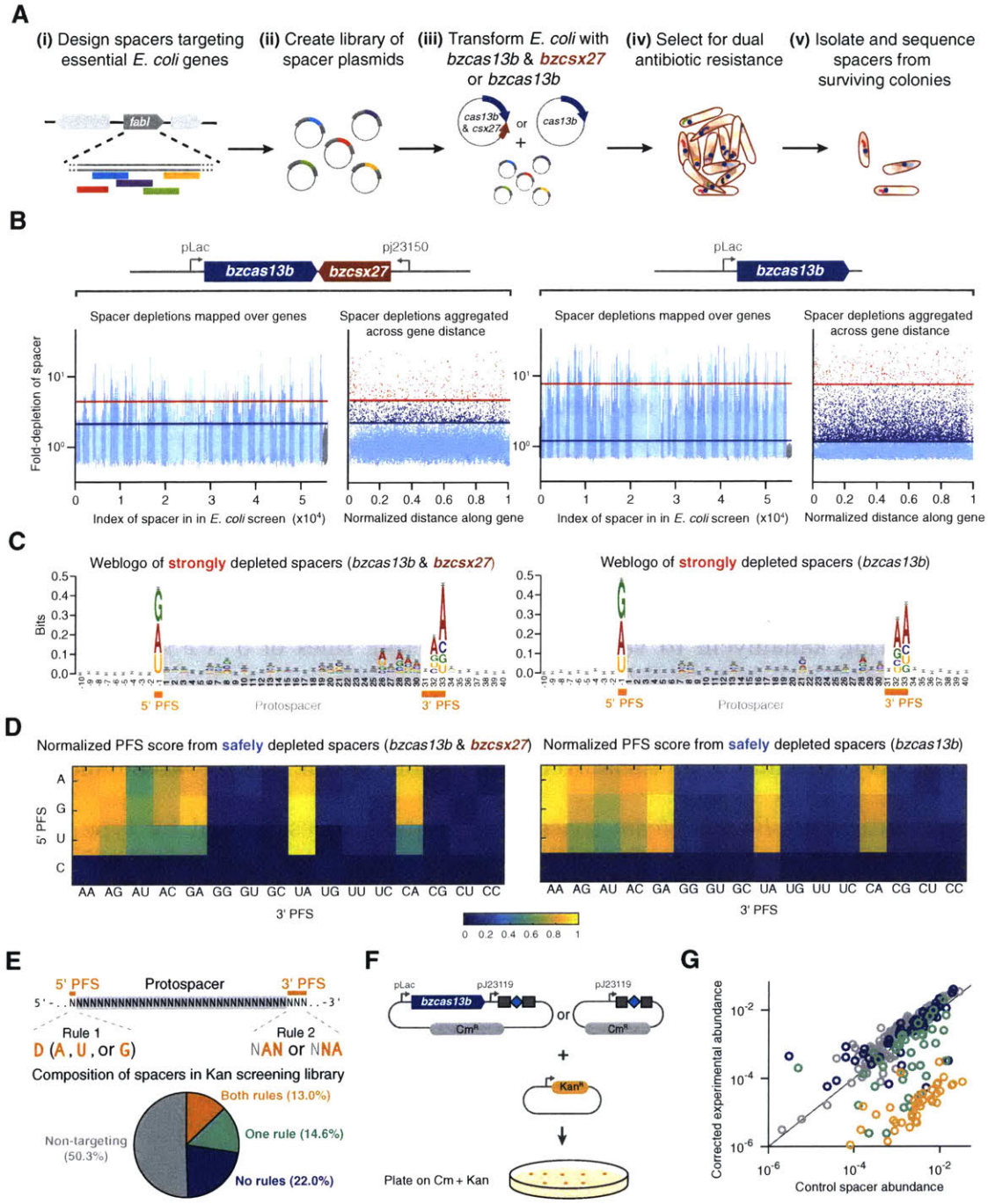

Figure 3 | Heterologous expression of Cas13b mediates knockdown of *E. coli* essential genes by a double-sided PFS. (A) Design of *E. coli* essential gene screen to determine targeting rules of nucleic acid interference. **(B)** Manhattan plots of mean

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 1672 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 6374 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. **(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 and Tables S2, S3, S4, and S5.

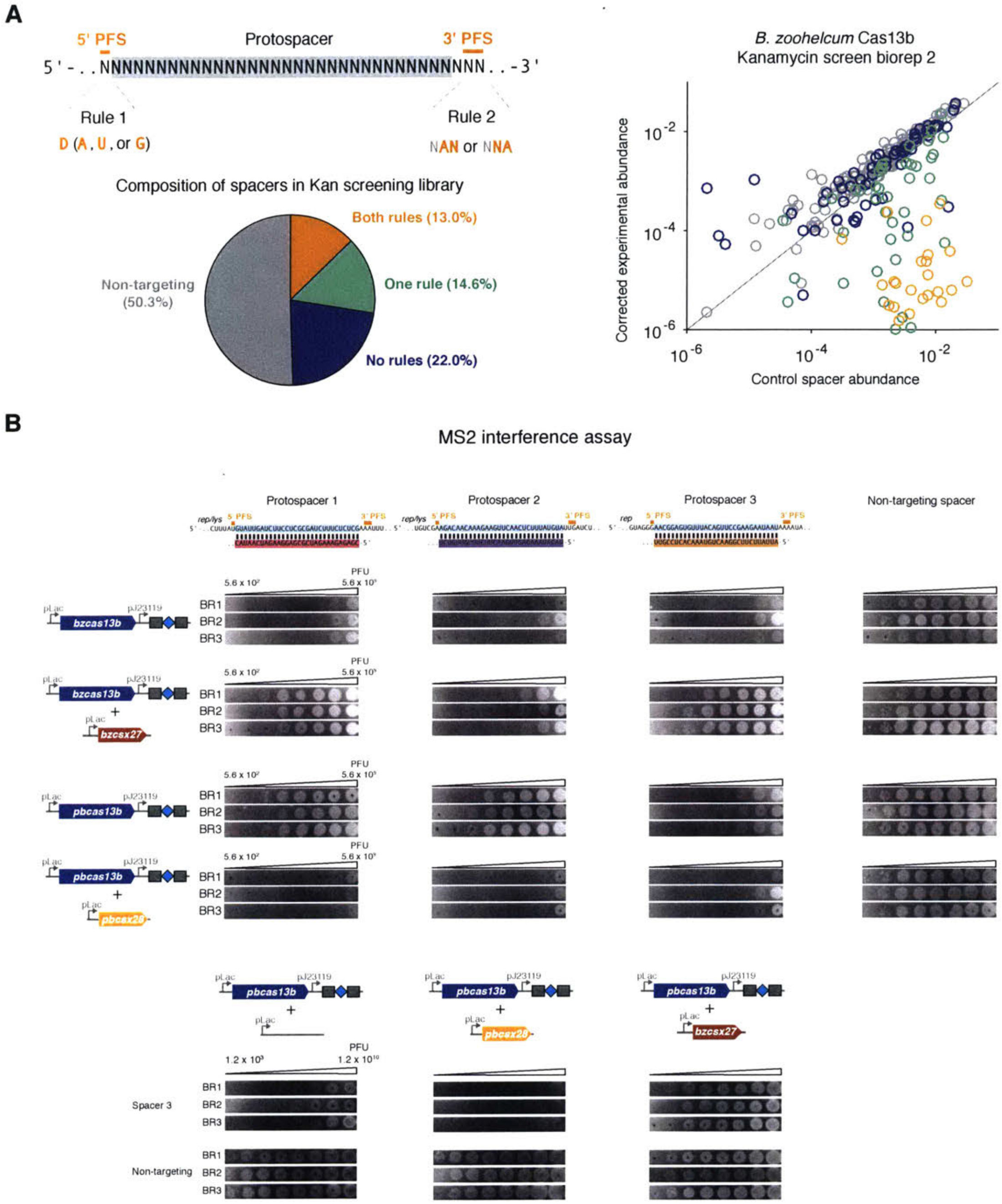
A
5' PFS
Protospacer
3' PFS
5'-..NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN..-3'
Rule 1
D (A, U, or G)
Rule 2
NAN or NNA
Composition of spacers in Kan screening library
Both rules (13.0%)
One rule (14.6%)
No rules (22.0%)
Non-targeting (50.3%)
B. zoohelcum Cas13b
Kanamycin screen biorep 2
Corrected experimental abundance
Control spacer abundance
B
MS2 interference assay
Protospacer 1
Protospacer 2
Protospacer 3
Non-targeting spacer
PFU
BR1
BR2
BR3
bzcas13b
bzcsx27
pbcas13b
pbcsx28
pLac
pJ23119
Spacer 3
Non-targeting

Figure S5 | Targeting rule validation of BzCas13b and MS2 interference assay of BzCas13b and PbCas13b. Related to Figures 3, 5, and 7. (A) Spacers targeting kanamycin to validate PFS targeting rules of 5' PFS (D) and 3' PFS (NAN or NNA) (left). Second kanamycin validation screen bioreplicate of spacer abundances versus control for individual *B. zoohelcum* spacers, with abundances colored by type of spacer (right). **(B)** Plaque drop assay with bioreplicates for *B. zoohelcum* VI-B1 locus and *cas13b*, for *P. buccae* VI-B2 locus and *cas13b*, and for *P. buccae cas13b* with pUC19, *B. zoohelcum csx27*, and *P. buccae csx28*.

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 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).

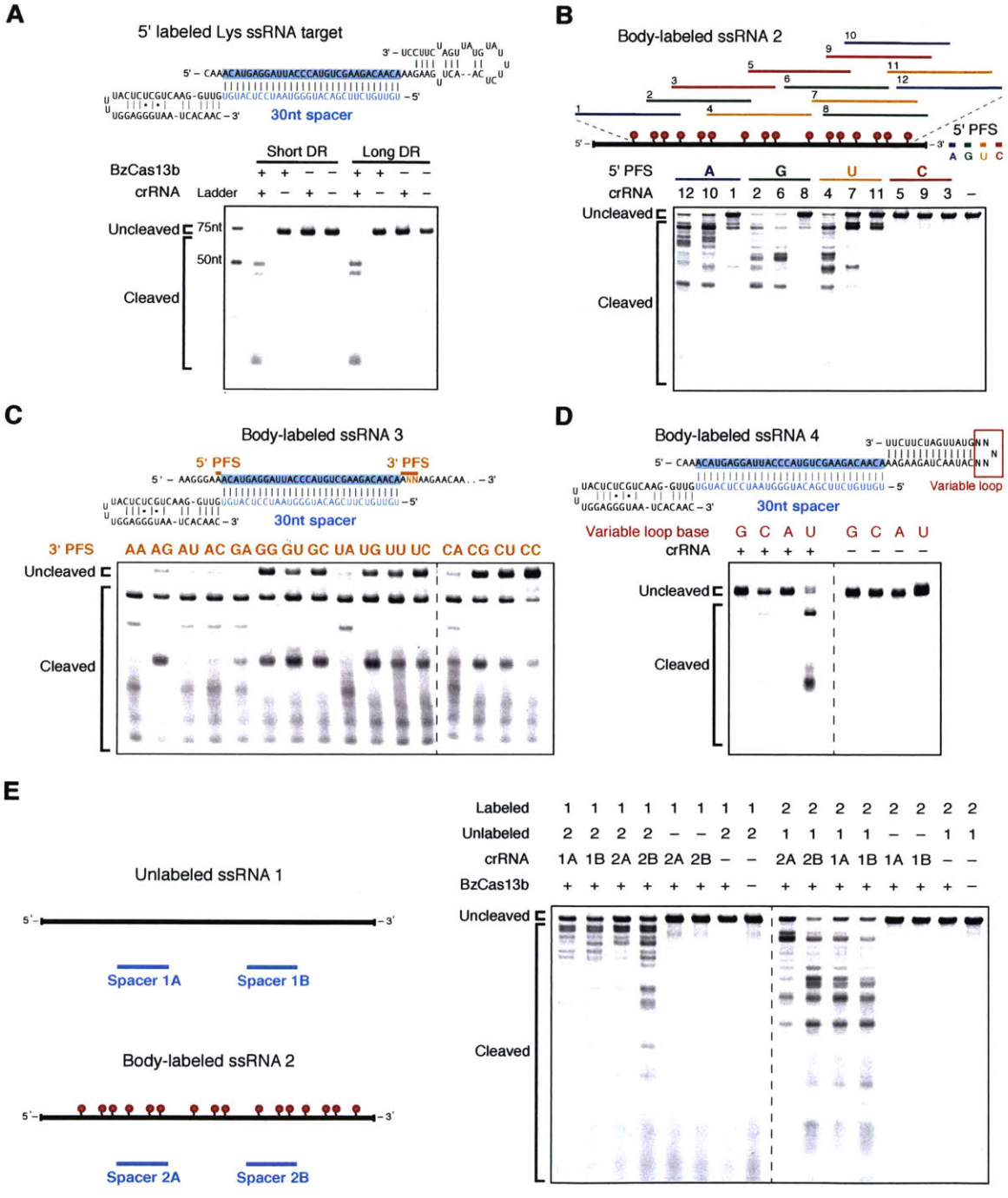

Figure 4 | Cas13b is a programmable single-stranded RNase with collateral activity. (A) Schematic showing the RNA secondary structure of the cleavage target in

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 minutes. 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 minutes. crRNAs correspond to protospacer numbered from the 5' to the 3' end of the target. Gel lane containing RNA ladder 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 minutes 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 not shown. Dashed line indicates two separate gels shown side by side. **(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 minutes. Dashed line indicates where the image was stitched together to remove U/A heteropolymer RNA lanes (shown in Figure S7B). Gel lane containing RNA ladder not shown. **(E)** Denaturing gel showing BzCas13b collateral cleavage activity after 30 minutes 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 not shown. Dashed line indicates two separate gels shown side by side. See also Figures S4 and S6, and Table S1.

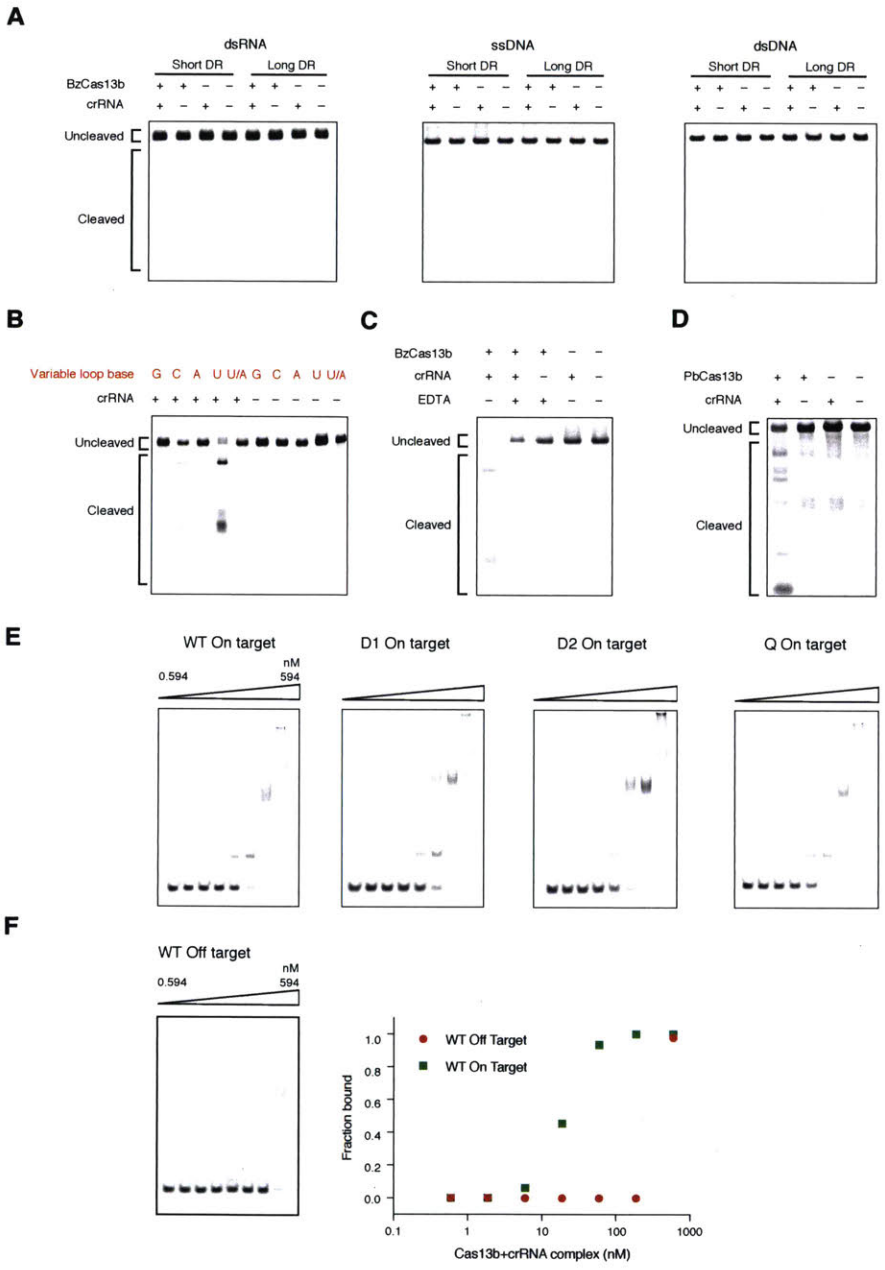

Figure S6 | Cas13b cleaves and binds to single-stranded RNA. Related to Figures 4 and 5. (A) Denaturing gels demonstrating no cleavage of dsRNA, ssDNA, or dsDNA by BzCas13b with either the short DR or long DR. Reactions were incubated for

10 minutes, the same amount of time which results in robust ssRNA cleavage for this target and crRNA pair. The ssDNA and top strand of the dsDNA target is 5′ labeled with IRDye 800. The dsRNA target is body labeled. Gel lane containing RNA ladder not shown. **(B)** Denaturing gel showing cleavage bands from the variable loop target as shown in Figure 4D. The U/A heteropolymer consists of the N^5 variable loop of alternating U and A residues (5' AUAUA 3'). **(C)** ssRNA cleavage requires BzCas13b and a targeting crRNA, and this cleavage activity is abolished by addition of EDTA. Gel lane containing RNA ladder not shown. **(D)** Denaturing gel showing PbCas13b cleavage activity of an ssRNA targeted substrate. The ssRNA is 5′ labeled with IRDye 800 and incubated for 30 minutes. Gel lane containing RNA ladder not shown. **(E)** EMSA gels that were used to quantify the K_D of the WT and mutant BzCas13b proteins, using an on-target crRNA complementary to the targeted ssRNA. **(F)** EMSA gel of WT BzCas13b with an off-target crRNA. The off-target crRNA is non-complementary to the targeted ssRNA.

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 *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 modified an existing plasmid interference assay 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 comparable rate to 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 ~2 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 obtained 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 wildtype and all three HEPN mutant Cas13b proteins showed comparable binding affinities for a single-stranded target RNA substrate, with K_D values ranging from 27nM to 42nM (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 wildtype 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.

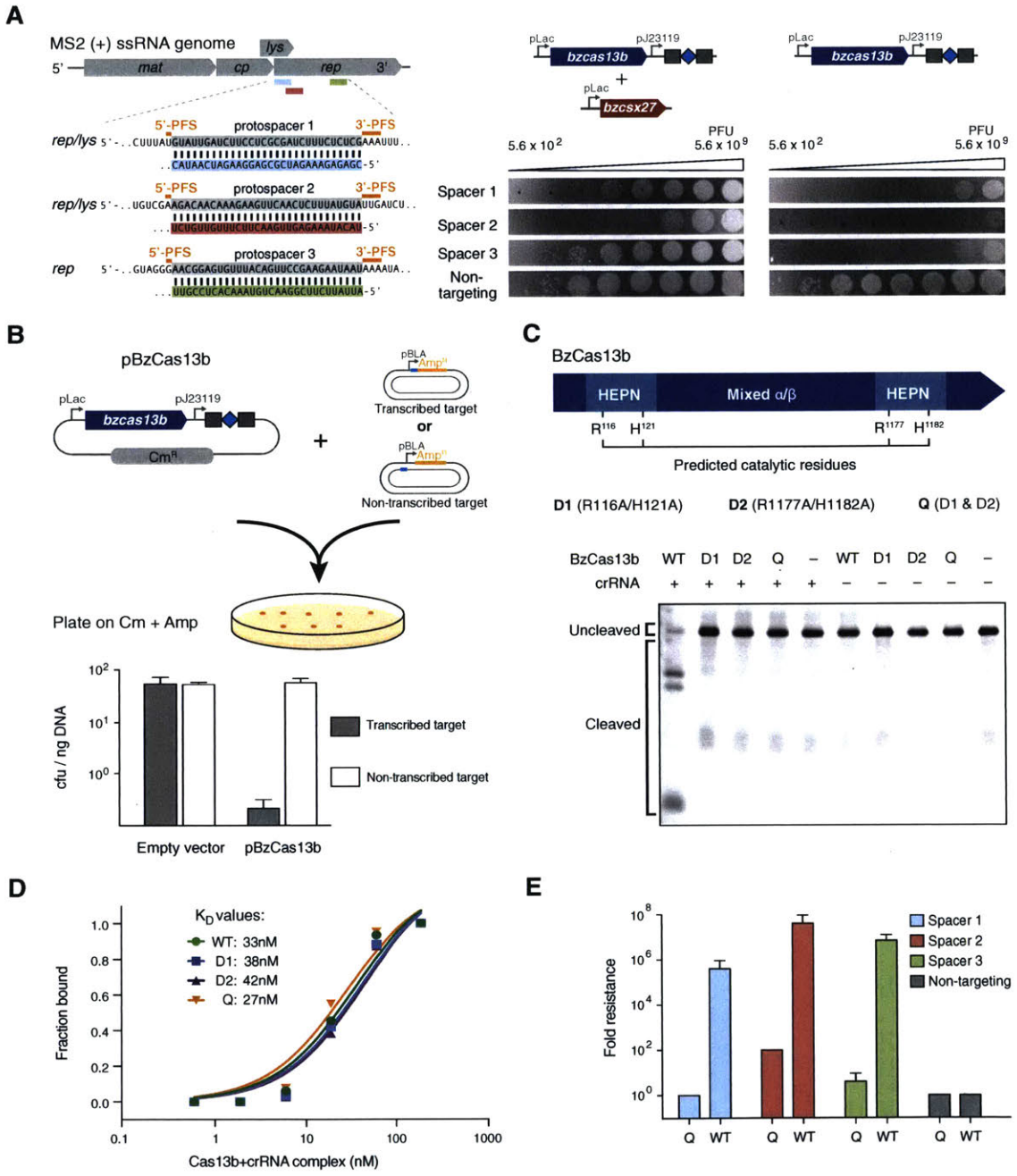

Figure 5 | HEPN domains mediate RNA cleavage by Cas13b, whose activity is repressed by Csx27. (A) Protospacer design for MS2 phage plaque drop assay to test RNA interference (left). Plaque drop assay for full *B. zoohelcum* VI-B1 locus (center) and *bzcas13b* (right). **(B)** DNA interference assay schematic (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 minutes. Gel lane containing RNA ladder 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. The EMSA assay was performed with supplemental EDTA to reduce any cleavage activity. **(E)** Quantification of MS2 phage plaque drop assay with *B. zoohelcum* wildtype and Q (R116A/H121A/R1177A/H1182A) mutant Cas13b. See also Figures S4, S5, and S6, and Tables S1, S2, S6, and S7.

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 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 only a marginal underrepresentation of spacers in the 3' UTR compared to a uniform distribution, suggesting that collateral activity may occur in vivo.

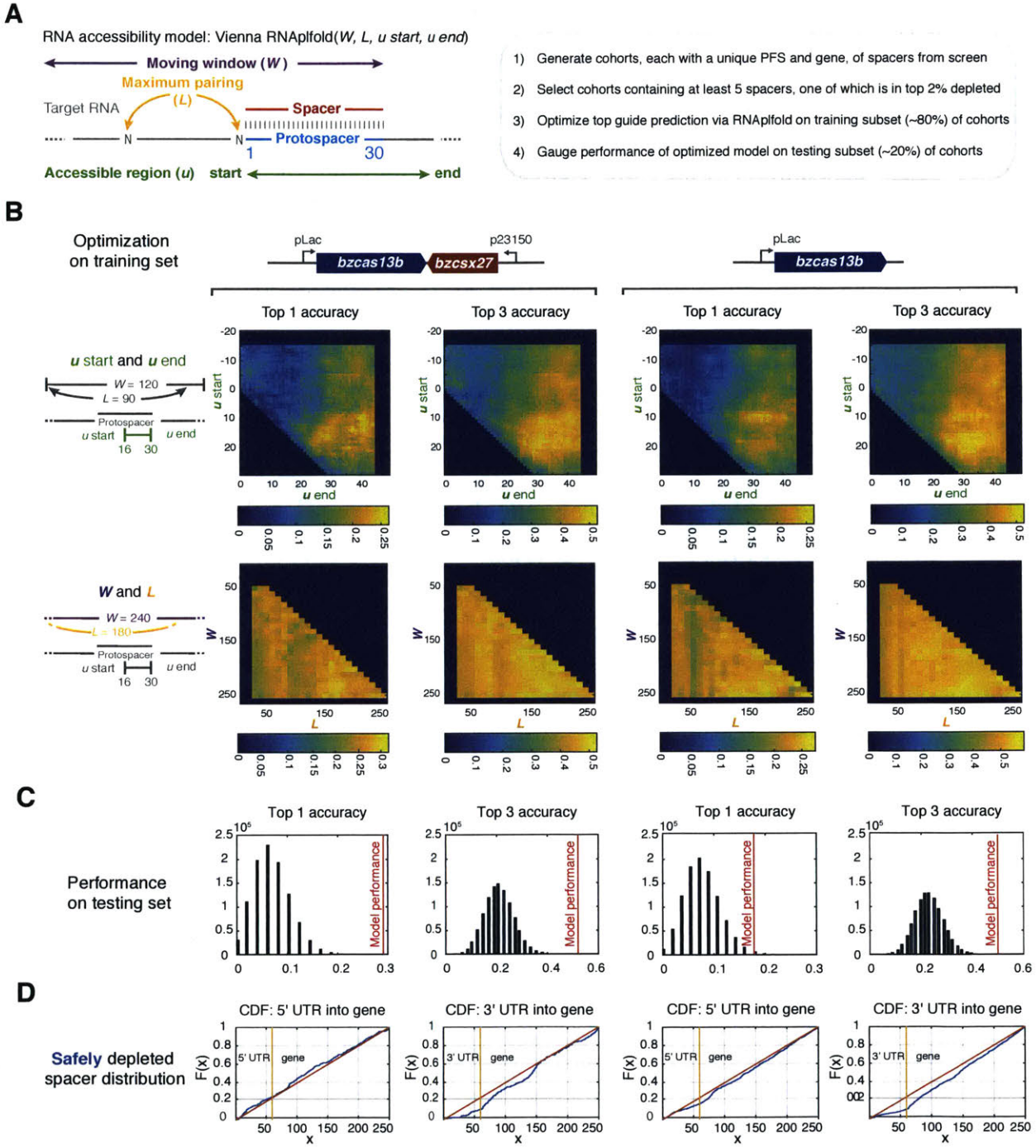

Figure 6 | Efficient RNA targeting by Cas13b is correlated with local RNA accessibility. **(A)** Methodology of secondary structure-mediated spacer efficiency analysis of *E. coli* essential gene screen data with Vienna RNAplfold. **(B)** Optimization of

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 3 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, 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 Tables S3 and S4.

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 PFS's 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 plaque drop 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.

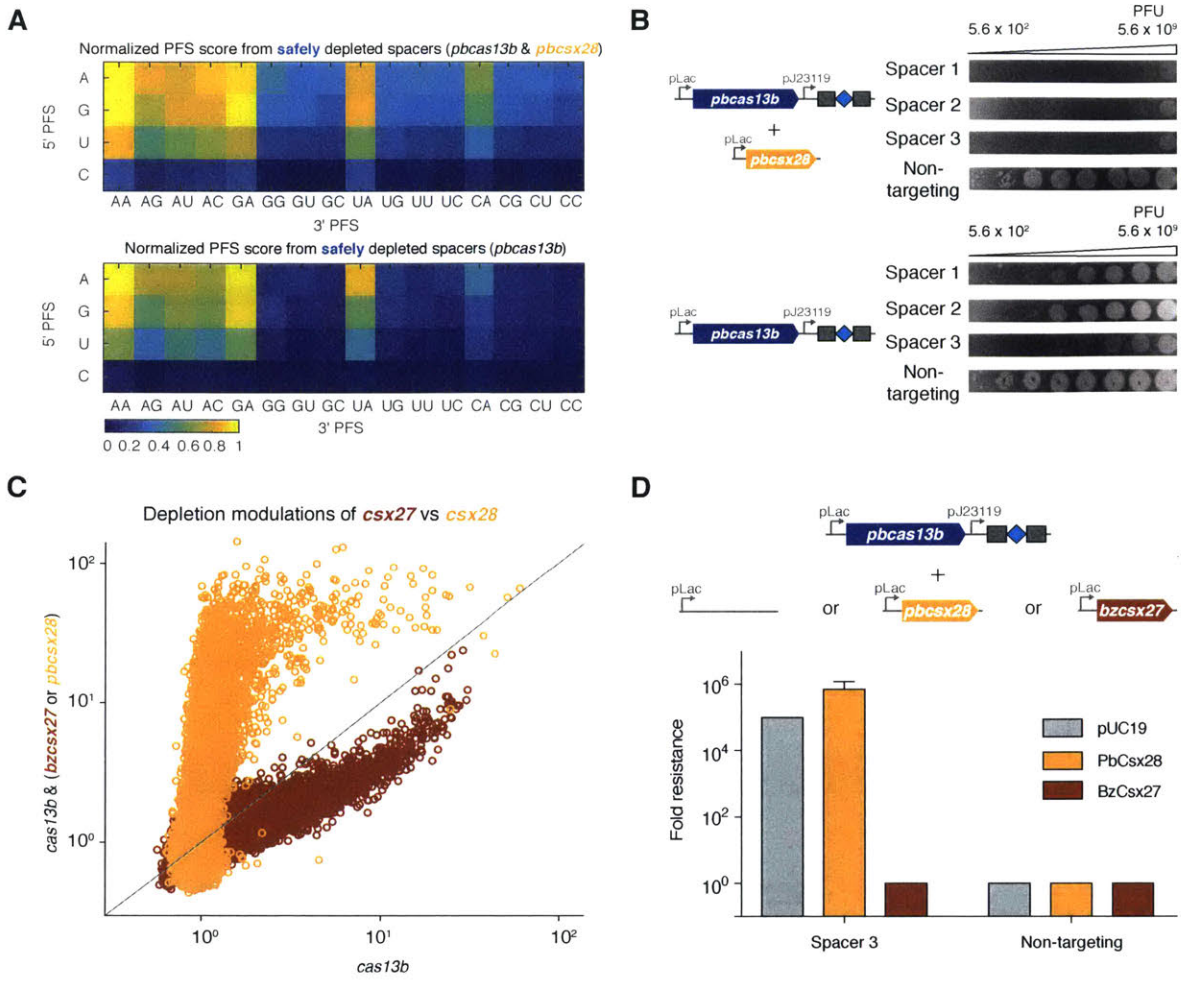

Figure 7 | Class 2 type VI-B systems are differentially regulated across two loci by Csx27 and Csx28. (A) Normalized PFS matrix, for *P. buccae* VI-B2 locus (left) and *pbcas13b* (right). **(B)** MS2 Plaque drop assay for full *P. buccae* VI-B2 locus (left) and *pbcas13b* (right). **(C)** Spacer depletions 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, and Tables S2, S3, S4, and S6.

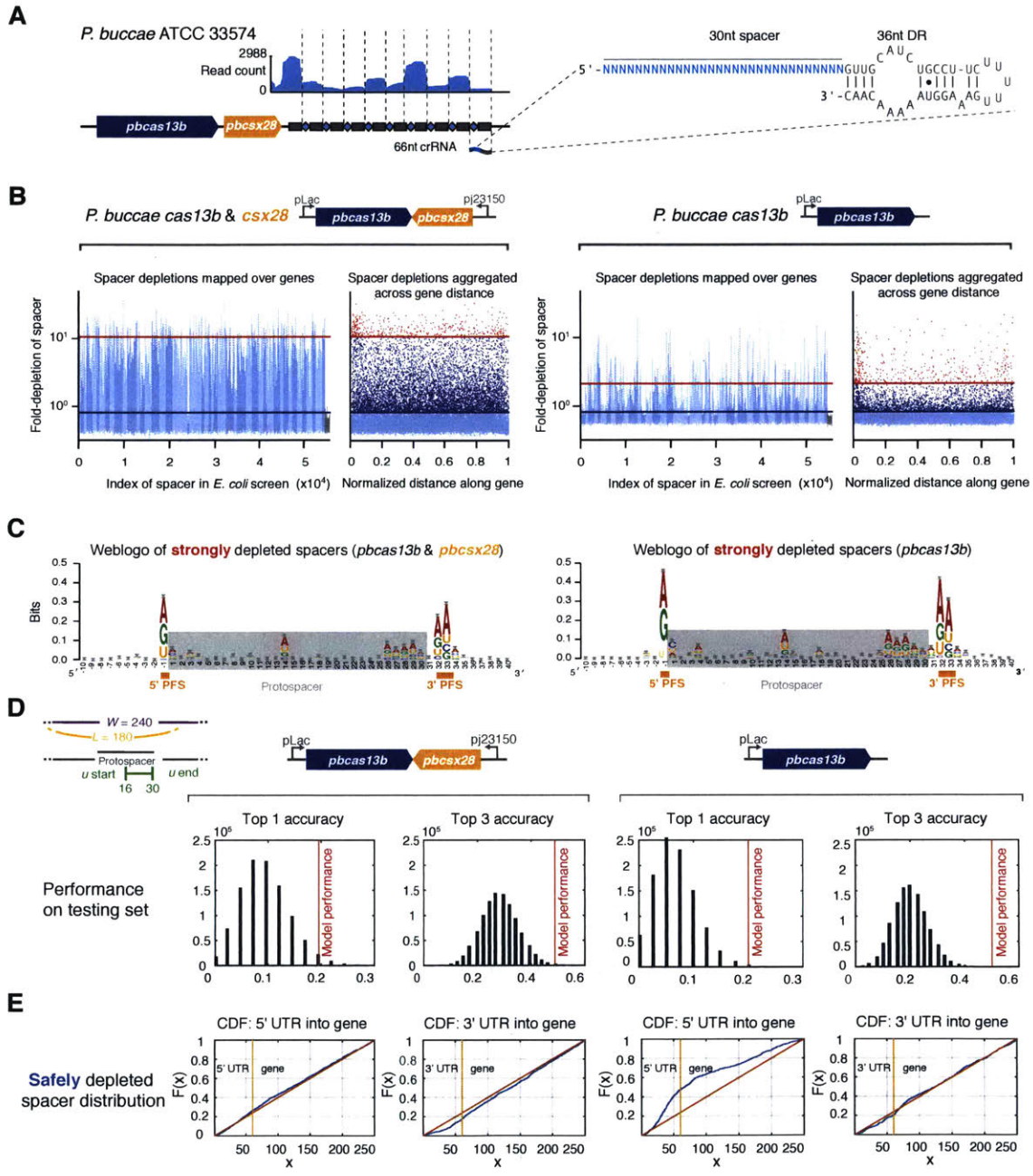
A
P. buccae ATCC 33574
2988
Read count
0
pbcas13b
pbcsx28
66nt crRNA
30nt spacer
36nt DR
5'-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGUUG
3'-CAAC
B
P. buccae cas13b & csx28
pLac
pbcas13b
pbcsx28
pj23150
P. buccae cas13b
Spacer depletions mapped over genes
Spacer depletions aggregated across gene distance
Fold-depletion of spacer
Index of spacer in E. coli screen (x10^4)
Normalized distance along gene
C
Weblogo of strongly depleted spacers (pbcas13b & pbcsx28)
Weblogo of strongly depleted spacers (pbcas13b)
Bits
5' PFS
Protospacer
3' PFS
D
W = 240
L = 180
Protospacer
u start
u end
16
30
Performance on testing set
Top 1 accuracy
Top 3 accuracy
Model performance
E
Safely depleted spacer distribution
CDF: 5' UTR into gene
CDF: 3' UTR into gene
5' UTR
3' UTR
gene
F(x)
x

Figure S7 | RNA-targeting of *P. buccae* VI-B2 CRISPR locus. Related to Figure 7. **(A)** RNA-Sequencing of heterologously expressed VI-B2 locus from *P. buccae* ATCC 33574 in *E. coli.* **(B)** Manhattan plots of spacer depletions mapped over 45 genes and aggregated across normalized gene distance for full *P. buccae* VI-B2 locus (left) and *cas13b* (right), with non-targeting spacers in gray, safely depleted (>5σ above mean depletion of non-targeting spacers) spacers above blue line, and strongly depleted (top 1% depleted) spacers above red line. For the full locus, 36,141 targeting spacers and 859 non-targeting spacers passed QC filter. Of the targeting, 370 are strongly depleted and 8065 are safely depleted. For *cas13b* alone, 41,126 targeting spacers and 824 non-targeting spacers passed QC filter. Of the targeting, 419 are strongly depleted and 3295 are safely depleted. **(C)** Sequence weblogos of strongly depleted *P. buccae* spacers, revealing double-sided PFS (protospacer flanking sequence). **(D)** Performance of optimized RNAplfold model (W=240, L=180, *u start*=16, *u end*=30) on randomly selected *P. buccae* testing dataset (41 cohorts for full *P. buccae* VI-B2 locus, 40 cohorts for *pbcas13b*) against 10^6 Monte Carlo simulations: empirical P-values from left to right of 3.3e-2, 2.7e-3, 3.9e-3, 1.5e-5. **(E)** Empirical cumulative distribution function of safely depleted *P. buccae* 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 *P. buccae* spacers.

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 exhibit 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, at least in the CRISPR-Cas context, 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; 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 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.

AUTHOR CONTRIBUTIONS

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

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., *et al.* (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.

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., Balazsi, 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 in molecular biology *1311*, 1-21.

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

Hildebrand, A., Remmert, M., Biegert, A., and Soding, 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., Honer 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. Nature structural & molecular biology *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.

Moller, 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 Soding, J. (2012). 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. Molecular 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. Nature reviews Microbiology *15*, 169-182.

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, C., 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., Nunez, 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*, D710-716.

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.

Supplementary Tables

Excel files for supplementary tables can be downloaded from the following link: http://www.cell.com/molecular-cell/fulltext/S1097-2765(16)30866-8

Table S1 | All crRNAs, nucleic acid targets, and primers used in biochemical experiments. Related to Figures 2, 4 and 5.

Table S2 | All Cas13b plasmids used in this study. Related to Figures 3, 5, and 7.

Table S3 | *E. coli* essential genes represented in *E. coli* essential gene screen library of spacers. Related to Figures 3, 6, and 7.

Table S4 | Spacers from *E. coli* essential gene screen. Related to Figures 3, 6, and 7.

Table S5 | Spacers from kanamycin validation screen. Related to Figure 3.

Table S6 | Spacers targeting MS2 and pBLA plasmids. Related to Figures 5 and 7.

Table S7 | EMSA raw data. Related to Figure 5.

Materials and Methods

***P. buccae* (Holdeman et al.) Shah and Collins:** *P. buccae* was not grown in this study.

***B. zoohelcum* (Holmes et al.) Vandamme et al.:** *B. zoohelcum* ATCC 43767 was grown in ATCC medium 44 (Brain Heart Infusion broth) at 37°C at 250 rpm overnight.

***E. coli* (Migula) Castellani and Chalmers (C3000):** *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 Electrocomp™ 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 liters 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 hours 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 hours.

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., 2012) 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 (Moller 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 minute 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 hours) 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™-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 minute 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 liters 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 hours 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 10,000g for 1 hour. Filtered lysate was incubated with StrepTactin Sepharose High Performance (GE Healthcare Life Sciences) at 4°C for 1 hour 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 hours. 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 minutes at 25°C. The reactions were mixed with equal parts of RNA loading dye (New England Biolabs) and denatured at 95°C for 5 minutes and then cooled on ice for 2 minutes. 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 .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 preincubated at 37°C for 15 minutes, after which the 5'-labeled target was added. Reactions were then incubated at 37°C for 10 minutes 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 5000g and resuspended in PBS with 1% methanol-free formaldehyde. After 30 minutes 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-minute 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 (Epicentre Bio) to convert 5'-triphosphates to 5'-monophosphates for adapter ligation. Samples were then polyA tailed with *E. coli* Poly(A) polymerase (New

England Biolabs), and a 5' RNA Illumina sequencing adapter 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' adapter 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 data set, 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 data set 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* (Tables S2, 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 (Tables S2, 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 minutes on ice, heatshocked for 30 seconds at 42°C and incubated on ice for 2 minutes. Then, 95 uL of SOC was added to cells and they were incubated with shaking at 37°C for 90 minutes, 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 $1/3N$ 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 data set 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

Chapter 3

Adapted from *Science*:

RNA Editing with CRISPR-Cas13

David B.T. Cox[1,2,3,4,5,6*], Jonathan S. Gootenberg[1,2,3,4,7*], Omar O. Abudayyeh[1,2,3,4,6*], Brian Franklin[1,2,3,4], Max J. Kellner[1,2,3,4], Julia Joung[1,2,3,4], Feng Zhang[1,2,3,4†]

[1] Broad Institute of MIT and Harvard

Cambridge, MA 02142, USA

[2] McGovern Institute for Brain Research

[3] Department of Brain and Cognitive Science

[4] Department of Biological Engineering

[5] Department of Biology

[6] Harvard-MIT Division of Health Sciences and Technology

Massachusetts Institute of Technology

Cambridge, MA 02139, USA

[7] Department of Systems Biology

Harvard Medical School

Boston, MA 02115, USA

* These authors contributed equally to this work.

† Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

Abstract

Nucleic acid editing holds promise for treating genetic disease, particularly at the RNA level, where disease-relevant sequences can be rescued to yield functional protein products. Type VI CRISPR-Cas systems contain the programmable single-effector RNA-guided RNases Cas13. Here, we profile Type VI systems to engineer a Cas13 ortholog capable of robust knockdown and demonstrate RNA editing by using catalytically inactive Cas13 (dCas13) to direct adenosine-to-inosine deaminase activity by ADAR2 to transcripts in mammalian cells. This system, referred to as RNA Editing for Programmable A-to-I Replacement (REPAIR), which has no strict sequence constraints, can be used to edit full-length transcripts containing pathogenic mutations. We further engineer this system to create a high-specificity variant and minimize the system to facilitate viral delivery. REPAIR presents a promising RNA-editing platform with broad applicability for research, therapeutics, and biotechnology.

* * * * *

Introduction

Precise nucleic acid editing technologies are valuable for studying cellular function and as novel therapeutics. Current editing tools, based on programmable nucleases such as the prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR)-associated nucleases Cas9 (*1-4*) or Cpf1 (*5*), have been widely adopted for mediating targeted DNA cleavage which in turn drives targeted gene disruption through non-homologous end joining (NHEJ) or precise gene editing through template-dependent homology-directed repair (HDR) (*6*). NHEJ utilizes host machineries that are active in both dividing and post-mitotic cells and provides efficient gene disruption by generating a

mixture of insertion or deletion (indel) mutations that can lead to frame shifts in protein coding genes. HDR, in contrast, is mediated by host machineries whose expression is largely limited to replicating cells. Accordingly, the development of gene-editing capabilities for post-mitotic cells remains a major challenge. DNA base editors, consisting of a fusion between Cas9 nickase and cytidine deaminase can mediate efficient cytidine-to-uridine conversions within a target window and significantly reduce the formation of double-strand break induced indels (*7, 8*). However the potential targeting sites of DNA base editors are limited by the requirement of Cas9 for a protospacer adjacent motif (PAM) at the editing site (*9*). Here, we describe the development of a precise and flexible RNA base editing technology using the type VI CRISPR-associated RNA-guided RNase Cas13 (*10-13*).

Cas13 enzymes have two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) endoRNase domains that mediate precise RNA cleavage with a preference for targets with protospacer flanking site (PFS) motif observed biochemically and in bacteria (*10, 11*). Three Cas13 protein families have been identified to date: Cas13a (previously known as C2c2), Cas13b, and Cas13c (*12, 13*). We recently reported that Cas13a enzymes can be adapted as tools for nucleic acid detection (*14*) as well as mammalian and plant cell RNA knockdown and transcript tracking (*15*). Interestingly, the biochemcial PFS was not required for RNA interference with Cas13a (*15*). The programmable nature of Cas13 enzymes makes them an attractive starting point to develop tools for RNA binding and perturbation applications.

The adenosine deaminase acting on RNA (ADAR) family of enzymes mediates endogenous editing of transcripts via hydrolytic deamination of adenosine to inosine, a nucleobase that is functionally equivalent to guanosine in translation and splicing (*16, 17*). There are two functional human ADAR orthologs, ADAR1 and ADAR2, which

consist of N-terminal double-stranded-RNA-binding domains and a C-terminal catalytic deamination domain. Endogenous target sites of ADAR1 and ADAR2 contain substantial double-stranded identity, and the catalytic domains require duplexed regions for efficient editing *in vitro* and *in vivo* (*18, 19*). Importantly, the ADAR catalytic domain is capable of deaminating target adenosines without any protein co-factors *in vitro* (*20*). ADAR1 has been found to target mainly repetitive regions whereas ADAR2 mainly targets non-repetitive coding regions (*17*). Although ADAR proteins have preferred motifs for editing that could restrict the potential flexibility of targeting, hyperactive mutants, such as ADAR2(E488Q) (*21*), relax sequence constraints and increase adenosine-to-inosine editing rates. ADARs preferentially deaminate adenosines mispaired with cytidine bases in RNA duplexes (*22*), providing a promising opportunity for precise base editing. Although previous approaches have engineered targeted ADAR fusions via RNA guides (*23-26*), the specificity of these approaches has not been reported and their respective targeting mechanisms rely on RNA-RNA hybridization without the assistance of protein partners that may enhance target recognition and stringency.

Here we assay a subset of the family of Cas13 enzymes for RNA knockdown activity in mammalian cells and identify the Cas13b ortholog from *Prevotella sp. P5-125* (PspCas13b) as the most efficient and specific for mammalian cell applications. We then fuse the ADAR2 deaminase domain ($ADAR2_{DD}$) to catalytically inactive PspCas13b and demonstrate RNA editing for programmable A-to-I (G) replacement (REPAIR) of reporter and endogenous transcripts as well as disease-relevant mutations. Lastly, we employ a rational mutagenesis scheme to improve the specificity of dCas13b-$ADAR2_{DD}$ fusions to generate REPAIRv2 with more than 919-fold higher specificity.

Comprehensive Characterization of Cas13 Family Members in Mammalian Cells

We previously developed LwaCas13a for mammalian knockdown applications, but it required an monomeric superfolder GFP (msfGFP) stabilization domain for efficient knockdown and, although the specificity was high, knockdown levels were not consistently below 50% (*15*). We sought to identify a more robust RNA-targeting CRISPR system by characterizing a genetically diverse set of Cas13 family members to assess their RNA knockdown activity in mammalian cells (Fig. 1A). We generated mammalian codon-optimized versions of multiple Cas13 proteins, including 21 orthologs of Cas13a, 15 of Cas13b and 7 of Cas13c, and cloned them into an expression vector with N- and C-terminal nuclear export signal (NES) sequences and a C-terminal msfGFP to enhance protein stability (Supplementary Table 1). To assay interference in mammalian cells, we designed a dual reporter construct expressing the independent *Gaussia* (*Gluc*) and *Cypridinia* (*Cluc*) luciferases under separate promoters, which allows one luciferase to function as a measure of Cas13 interference activity and the other to serve as an internal control. For each Cas13 ortholog, we designed protospacer flanking site (PFS)-compatible guide RNAs, using the Cas13b PFS motifs derived from an ampicillin-interference assay (fig S1; Supplementary Table 2; Supplementary Information) and the 3' H (not G) PFS from previous reports of Cas13a activity (*10*).

We transfected HEK293FT cells with Cas13-expression, guide RNA, and reporter plasmids and then quantified levels of Cas13 expression and the targeted *Gluc* 48 hours later (Fig. 1B, fig. S2A). Testing two guide RNAs for each Cas13 ortholog revealed a range of activity levels, including five Cas13b orthologs with similar or increased interference across both guide RNAs relative to the recently characterized LwaCas13a (Figure 1B), and we observed only a weak correlation between Cas13 expression and

interference activity (fig. S2B-D). We selected the top five Cas13b orthologs, as well as the top two Cas13a orthologs for further engineering.

We next tested Cas13-mediated knockdown of Gluc without msfGFP, to select orthologs that do not require stabilization domains for robust activity. We hypothesized that Cas13 activity could be affected by subcellular localization, as we previously reported for optimization of LwaCas13a (*15*). Therefore, we tested the interference activity of the seven selected Cas13 orthologs C-terminally fused to one of six different localization tags without msfGFP. Using the luciferase reporter assay, we identified the top three Cas13b designs with the highest level of interference activity: Cas13b from *Prevotella sp. P5-125* (PspCas13b) and Cas13b from *Porphyromonas gulae* (PguCas13b) C-terminally fused to the HIV Rev gene NES and Cas13b from *Riemerella anatipestifer* (RanCas13b) C-terminally fused to the MAPK NES (fig. S3A). To further distinguish activity levels of the top orthologs, we compared the three optimized Cas13b constructs to the optimal LwaCas13a-msfGFP fusion and to shRNA for their ability to knockdown the endogenous *KRAS* transcript using position-matched guides (fig. S3B). We observed the highest levels interference for PspCas13b (average knockdown 62.9%) and thus selected this for further comparison to LwaCas13a.

To more rigorously define the activity of PspCas13b and LwaCas13a, we designed position-matched guides tiling along both *Gluc* and *Cluc* transcripts and assayed their activity using our luciferase reporter assay. We tested 93 and 20 position-matched guides targeting *Gluc* and *Cluc*, respectively, and found that PspCas13b had consistently increased levels of knockdown relative to LwaCas13a (average of 92.3% for PspCas13b vs. 40.1% knockdown for LwaCas13a) (Fig. 1C,D).

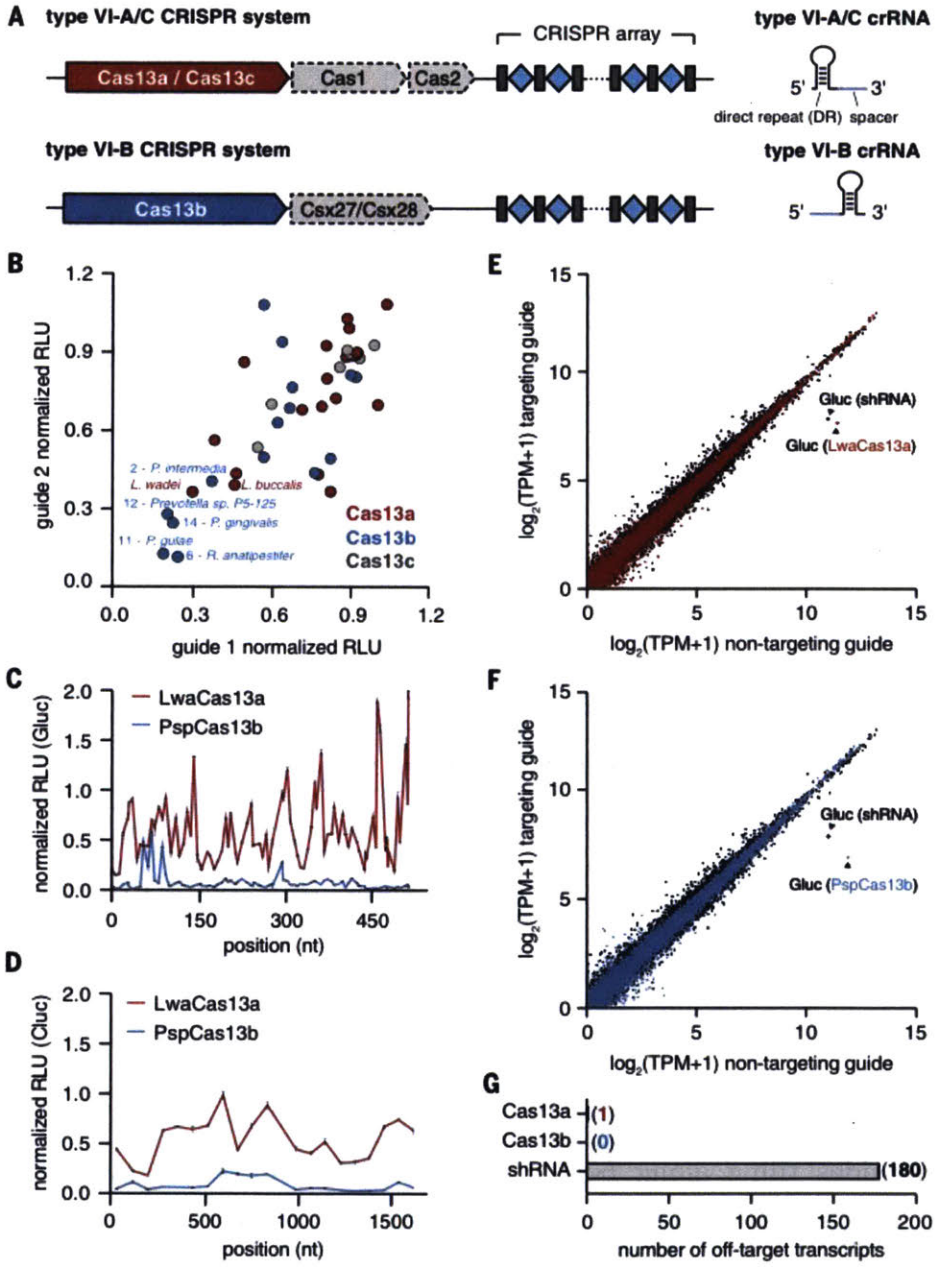

Figure 1: Characterization of a highly active Cas13b ortholog for RNA knockdown

A) Schematic of stereotypical Cas13 loci and corresponding crRNA structure.

B) Evaluation of 19 Cas13a, 15 Cas13b, and 7 Cas13c orthologs for luciferase knockdown using two different guides. Orthologs with efficient knockdown using

both guides are labeled with their host organism name. Values are normalized to a non-targeting guide with designed against the *E. coli LacZ* transcript, with no homology to the human transcriptome.

C) PspCas13b and LwaCas13a knockdown activity (as measured by luciferase activity) using tiling guides against *Gluc*. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 1B.

D) PspCas13b and LwaCas13a knockdown activity (as measured by luciferase activity) using tiling guides against *Cluc*. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 1B.

E) Expression levels in $\log_2$(transcripts per million (TPM+1)) values of all genes detected in RNA-seq libraries of non-targeting control (x-axis) compared to *Gluc*-targeting condition (y-axis) for LwaCas13a (red) and shRNA (black). Shown is the mean of three biological replicates. The *Gluc* transcript data point is labeled. Non-targeting guide is the same as in Fig1B.

F) Expression levels in $\log_2$(transcripts per million (TPM+1)) values of all genes detected in RNA-seq libraries of non-targeting control (x-axis) compared to *Gluc*-targeting condition (y-axis) for PspCas13b (blue) and shRNA (black). Shown is the mean of three biological replicates. The *Gluc* transcript data point is labeled. Non-targeting guide is the same as in Fig. 1B.

G) Number of significant off-targets from *Gluc* knockdown for LwaCas13a, PspCas13b, and shRNA from the transcriptome wide analysis in E and F.

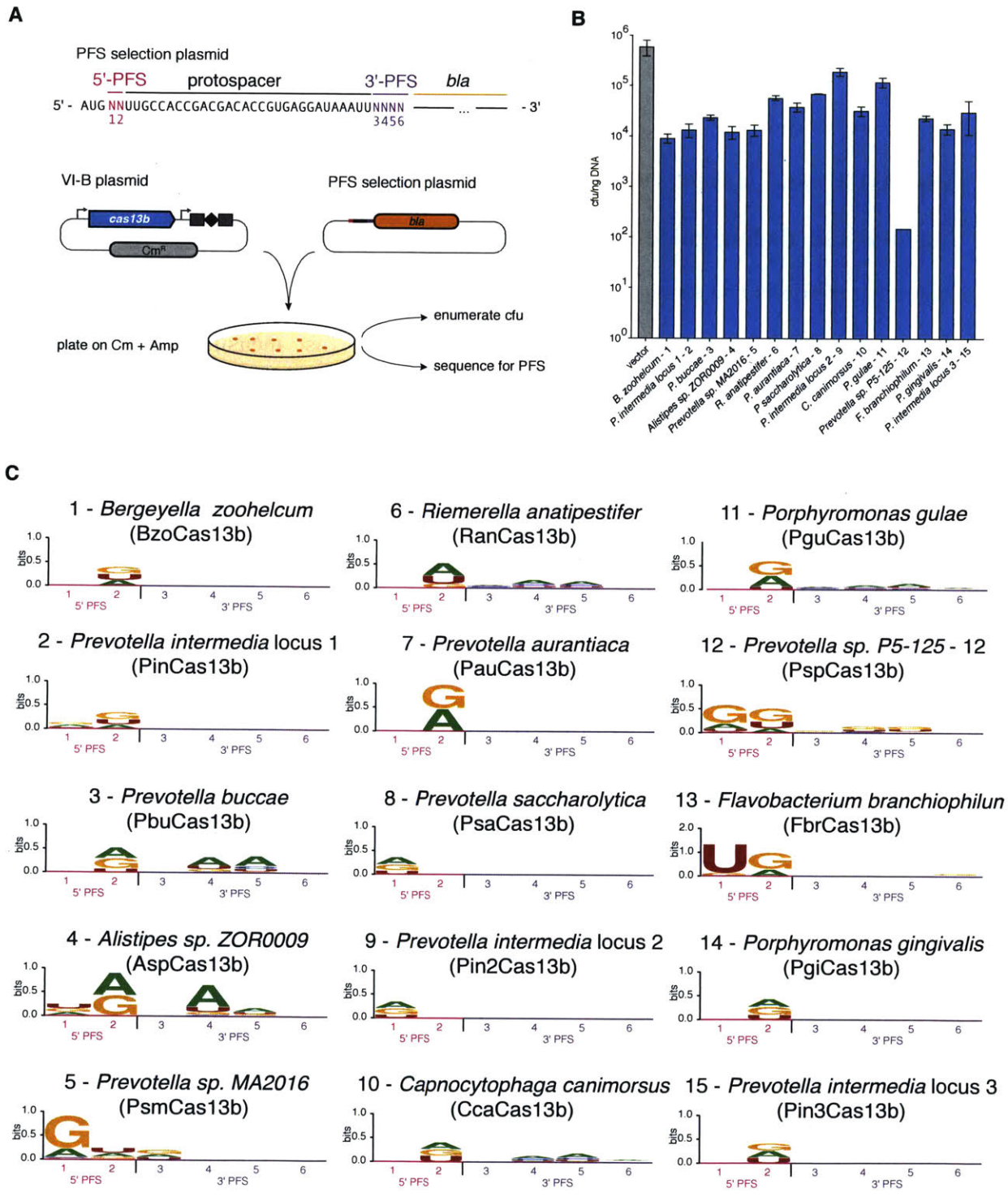

Figure S1: Bacterial screening of Cas13b orthologs for *in vivo* efficiency and PFS determination.

A) Schematic of bacterial assay for determining the PFS of Cas13b orthologs. Cas13b orthologs with beta-lactamase targeting spacers are co-transformed with beta-lactamase expression plasmids containing randomized PFS sequences and subjected to dual antibiotic selection. PFS sequences that are depleted during co-transformation with Cas13b suggest targeting activity and are used to infer PFS preferences.

B) Quantification of interference activity of Cas13b orthologs targeting beta-lactamase as measured by colony forming units (cfu). Values represent mean +/– S.D.

C) PFS weblogos for Cas13b orthologs as determined by depleted sequences from the bacterial assay. PFS preferences are derived from sequences depleted in the Cas13b condition relative to empty vector controls. Depletion values used to calculate PFS weblogos are listed in table S2.

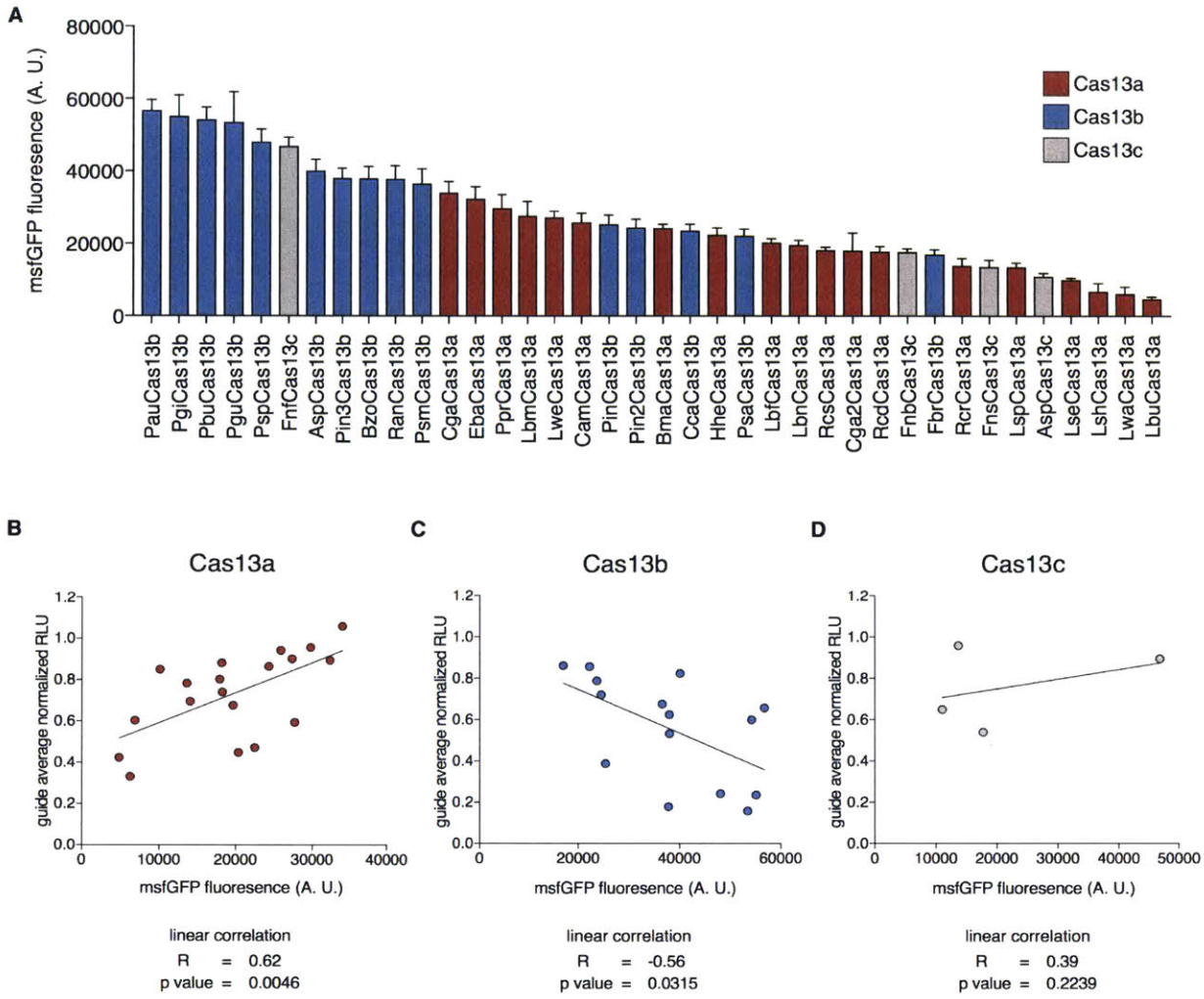

Figure S2: Relative expression of Cas13 orthologs in mammalian cells and correlation of expression with interference activity.

A) Expression of Cas13 orthologs as measured by msfGFP fluoresence. Cas13 orthologs C-terminally tagged with msfGFP were transfected into HEK293FT cells and their fluorescence measured 48 hours post transfection.

B) Correlation of Cas13 expression to interference activity. The average RLU of two *Gluc* targeting guides for Cas13 orthologs, separated by subfamily, is plotted versus expression as determined by msfGFP fluoresence. The RLU for targeting guides are normalized to RLU for a non-targeting guide, whose value is set to 1. The non-targeting guide is the same as in Figure 1B for Cas13b.

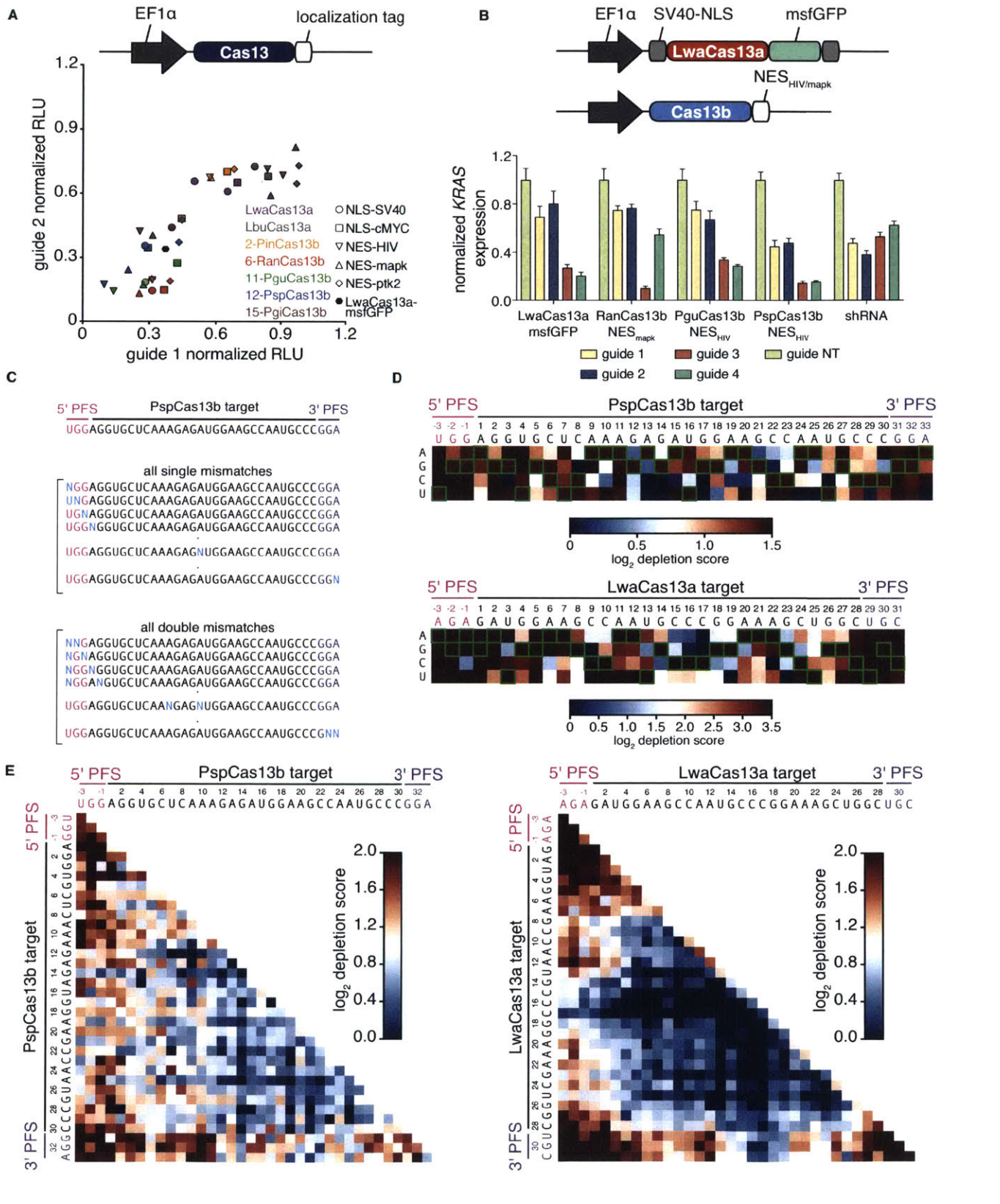
A
EF1α
localization tag
Cas13
guide 2 normalized RLU
guide 1 normalized RLU
LwaCas13a
LbuCas13a
2-PinCas13b
6-RanCas13b
11-PguCas13b
12-PspCas13b
15-PgiCas13b
NLS-SV40
NLS-cMYC
NES-HIV
NES-mapk
NES-ptk2
LwaCas13a-msfGFP
B
EF1α
SV40-NLS
msfGFP
LwaCas13a
NES HIV/mapk
Cas13b
normalized KRAS expression
LwaCas13a msfGFP
RanCas13b NES mapk
PguCas13b NES HIV
PspCas13b NES HIV
shRNA
guide 1
guide 2
guide 3
guide 4
guide NT
C
5' PFS
PspCas13b target
3' PFS
UGGAGGUGCUCAAAGAGAUGGAAGCCAAUGCCCGGA
all single mismatches
all double mismatches
D
PspCas13b target
LwaCas13a target
log2 depletion score
E
PspCas13b target
LwaCas13a target
log2 depletion score

Figure S3: Optimization of Cas13b knockdown and further characterization of mismatch specificity.

A) Gluc knockdown with two different guides is measured using the top two Cas13a and top four Cas13b orthologs fused to a variety of C-terminal nuclear-localization and nuclear-export tags.

B) Knockdown of *KRAS* is measured for LwaCas13a, RanCas13b, PguCas13b, PspCas13b and shRNA with four position-matched guides. Non-targeting guide is the same as in Figure 1B. shRNA non-targeting guide sequence is listed in table S6.

C) Schematic of the single and double mismatch plasmid libraries used for evaluating the specificity of LwaCas13a and PspCas13b knockdown. Every possible single and double mismatch is present in the target sequence as well as in three positions directly flanking the 5' and 3' ends of the target site.

D) The depletion levels of transcripts with the indicated single mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions. The wildtype base is outlined by a green box.

E) The depletion levels of transcripts with the indicated double mismatches are plotted as a heatmap for both the LwaCas13a and PspCas13b conditions. Each box represents the average of all possible double mismatches for the indicated position.

Specificity of Cas13 mammalian interference activity

To characterize the interference specificities of PspCas13b and LwaCas13a we designed a plasmid library of luciferase targets containing single mismatches and double mismatches throughout the target sequence and the three flanking 5' and 3' base pairs (fig. S3C). We transfected HEK293FT cells with either LwaCas13a or PspCas13b, a fixed guide RNA targeting the unmodified target sequence, and the mismatched-target library corresponding to the appropriate system. We then performed targeted RNA sequencing of uncleaved transcripts to quantify depletion of mismatched-target sequences. We found that LwaCas13a and PspCas13b had a central region that was relatively intolerant to single mismatches, extending from base pairs 12-26 for the PspCas13b target and 13-24 for the LwaCas13a target (fig. S3D). Double mismatches were even less tolerated than single mutations, with little knockdown activity observed over a larger window, extending from base pairs 12-29 for PspCas13b and 8-27 for LwaCas13a in their respective targets (fig. S3E). Additionally, because there are mismatches included in the three nucleotides flanking the 5' and 3' ends of the target sequence, we could assess PFS constraints on Cas13 knockdown activity. Sequencing showed that almost all PFS combinations allowed robust knockdown, indicating that a PFS constraint for interference in mammalian cells likely does not exist for either enzyme tested. These results indicate that Cas13a and Cas13b display similar sequence constraints and sensitivities against mismatches.

We next characterized the interference specificity of PspCas13b and LwaCas13a across the mRNA fraction of the transcriptome. We performed transcriptome-wide mRNA sequencing to detect significant differentially expressed genes. LwaCas13a and PspCas13b demonstrated robust knockdown of *Gluc* (Fig. 1E,F) and were highly specific compared to a position-matched shRNA, which showed hundreds of off-targets (Fig. 1G),

consistent with our previous characterization of LwaCas13a specificity in mammalian cells (*15*).

Cas13-ADAR fusions enable targeted RNA editing

Given that PspCas13b achieved consistent, robust, and specific knockdown of mRNA in mammalian cells, we envisioned that it could be adapted as an RNA binding platform to recruit RNA modifying domains, such as the deaminase domain of ADARs ($ADAR_{DD}$) for programmable RNA editing. To engineer a PspCas13b lacking nuclease activity (dPspCas13b, referred to as dCas13b hereafter), we mutated conserved catalytic residues in the HEPN domains and observed loss of luciferase RNA knockdown (fig. S4A). We hypothesized that a dCas13b-$ADAR_{DD}$ fusion could be recruited by a guide RNA to target adenosines, with the hybridized RNA creating the required duplex substrate for ADAR activity (Fig. 2A). To enhance target adenosine deamination rates we introduced two additional modifications to our initial RNA editing design: we introduced a mismatched cytidine opposite the target adenosine, which has been previously reported to increase deamination frequency, and fused dCas13b with the deaminase domains of human ADAR1 or ADAR2 containing hyperactivating mutations to enhance catalytic activity ($ADAR1_{DD}$(E1008Q) (*27*) or $ADAR2_{DD}$(E488Q) (*21*)).

To test the activity of dCas13b-$ADAR_{DD}$ we generated an RNA-editing reporter on *Cluc* by introducing a nonsense mutation (W85X (UGG->UAG)), which could functionally be repaired to the wildtype codon through A->I editing (Fig. 2B) and then be detected as restoration of Cluc luminescence. We evenly tiled guides with spacers of 30, 50, 70 or 84 nucleotides in length across the target adenosine to determine the optimal guide placement and design (Fig. 2C). We found that dCas13b-$ADAR1_{DD}$ required longer guides

to repair the *Cluc* reporter, while dCas13b-ADAR2$_{DD}$ was functional with all guide lengths tested (Fig. 2C). We also found that the hyperactive E488Q mutation improved editing efficiency, as luciferase restoration with the wildtype ADAR2$_{DD}$ was reduced (fig. S4B). From this demonstration of activity, we chose dCas13b-ADAR2$_{DD}$(E488Q) for further characterization and designated this approach as RNA Editing for Programmable A-to-I Replacement version 1 (REPAIRv1).

To validate that restoration of luciferase activity was due to *bona fide* editing events, we directly measured REPAIRv1-mediated editing of *Cluc* transcripts via reverse transcription and targeted next-generation sequencing. We tested 30- and 50-nt spacers around the target site and found that both guide lengths resulted in the expected A-to-I edit, with 50-nt spacers achieving higher editing percentages (Fig. 2D,E, fig. S4C). We also observed that 50-nt spacers had an increased propensity for editing at non-targeted adenosines within the sequencing window, likely due to increased regions of duplex RNA (Fig. 2E, fig. S4C).

We next targeted an endogenous gene, *PPIB*. We designed 50-nt spacers tiling *PPIB* and found that we could edit the *PPIB* transcript with up to 28% editing efficiency (Fig. S4D). To test if REPAIR could be further optimized, we modified the linker between dCas13b and ADAR2$_{DD}$(E488Q) (fig. S4E, Supplementary Table 3) and found that linker choice modestly affected luciferase activity restoration. Additionally, we tested the ability of dCas13b and guide alone to mediate editing events, finding that the ADAR deaminase domain is required for editing (fig. S5A-D).

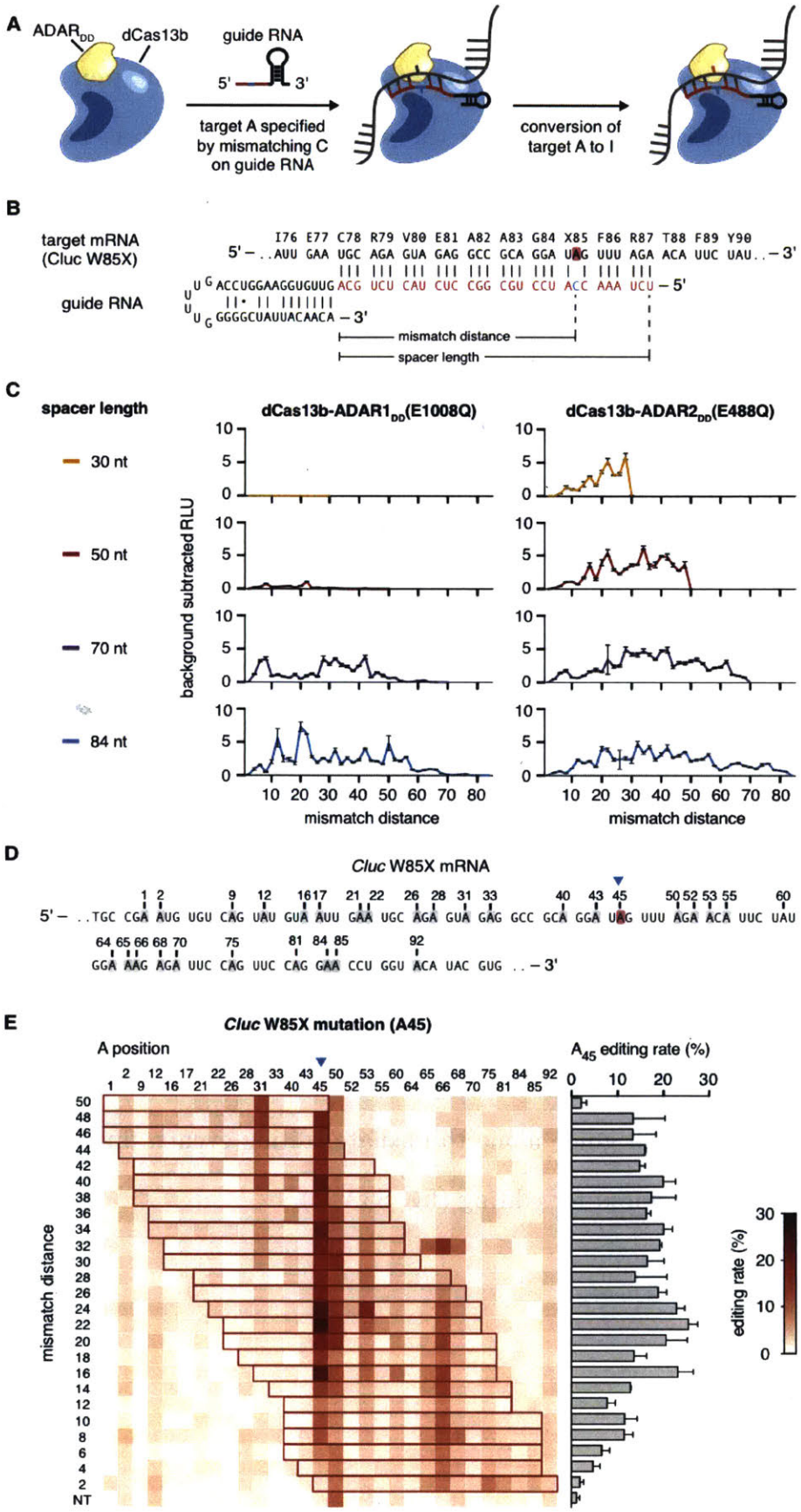
A
ADAR$_{DD}$
dCas13b
guide RNA
5'
3'
target A specified by mismatching C on guide RNA
conversion of target A to I
B
target mRNA (Cluc W85X)
guide RNA
I76 E77 C78 R79 V80 E81 A82 A83 G84 X85 F86 R87 T88 F89 Y90
5' – ..AUU GAA UGC AGA GUA GAG GCC GCA GGA UAG UUU AGA ACA UUC UAU.. – 3'
ACCUGGAAGGUGUUG ACG UCU CAU CUC CGG CGU CCU ACC AAA UCU – 5'
GGGGCUAUUACAACA – 3'
mismatch distance
spacer length
C
spacer length
30 nt
50 nt
70 nt
84 nt
dCas13b-ADAR1$_{DD}$(E1008Q)
dCas13b-ADAR2$_{DD}$(E488Q)
background subtracted RLU
mismatch distance
D
Cluc W85X mRNA
5' – ..TGC CGA AUG UGU CAG UAU GUA AUU GAA UGC AGA GUA GAG GCC GCA GGA UAG UUU AGA ACA UUC UAU
GGA AAG AGA UUC CAG UUC CAG GAA CCU GGU ACA UAC GUG .. – 3'
E
Cluc W85X mutation (A45)
A position
A$_{45}$ editing rate (%)
mismatch distance
editing rate (%)
NT

Figure 2: Engineering dCas13b-ADAR fusions for RNA editing

A) Schematic of RNA editing by dCas13b-ADAR$_{DD}$ fusion proteins. Catalytically dead Cas13b (dCas13b) is fused to the deaminase domain of human ADAR (ADAR$_{DD}$), which naturally deaminates adenosines to insosines in dsRNA. The crRNA specifies the target site by hybridizing to the bases surrounding the target adenosine, creating a dsRNA structure for editing, and recruiting the dCas13b-ADAR$_{DD}$ fusion. A mismatched cytidine in the crRNA opposite the target adenosine enhances the editing reaction, promoting target adenosine deamination to inosine, a base that functionally mimics guanosine in many cellular reactions.

B) Schematic of *Cypridina* luciferase W85X target and targeting guide design. Deamination of the target adenosine restores the stop codon to the wildtype tryptophan. Spacer length is the region of the guide that contains homology to the target sequence. Mismatch distance is the number of bases between the 3' end of the spacer and the mismatched cytidine. The cytidine mismatched base is included as part of the mismatch distance calculation.

C) Quantification of luciferase activity restoration for dCas13b-ADAR1$_{DD}$(E1008Q) (left) and dCas13b-ADAR2$_{DD}$(E488Q) (right) with tiling guides of length 30, 50, 70, or 84 nt. All guides with even mismatch distances are tested for each guide length. Values are background subtracted relative to a 30nt non-targeting guide that is randomized with no sequence homology to the human transcriptome.

D) Schematic of the sequencing window in which A-to-I edits were assessed for *Cypridinia* luciferase W85X.

E) Sequencing quantification of A-to-I editing for 50-nt guides targeting *Cypridinia* luciferase W85X. Blue triangle indicates the targeted adenosine. For each guide,

the region of duplex RNA is outlined in red. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 2C.

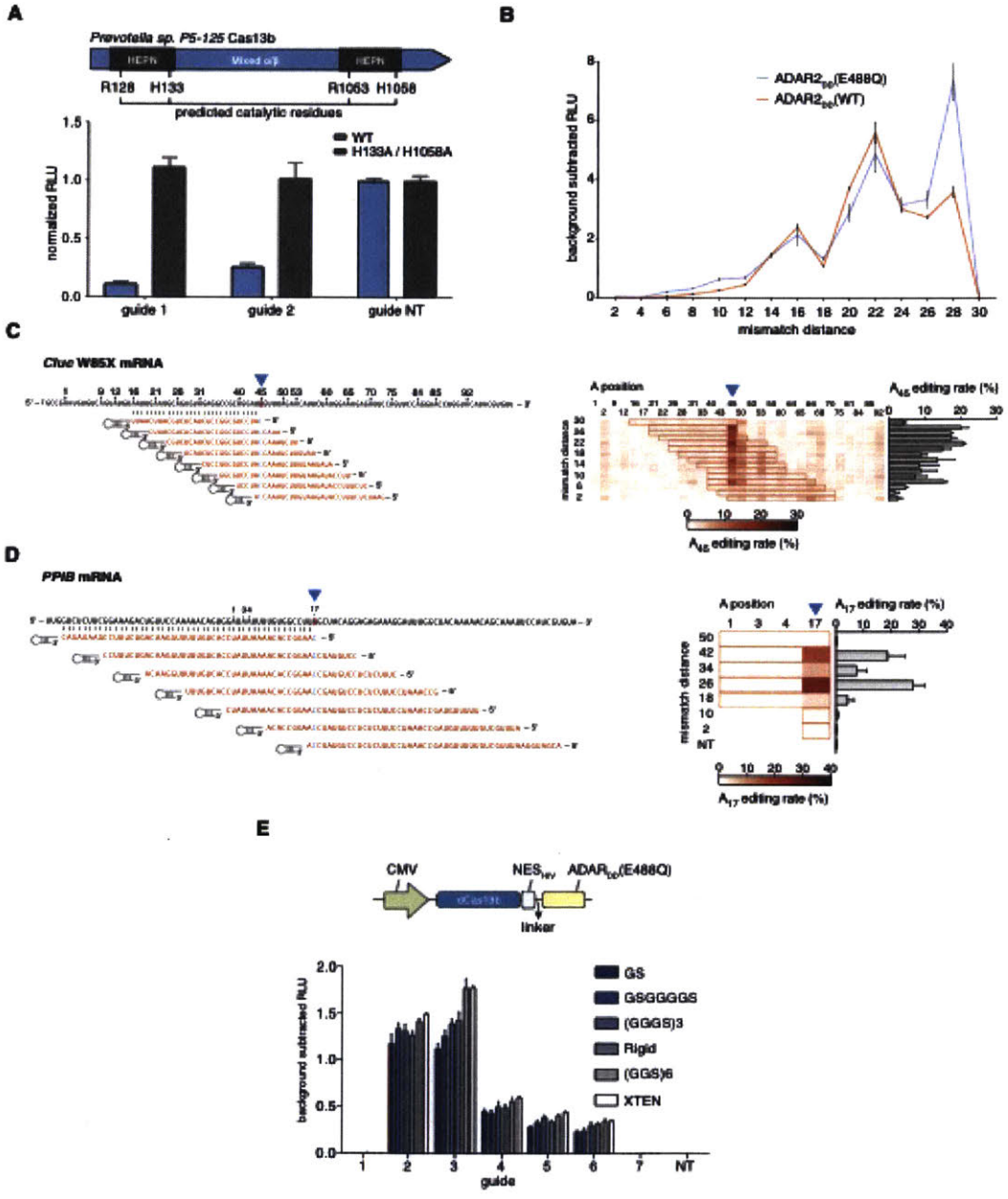

Figure S4: Characterization of design parameters for REPAIRv1.

A) Knockdown efficiency of Gluc with wild-type Cas13b or catalytically inactive H133A/H1058A Cas13b (dCas13b).

B) Quantification of luciferase activity restoration by dCas13b fused to either the wild-type ADAR2 deaminase domain ($ADAR2_{DD}$) or the hyperactive E488Q mutant $ADAR2_{DD}$(E488Q) deaminase domain, tested with tiling *Cluc* targeting guides.

C) Guide design and sequencing quantification of A-to-I editing for 30-nt guides targeting *Cluc* W85X.

D) Guide design and sequencing quantification of A-to-I editing for 50-nt guides targeting *PPIB*.

E) Influence of linker choice on luciferase activity restoration by REPAIRv1. Values represent mean +/− S.E.M.

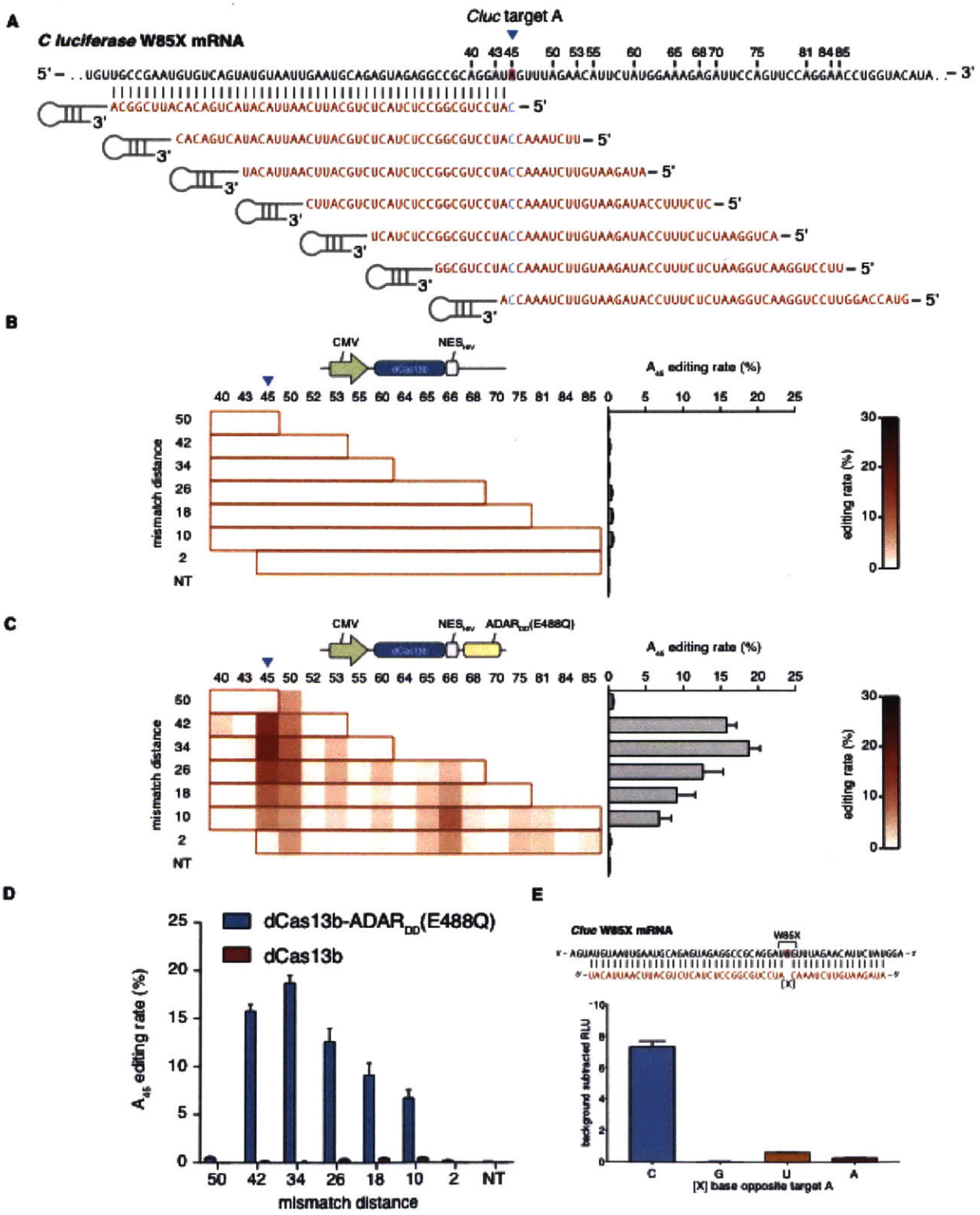

Figure S5: Comparison of RNA editing activity of dCas13b and REPAIRv1.

A) Schematic of guides used to target the W85X mutation in the *Cluc* reporter.

B) Sequencing quantification of A-to-I editing for indicated guides transfected with dCas13b. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig2C.

C) Sequencing quantification of A-to-I editing for indicated guides transfected with REPAIRv1. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig2C.

D) Comparison of on-target A-to-I editing rates for dCas13b and dCas13b-$ADAR2_{DD}$(E488Q) for guides tested in panel B and C.

E) Influence of base identify opposite the targeted adenosine on luciferase activity restoration by REPAIRv1. Values represent mean +/− S.E.M.

Defining the sequence parameters for RNA editing

Given that we could achieve precise RNA editing at a test site, we wanted to characterize the sequence constraints for programming the system against any RNA target in the transcriptome. Sequence constraints could arise from dCas13b targeting limitations, such as the PFS, or from ADAR sequence preferences (*28*). To investigate PFS constraints on REPAIRv1, we designed a plasmid library carrying a series of four randomized nucleotides at the 5' end of a target site on the *Cluc* transcript (Fig. 3A). We targeted the center adenosine within either a UAG or AAC motif and found that for both motifs, all PFSs demonstrated detectable levels of RNA editing, with a majority of the PFSs having greater than 50% editing at the target site (Fig. 3B). Next, we sought to determine if the $ADAR2_{DD}$ in REPAIRv1 had any sequence constraints immediately flanking the targeted base, as has been reported previously for $ADAR2_{DD}$ (*28*). We tested every possible combination of 5' and 3' flanking nucleotides directly surrounding the target adenosine (Fig. 3C), and found that REPAIRv1 was capable of editing all motifs (Fig. 3D). Lastly, we analyzed whether the identity of the base opposite the target A in the spacer sequence affected editing efficiency and found that an A-C mismatch had the highest luciferase restoration, in agreement with previous reports of ADAR2 activity, with A-G, A-U, and A-A having drastically reduced REPAIRv1 activity (fig. S5E).

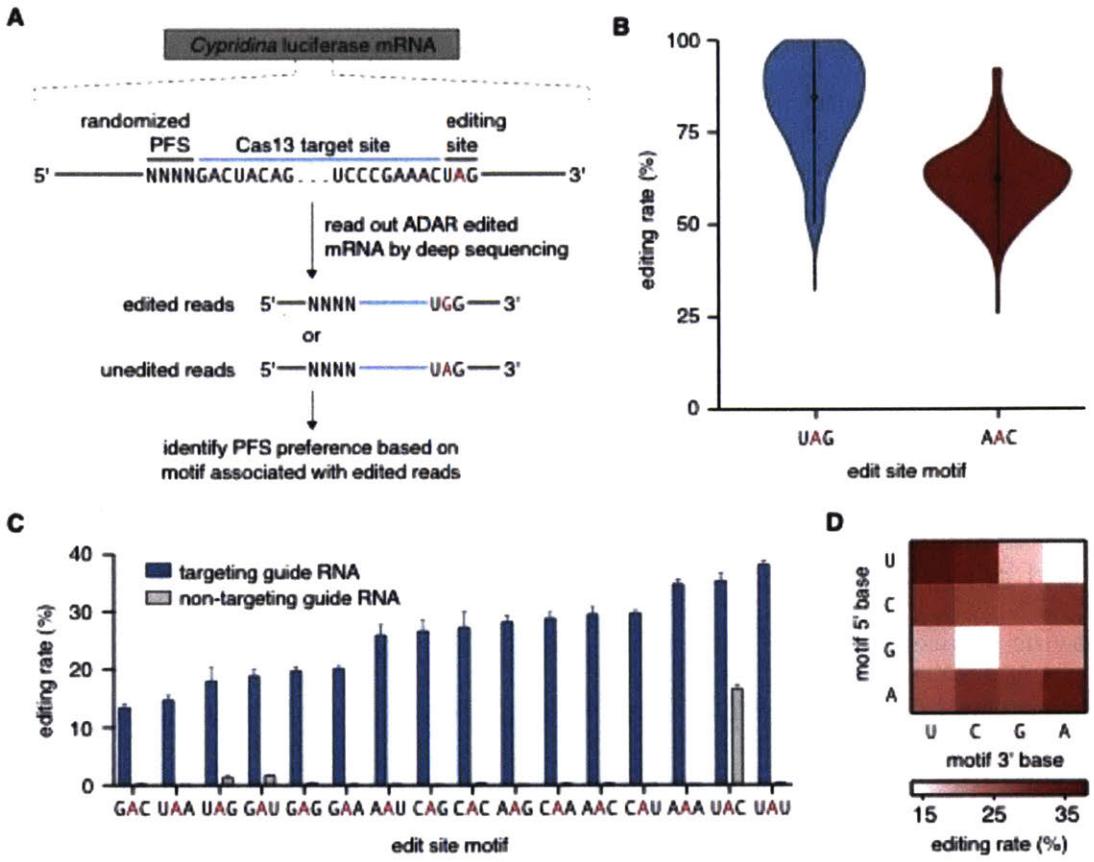

Figure 3: Measuring sequence flexibility for RNA editing by REPAIRv1

A) Schematic of screen for determining Protospacer Flanking Site (PFS) preferences of RNA editing by REPAIRv1. A randomized PFS sequence is cloned 5' to a target site for REPAIR editing. Following exposure to REPAIR, deep sequencing of reverse transcribed RNA from the target site and PFS is used to associate edited reads with PFS sequences.

B) Distributions of RNA editing efficiencies for all 4-N PFS combinations at two different editing sites

F) Quantification of the percent editing of REPAIRv1 at Cluc W85 across all possible 3 base motifs. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 2C.

C) Heatmap of 5' and 3' base preferences of RNA editing at Cluc W85 for all possible 3 base motifs.

Correction of disease-relevant human mutations using REPAIRv1

To demonstrate the broad applicability of the REPAIRv1 system for RNA editing in mammalian cells, we designed REPAIRv1 guides against two disease-relevant mutations: 878G>A (*AVPR2* W293X) in X-linked Nephrogenic diabetes insipidus and 1517G>A (*FANCC* W506X) in Fanconi anemia. We transfected expression constructs for cDNA of genes carrying these mutations into HEK293FT cells and tested whether REPAIRv1 could correct the mutations. Using guide RNAs containing 50-nt spacers, we were able to achieve 35% correction of *AVPR2* and 23% correction of *FANCC* (Fig. 4A-D). We then tested the ability of REPAIRv1 to correct 34 different disease-relevant G>A mutations (Supplementary Table 4) and found that we were able to achieve significant editing at 33 sites with up to 28% editing efficiency (Fig. 4E). The mutations we chose are only a fraction of the pathogenic G to A mutations (5,739) in the ClinVar database, which also includes an additional 11,943 G to A variants (Fig. 4F and fig. S6). Because there are no sequence constraints (Fig. 3), REPAIRv1 is capable of potentially editing all these disease relevant mutations, especially given that we observed editing regardless of the target motif (Fig. 3C and Fig. 4G).

Delivering the REPAIRv1 system to diseased cells is a prerequisite for therapeutic use, and we therefore sought to design REPAIRv1 constructs that could be packaged into therapeutically relevant viral vectors, such as adeno-associated viral (AAV) vectors. AAV vectors have a packaging limit of 4.7kb, which cannot accommodate the large size of dCas13b-ADAR$_{DD}$ (4,473 bp) along with promoter and expression-regulatory elements. To reduce the size, we tested a variety of N-terminal and C-terminal truncations of dCas13 fused to ADAR2$_{DD}$(E488Q) for RNA editing activity. We found that all C-terminal truncations tested were still functional and able to restore luciferase signal (fig. S7), and

the largest truncation, C-terminal Δ984-1090 (total size of the fusion protein 4,152bp) was small enough to fit within the packaging limit of AAV vectors.

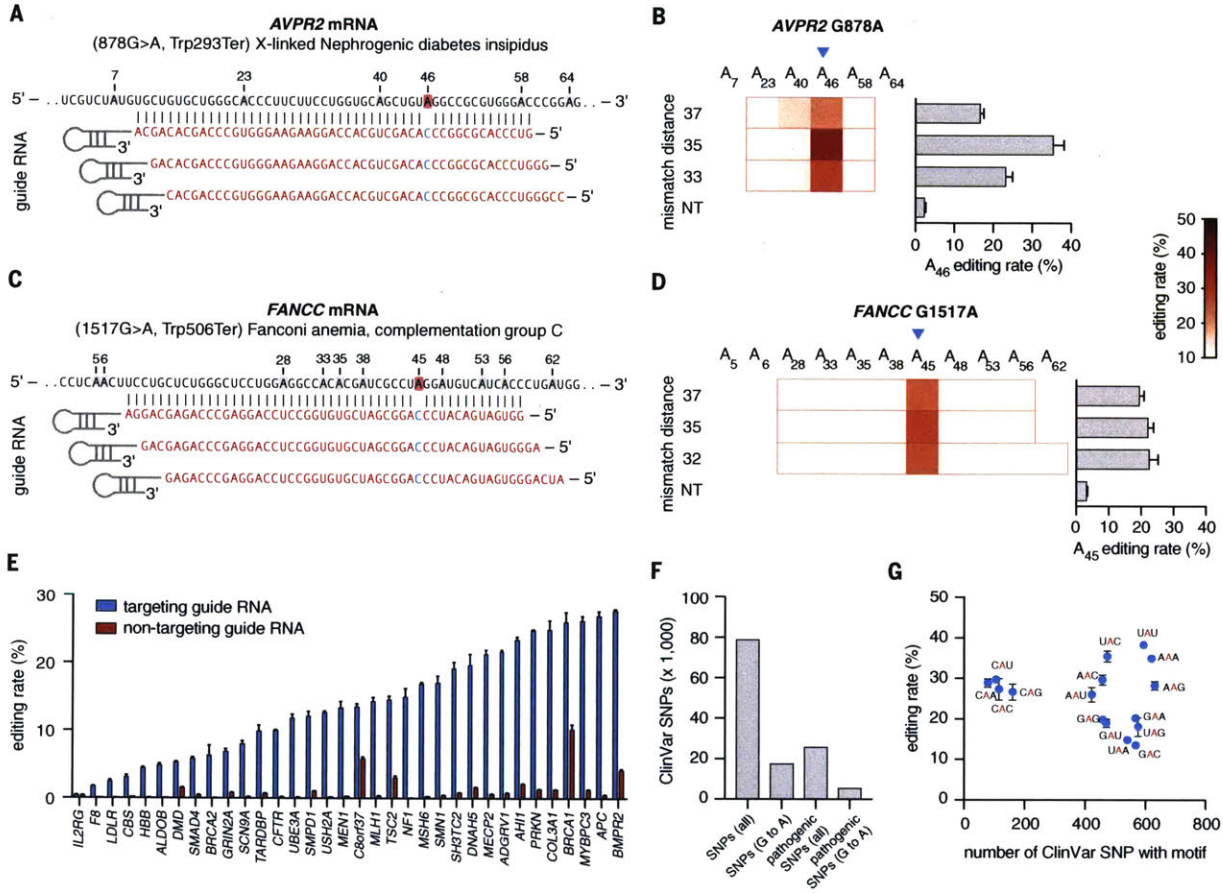

Figure 4: Correction of disease-relevant mutations with REPAIRv1

A) Schematic of target and guide design for targeting *AVPR2* 878G>A.

B) The 878G>A mutation (indicated by blue triangle) in *AVPR2* is corrected to varying levels using REPAIRv1 with three different guide designs. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 2C.

C) Schematic of target and guide design for targeting *FANCC* 1517G>A.

D) The 1517G>A mutation (indicated by blue triangle) in *FANCC* is corrected to varying levels using REPAIRv1 with three different guide designs. For each guide, the region of duplex RNA is outlined in red. The heatmap scale bar is the same as

in panel B. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 2C.

E) Quantification of the percent editing of 34 different disease-relevant G>A mutations selected from ClinVar using REPAIRv1. Non-targeting guide is the same as in Fig. 2C.

F) Analysis of all the possible G>A mutations that could be corrected using REPAIR as annotated in the ClinVar database.

G) The distribution of editing motifs for all G>A mutations in ClinVar is shown versus the editing efficiency by REPAIRv1 per motif as quantified on the *Gluc* transcript. Values represent mean +/− S.E.M.

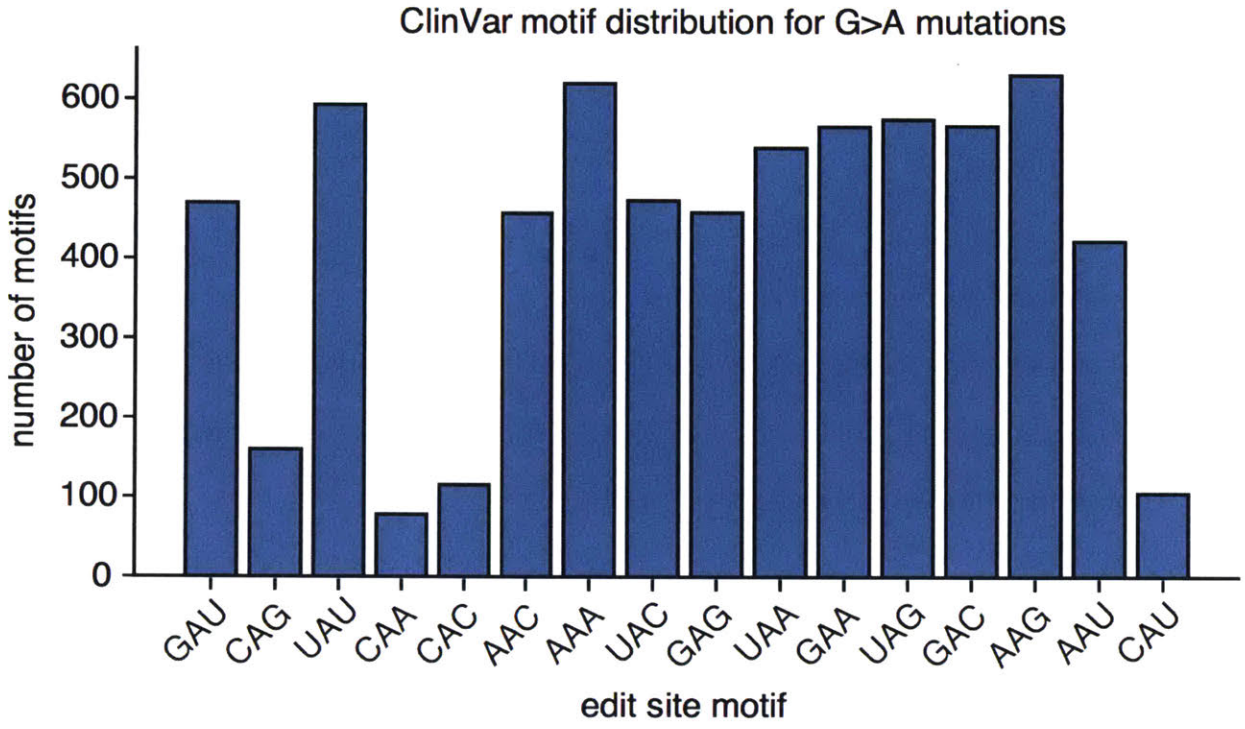

Figure S6: ClinVar motif distribution for G>A mutations.

The number of each possible triplet motif observed in the ClinVar database for all G>A mutations.

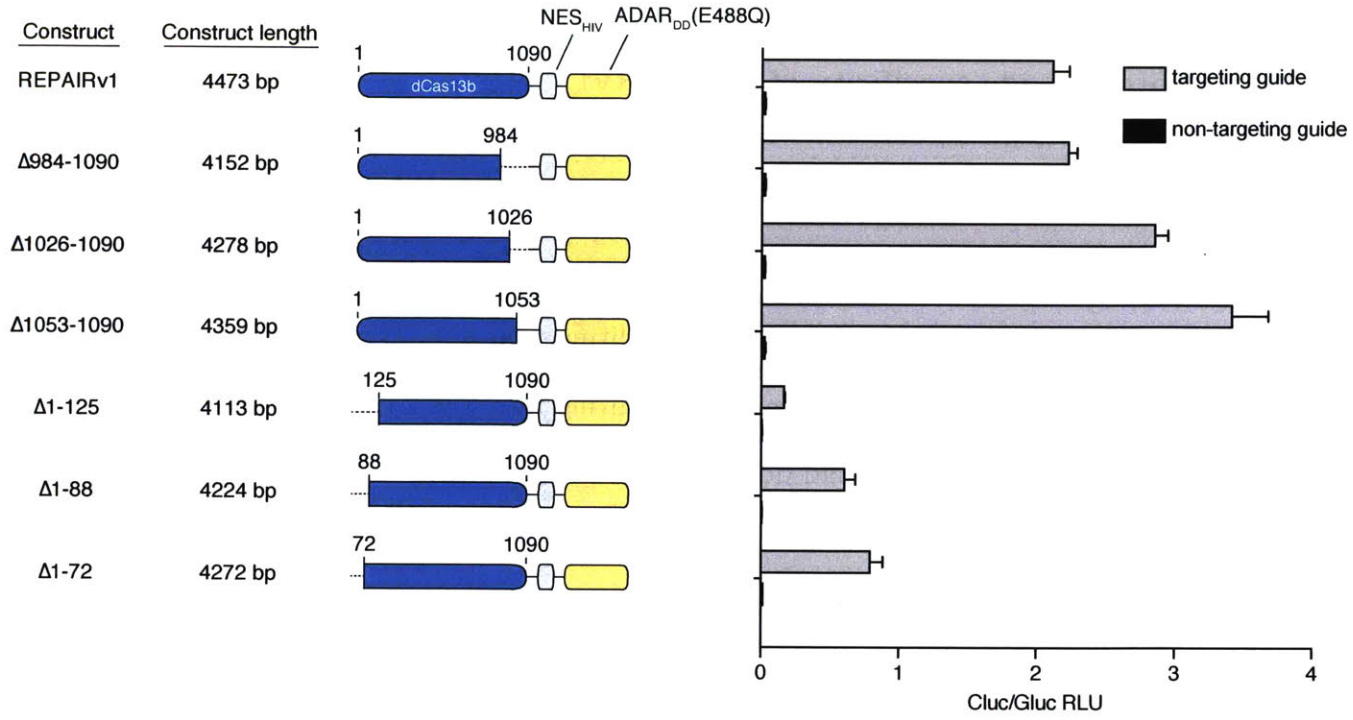

Figure S7: Truncations of dCas13b support functional RNA editing.

N-terminal and C-terminal truncations of dCas13b allow for RNA editing as measured by restoration of luciferase signal for the *Cluc* W85X reporter. Values represent mean +/− S.E.M. The construct length refers to the coding sequence of the REPAIR constructs.

Transcriptome-wide specificity of REPAIRv1

Although RNA knockdown with PspCas13b was highly specific in our luciferase tiling experiments, we observed off-target adenosine editing within the guide:target duplex (Fig. 2E). To see if this was a widespread phenomenon, we tiled an endogenous transcript, *KRAS*, and measured the degree of off-target editing near the target adenosine (Fig. 5A). We found that for *KRAS*, while the on-target editing rate was 23%, there were many sites around the target site that also had detectable A-to-I edits (Fig. 5B).

Because of the observed off-target editing within the guide:target duplex, we initially evaluated transcriptome-wide off-targets by performing RNA sequencing on all mRNAs with 12.5X coverage. Of all the editing sites across the transcriptome, the on-target editing site had the highest editing rate, with 89% A-to-I conversion. We also found that there was a substantial number of A-to-I off-target events, with 1,732 off-targets in the targeting guide condition and 925 off-targets in the non-targeting guide condition, with 828 off-targets shared between the targeting and non-targeting guide conditions (Fig. 5C,D). Given the high number of overlapping off-targets between the targeting and non-targeting guide conditions, we reasoned that the off-targets may arise from $ADAR_{DD}$. To test this hypothesis, we repeated the Cluc targeting experiment, this time comparing transcriptome changes for REPAIRv1 with a targeting guide, REPAIRv1 with a non-targeting guide, REPAIRv1 alone, or $ADAR_{DD}$(E488Q) alone (fig. S8). We found differentially expressed genes and off-target editing events in each condition (fig. S8B,C). Interestingly, there was a high degree of overlap in the off-target editing events between $ADAR_{DD}$(E488Q) and all REPAIRv1 off-target edits, supporting the hypothesis that REPAIR off-target edits are driven by dCas13b-independent $ADAR_{DD}$(E488Q) editing events (fig. S8D).

Next, we sought to compare two RNA-guided ADAR systems that have been described previously (fig. S9A). The first utilizes a fusion of $ADAR2_{DD}$ to the small viral protein lambda N (λN), which binds to the BoxB-λ RNA hairpin (*24*). A guide RNA with double BoxB-λ hairpins guides $ADAR2_{DD}$ to edit sites encoded in the guide RNA (*25*). The second design utilizes full-length ADAR2 (ADAR2) and a guide RNA with a hairpin that the double strand RNA binding domains (dsRBDs) of ADAR2 recognize (*23, 26*). We analyzed the editing efficiency of these two systems compared to REPAIRv1 and found that the BoxB-ADAR2 and full-length ADAR2 systems demonstrated 50% and 34.5% editing rates, respectively, compared to the 89% editing rate achieved by REPAIRv1 (fig. S9B-E). Additionally, the BoxB and full-length ADAR2 systems created 1,814 and 66 observed off targets, respectively, in the targeting guide conditions, compared to the 2,111 off targets in the REPAIRv1 targeting guide condition. Notably, all the conditions with the two $ADAR2_{DD}$-based systems (REPAIRv1 and BoxB) showed a high percentage of overlap in their off-targets whereas the full-length ADAR2 system had a largely distinct set of off-targets (fig. S9F). The overlap in off-targets between the targeting and non-targeting conditions and between REPAIRv1 and BoxB conditions suggests $ADAR2_{DD}$ drives off-targets independent of dCas13 targeting (fig. S9F).

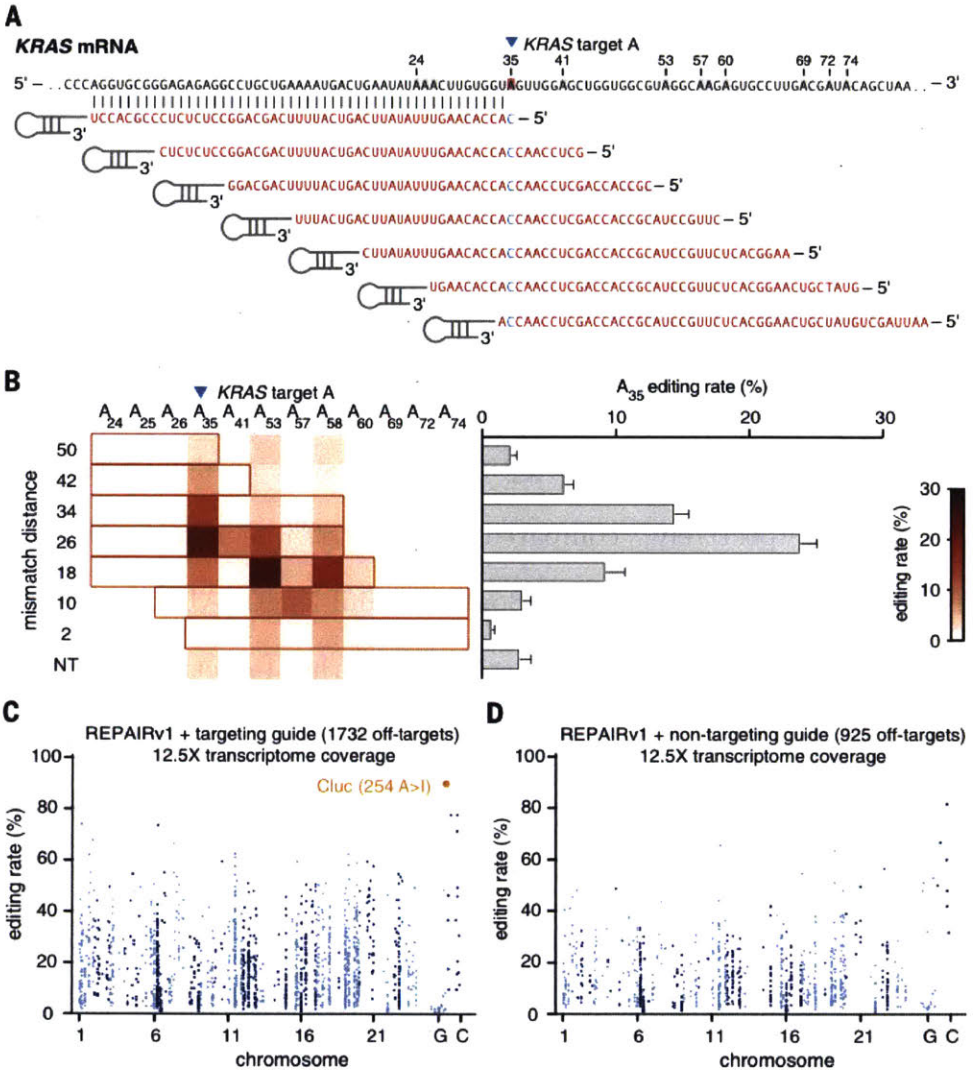

Figure 5: Characterizing specificity of REPAIRv1

A) Schematic of *KRAS* target site and guide design.

B) Quantification of percent A-to-I editing for tiled *KRAS*-targeting guides. Editing percentages are shown for the on-target (blue triangle) and neighboring adenosine sites. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/− S.E.M.

C) Transcriptome-wide sites of significant RNA editing by REPAIRv1 (150ng REPAIR vector transfected) with *Cluc* targeting guide. The on-target site *Cluc* site (254 A>I) is highlighted in orange.

D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 (150ng REPAIR vector transfected) with non-targeting guide. Non-targeting guide is the same as in Fig. 2C.

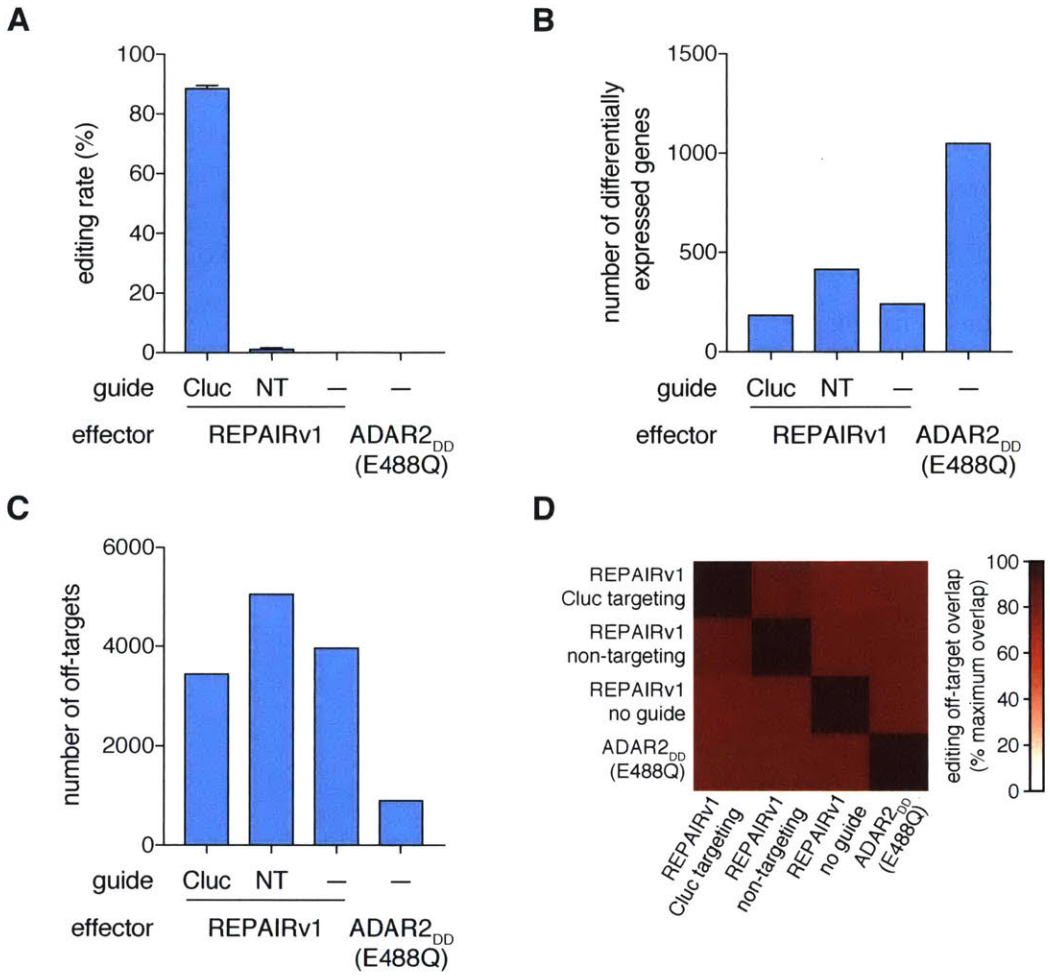

Figure S8: REPAIRv1 editing activity evaluated without a guide and in comparison to ADAR2 deaminase domain alone.

A) Quantification of A-to-I editing of the *Cluc* W85X mutation by REPAIRv1 with and without guide as well as the ADAR2 deaminase domain only without guide. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig2C.

B) Number of differentially expressed genes in the REPAIRv1 and $ADAR2_{DD}$ conditions from panel A.

C) The number of significant off-targets from the REPAIRv1 and $ADAR2_{DD}$ conditions from panel A.

D) Overlap of off-target A-to-I editing events between the REPAIRv1 and $ADAR2_{DD}$ conditions from panel A. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

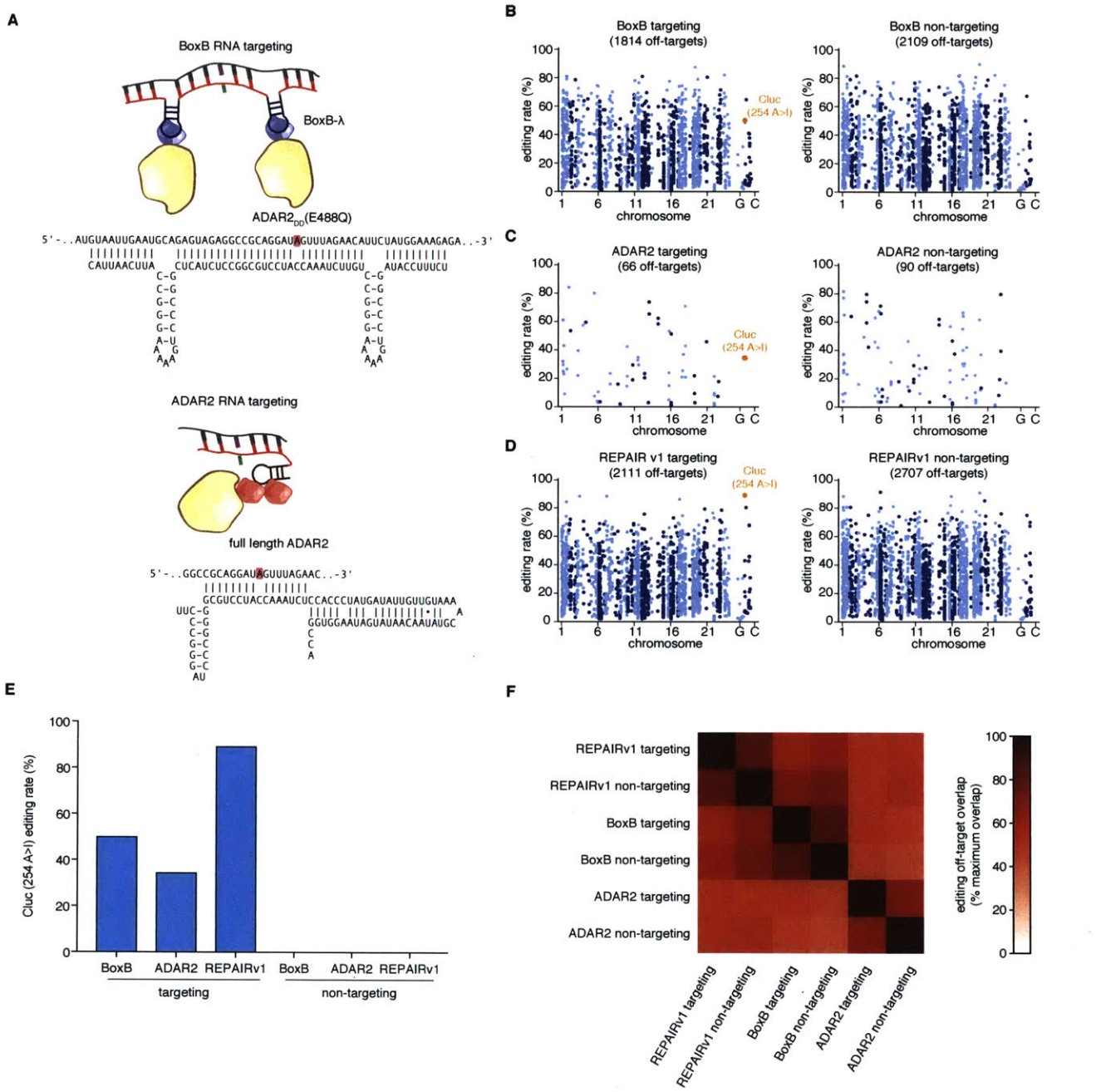

Figure S9: Comparison of REPAIRv1 to other programmable ADAR systems.

A) Schematic of two programmable ADAR schemes: BoxB-based targeting (top) and full length ADAR2 targeting (bottom). For BoxB-based targeting, $ADAR_{DD}$(E488Q) is fused to the viral protein lambda N (BoxB-λ), and the fusion protein is recruited to target adenosines by a guide RNA containing homology to

the target site and hairpins that BoxB-λ binds to. Full length ADAR2 targeting utilizes a guide RNA with homology to the target site and a motif recognized by the double strand RNA binding domains of ADAR2.

B) Transcriptome-wide sites of significant RNA editing by BoxB-ADAR2 $_{DD}$(E488Q) with a guide targeting *Cluc* and a non-targeting guide. The on-target *Cluc* site (254 A>I) is highlighted in orange.

C) Transcriptome-wide sites of significant RNA editing by full length ADAR2 with a guide targeting *Cluc* and a non-targeting guide. The on-target *Cluc* site (254 A>I) is highlighted in orange.

D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 with a guide targeting *Cluc* and a non-targeting guide. The on-target *Cluc* site (254 A>I) is highlighted in orange. The non-targeting guide is the same as in Fig2C.

E) Quantification of on-target editing rate percentage for BoxB-ADAR2 $_{DD}$(E488Q), ADAR2, and REPAIRv1 for targeting guides against *Cluc*.

F) Overlap of off-target sites between different targeting and non-targeting conditions for programmable ADAR systems. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

Improving specificity of REPAIR through rational protein engineering

To improve the specificity of REPAIRv1, we employed structure-guided protein engineering of $ADAR2_{DD}$(E488Q). Because of the guide-independent nature of the off-targets, we hypothesized that destabilizing $ADAR2_{DD}$(E488Q)-RNA binding would selectively decrease off-target editing, but maintain on-target editing due to increased local concentration from dCas13b tethering of $ADAR2_{DD}$(E488Q) to the target site. We mutated residues in $ADAR2_{DD}$(E488Q) previously determined to contact the duplex region of the target RNA (Fig. 6A) (*19*). To assess efficiency and specificity, we tested 17 single mutants with both targeting and non-targeting guides, under the assumption that background luciferase restoration in the non-targeting condition would be indicative of broader off-target activity. We found that mutations at the selected residues had significant effects on the luciferase activity for targeting and non-targeting guides (Fig. 6A,B, fig. S10A). A majority of mutants either significantly improved the luciferase activity for the targeting guide or increased the ratio of targeting to non-targeting guide activity, which we termed the specificity score (Fig. 6A,B).

We selected a subset of these mutants (Fig. 6B) for transcriptome-wide specificity profiling by next generation sequencing. As expected, off-targets measured from transcriptome-wide sequencing correlated with our specificity score (fig. S10B) for mutants. We found that with the exception of $ADAR2_{DD}$(E488Q/R455E), all sequenced REPAIRv1 mutants could effectively edit the reporter transcript (Fig. 6C), with many mutants showing reduction in the number of off-targets (Fig. 6C, fig S10C, S11). We further explored the surrounding motifs of off-targets for the various specificity mutants, and found that REPAIRv1 and most of the engineered variants exhibited a strong 3' G

preference for their off-target edits, in agreement with the characterized ADAR2 motif (fig. S12A) (*28*).

We focused on the mutant ADAR2$_{DD}$(E488Q/T375G), as it had the highest percent editing of the four mutants with the lowest numbers of transcriptome-wide off targets and termed it REPAIRv2. Compared to REPAIRv1, REPAIRv2 exhibited increased specificity, with a reduction from 18,385 to 20 transcriptome-wide off-targets by high-coverage sequencing (125X coverage, 10ng DNA transfection) (Fig. 6D). In the region surrounding the targeted adenosine in *Cluc*, REPAIRv2 also had reduced off-target editing, visible in sequencing traces (Fig. 6E). In motifs derived from the off-target sites, REPAIRv1 presented a strong preference towards 3' G, but showed off-targeting edits for all motifs (fig. S12B); by contrast, REPAIRv2 only edited the strongest off-target motifs (fig. S12C). The distribution of edits on transcripts was heavily skewed for REPAIRv1, with highly-edited genes having over 60 edits (fig. S13A), whereas REPAIRv2 only edited one transcript (*EEF1A1*) multiple times (fig. S13B). REPAIRv1 off-target edits were predicted to result in numerous variants, including 1000 missense base changes (fig. S13C) with 93 events in genes related to cancer processes (fig. S13D). In contrast, REPAIRv2 only had 6 predicted base changes (fig. S10E), none of which were in cancer-related genes (fig. S13F). Analysis of the sequence surrounding off-target edits for REPAIRv1 or v2 did not reveal homology to guide sequences, suggesting that off-targets are likely dCas13b-independent (fig. S14), consistent with the high overlap of off-targets between REPAIRv1 and the ADAR deaminase domain (fig. S8D). To directly compare REPAIRv2 against other programmable ADAR systems, we repeated our *Cluc* targeting experiments with all systems at two different dosages of ADAR vector, finding that REPAIRv2 had comparable on-target editing to BoxB and ADAR2 but with significantly fewer off-target editing events at both dosages (fig S15). REPAIRv2 had enhanced specificity compared

to REPAIRv1 at both dosages (fig. S15B), a finding that also extended to two guides targeting distinct sites on *PPIB* (fig. S16A-D). It is also worth noting that, in general, the lower dosage condition (10 ng) had fewer off-targets than the higher dosage condition (150 ng) (fig. S5).

To assess editing specificity with greater sensitivity, we sequenced the low dosage condition (10 ng of transfected DNA) of REPAIRv1 and v2 at significantly higher sequencing depth (125X coverage of the transcriptome). Increased numbers of off-targets were found at higher sequencing depths corresponding to detection of rarer off-target events (fig. S17). Furthermore, we speculated that different transcriptome states could also potentially alter the number of off-targeting events. Therefore, we tested REPAIRv2 activity in the osteosarcoma U2OS cell line, observing 6 and 7 off-targets for the targeting and non-targeting guide, respectively (fig. S18).

We targeted REPAIRv2 to endogenous genes to test if the specificity-enhancing mutations reduced nearby edits in target transcripts while maintaining high-efficiency on-target editing. For guides targeting either *KRAS* or *PPIB*, we found that REPAIRv2 had no detectable off-target edits, unlike REPAIRv1, and could effectively edit the on-target adenosine at efficiencies of 27.1% (*KRAS*) or 13% (*PPIB*) (Fig. 6F). This specificity extended to additional target sites, including regions that demonstrate high-levels of background in non-targeting conditions for REPAIRv1, such as other *KRAS* or *PPIB* target sites (fig. S19). Overall, REPAIRv2 eliminated off-targets in duplexed regions around the edited adenosine and showed dramatically enhanced transcriptome-wide specificity.

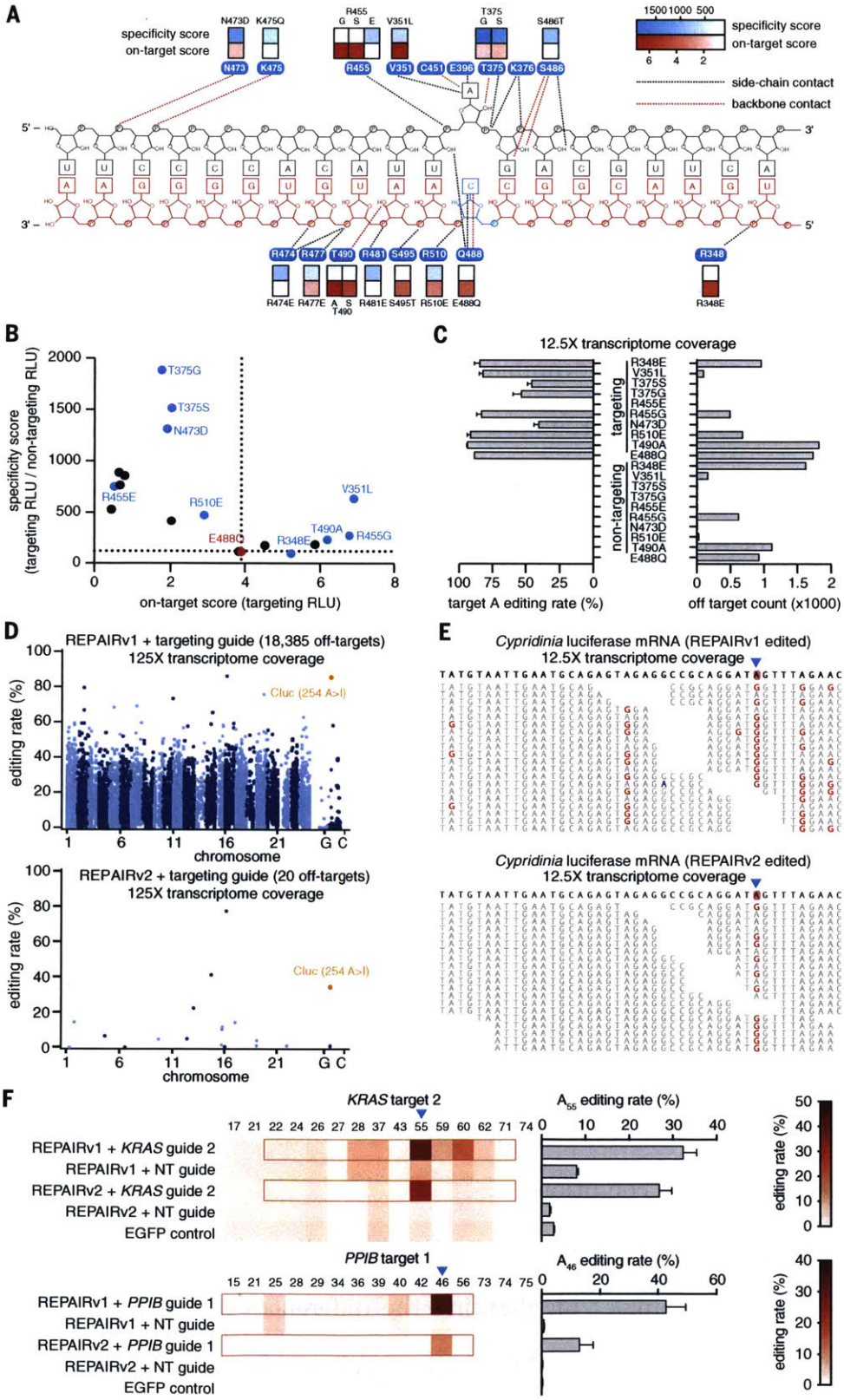
A
specificity score
on-target score
N473D K475Q
R455 G S E
V351L
T375 G S
S486T
1500 1000 500
specificity score
on-target score
6 4 2
N473 K475
R455 V351 C451 E396 T375 K376 S486
side-chain contact
backbone contact
5'
3'
3'
5'
R474 R477 T490 R481 S495 T510 Q488
R348
R474E R477E A S T490 R481E S495T R510E E488Q
R348E
B
specificity score (targeting RLU / non-targeting RLU)
on-target score (targeting RLU)
T375G
T375S
N473D
R455E
R510E
V351L
T490A
R455G
E488Q
R348E
C
12.5X transcriptome coverage
targeting
non-targeting
R348E
V351L
T375S
T375G
R455E
R455G
N473D
R510E
T490A
E488Q
target A editing rate (%)
off target count (x1000)
D
REPAIRv1 + targeting guide (18,385 off-targets)
125X transcriptome coverage
Cluc (254 A>I)
editing rate (%)
chromosome
REPAIRv2 + targeting guide (20 off-targets)
125X transcriptome coverage
Cluc (254 A>I)
editing rate (%)
chromosome
E
Cypridinia luciferase mRNA (REPAIRv1 edited)
12.5X transcriptome coverage
Cypridinia luciferase mRNA (REPAIRv2 edited)
12.5X transcriptome coverage
F
KRAS target 2
A55 editing rate (%)
REPAIRv1 + KRAS guide 2
REPAIRv1 + NT guide
REPAIRv2 + KRAS guide 2
REPAIRv2 + NT guide
EGFP control
editing rate (%)
PPIB target 1
A46 editing rate (%)
REPAIRv1 + PPIB guide 1
REPAIRv1 + NT guide
REPAIRv2 + PPIB guide 1
REPAIRv2 + NT guide
EGFP control
editing rate (%)

Figure 6: Rational mutagenesis of ADAR2 to improve the specificity of REPAIRv1

A) Quantification of luciferase signal restoration (on-target score, red boxes) by various dCas13-ADAR2$_{DD}$ mutants as well as their specificity score (blue boxes) plotted along a schematic of the contacts between key ADAR2 deaminase residues and the dsRNA target (target strand shown in gray; the non-target strand is shown in red). All deaminase mutations were made on the dCas13-ADAR2$_{DD}$(E488Q) background. The specificity score is defined as the ratio of the luciferase signal between targeting guide and non-targeting guide conditions. Schematic of ADAR2 deaminase domain contacts with dsRNA is adapted from ref (*20*).

B) Quantification of luciferase signal restoration by various dCas13-ADAR2 mutants versus their specificity score. Non-targeting guide is the same as in Fig. 2C.

C) Quantification of on-target editing and the number of significant off-targets for each dCas13-ADAR2$_{DD}$(E488Q) mutant by transcriptome wide sequencing of mRNAs. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 2C.

D) Transcriptome-wide sites of significant RNA editing by REPAIRv1 (top) and REPAIRv2 (bottom) with a guide targeting a pretermination site in *Cluc*. The on-target *Cluc* site (254 A>I) is highlighted in orange. 10 ng of REPAIR vector was transfected for each condition.

E) Representative RNA sequencing reads surrounding the on-target *Cluc* editing site (254 A>I; blue triangle) highlighting the differences in off-target editing between REPAIRv1 (top) and REPAIRv2 (bottom). A>I edits are highlighted in red;

sequencing errors are highlighted in blue. Gaps reflect spaces between aligned reads. Non-targeting guide is the same as in Fig. 2C.

F) RNA editing by REPAIRv1 and REPAIRv2 with guides targeting an out-of-frame UAG site in the endogenous *KRAS* and *PPIB* transcripts. The on-target editing fraction is shown as a sideways bar chart on the right for each condition row. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/− S.E.M. Non-targeting guide is the same as in Fig. 2C.

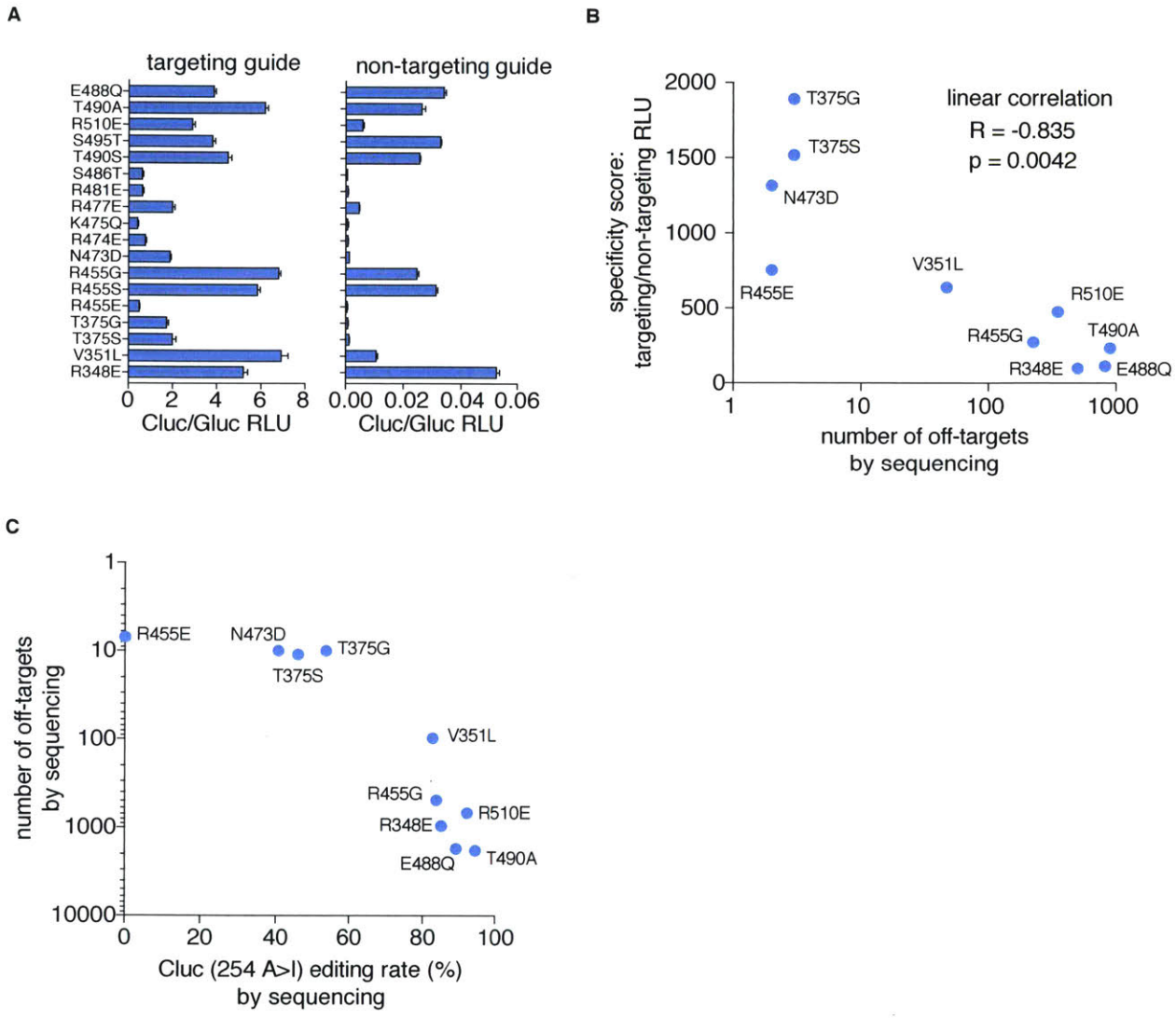

Figure S10: Efficiency and specificity of dCas13b-ADAR2$_{DD}$(E488Q) mutants.

A) Quantification of luciferase activity restoration by dCas13b-ADAR2$_{DD}$(E488Q) mutants for *Cluc*-targeting and non-targeting guides. Non-targeting guide is the same as in Fig2C.

B) Relationship between the ratio of targeting and non-targeting guide RLU and the number of RNA-editing off-targets as quantified by transcriptome-wide sequencing

C) Quantification of transcriptome-wide off-target RNA editing sites versus on-target

Cluc editing efficiency for dCas13b-ADAR2$_{DD}$(E488Q) mutants.

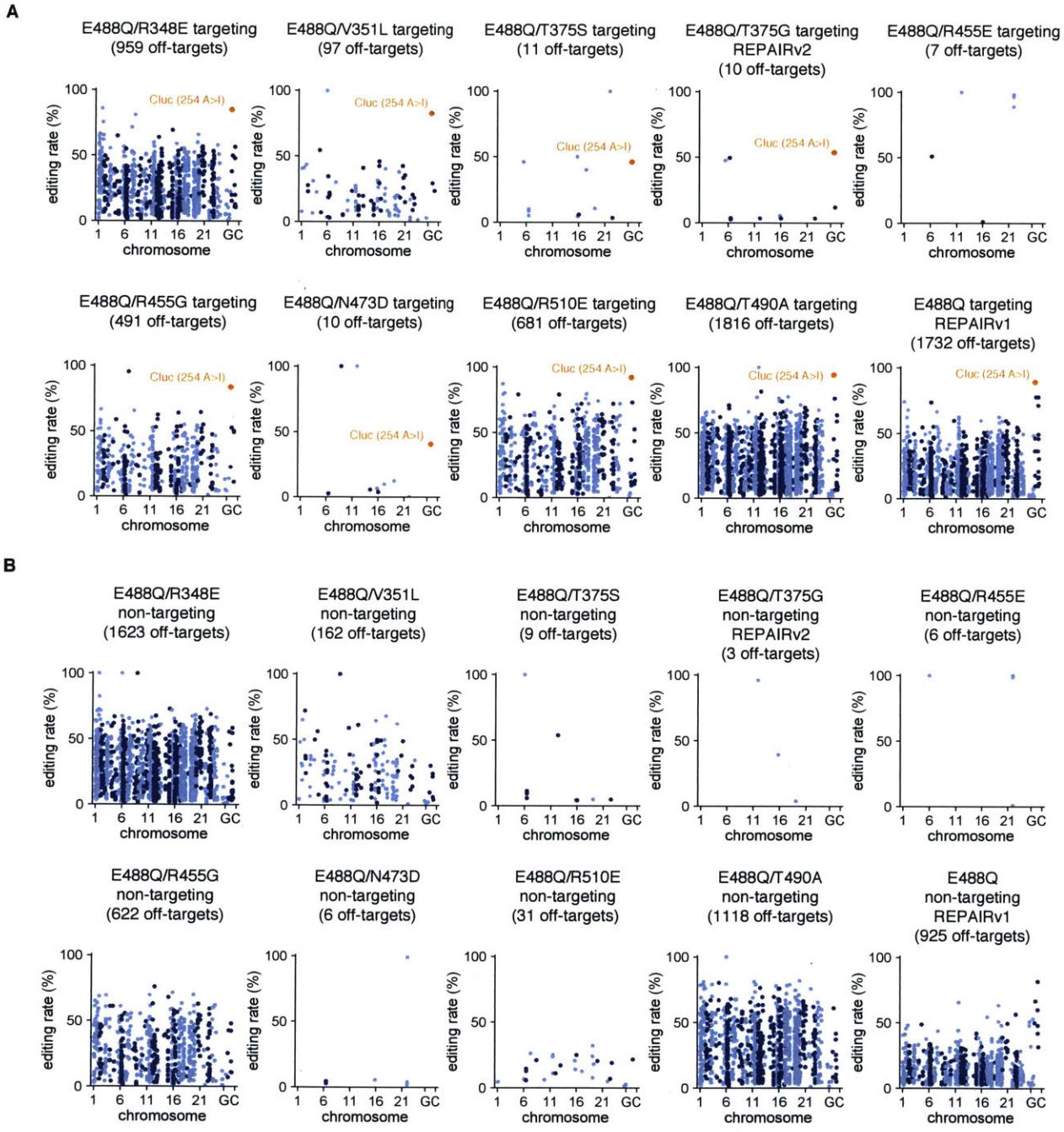

Figure S11: Transcriptome-wide specificity of RNA editing by dCas13b-ADAR2$_{DD}$(E488Q).

A) Transcriptome-wide sites of significant RNA editing by dCas13b-ADAR2$_{DD}$(E488Q) mutants with a guide targeting *Cluc*. The on-target *Cluc* site (254 A>I) is

highlighted in orange.

B) Transcriptome-wide sites of significant RNA editing by dCas13b-ADAR2$_{DD}$(E488Q) mutants with a non-targeting guide.

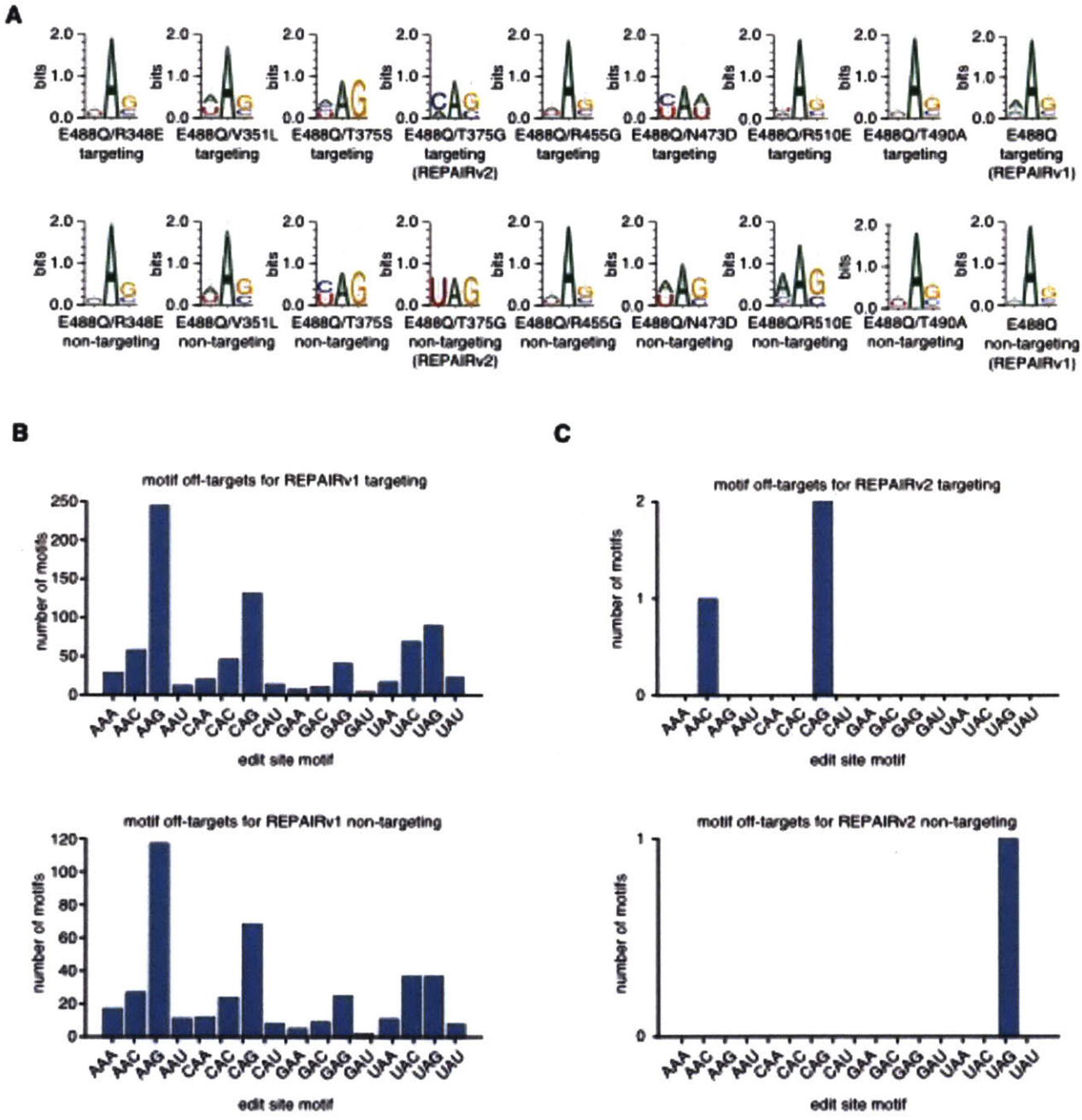

Figure S12: Characterization of motif biases in the off-targets of dCas13b-ADAR2$_{DD}$(E488Q) editing.

A) For each dCas13b-ADAR2$_{DD}$(E488Q) mutant, the motif present across all A>I off-target edits in the transcriptome is shown.

B) The distribution of off-target A>I edits per motif identity is shown for REPAIRv1 with targeting and non-targeting guide.

C) The distribution of off-target A>I edits per motif identity is shown for REPAIRv2 with targeting and non-targeting guide.

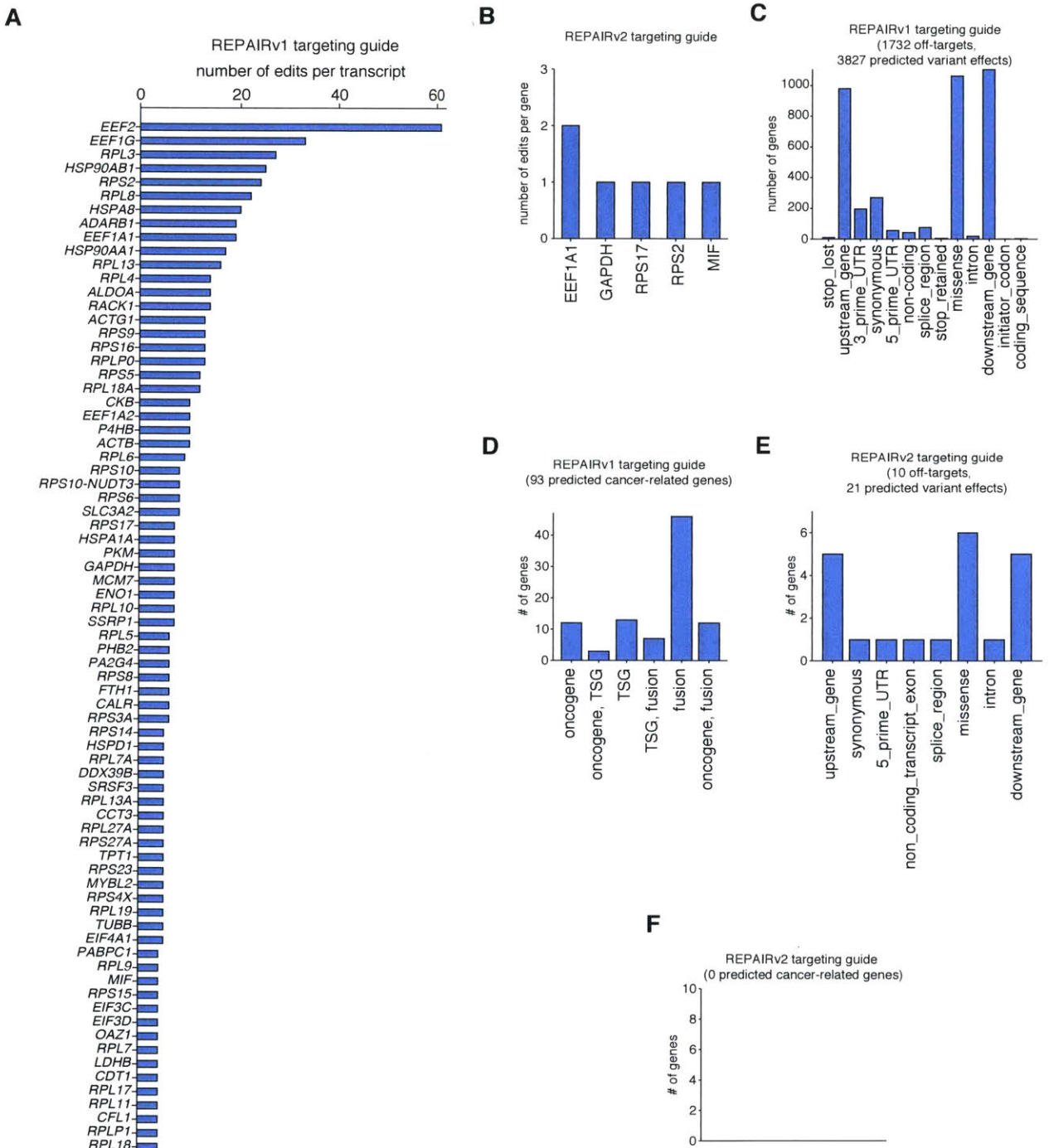
A
REPAIRv1 targeting guide
number of edits per transcript
0
20
40
60
EEF2
EEF1G
RPL3
HSP90AB1
RPS2
RPL8
HSPA8
ADARB1
EEF1A1
HSP90AA1
RPL13
RPL4
ALDOA
RACK1
ACTG1
RPS9
RPS16
RPLP0
RPS5
RPL18A
CKB
EEF1A2
P4HB
ACTB
RPL6
RPS10
RPS10-NUDT3
RPS6
SLC3A2
RPS17
HSPA1A
PKM
GAPDH
MCM7
ENO1
RPL10
SSRP1
RPL5
PHB2
PA2G4
RPS8
FTH1
CALR
RPS3A
RPS14
HSPD1
RPL7A
DDX39B
SRSF3
RPL13A
CCT3
RPL27A
RPS27A
TPT1
RPS23
MYBL2
RPS4X
RPL19
TUBB
EIF4A1
PABPC1
RPL9
MIF
RPS15
EIF3C
EIF3D
OAZ1
RPL7
LDHB
CDT1
RPL17
RPL11
CFL1
RPLP1
RPL18
RPSA
SRRM2
RPL17-C18orf32
RPS19
VIM
TUFM
B
REPAIRv2 targeting guide
number of edits per gene
0
1
2
3
EEF1A1
GAPDH
RPS17
RPS2
MIF
C
REPAIRv1 targeting guide
(1732 off-targets,
3827 predicted variant effects)
number of genes
0
200
400
600
800
1000
stop_lost
upstream_gene
3_prime_UTR
synonymous
5_prime_UTR
non-coding
splice_region
stop_retained
missense
intron
downstream_gene
initiator_codon
coding_sequence
D
REPAIRv1 targeting guide
(93 predicted cancer-related genes)
# of genes
0
10
20
30
40
oncogene
oncogene, TSG
TSG
TSG, fusion
fusion
oncogene, fusion
E
REPAIRv2 targeting guide
(10 off-targets,
21 predicted variant effects)
# of genes
0
2
4
6
upstream_gene
synonymous
5_prime_UTR
non_coding_transcript_exon
splice_region
missense
intron
downstream_gene
F
REPAIRv2 targeting guide
(0 predicted cancer-related genes)
# of genes
0
2
4
6
8
10

Figure S13: Further characterization of REPAIRv1 and REPAIRv2 off-targets.

A) Histogram of the number of off-targets per transcript for REPAIRv1.

B) Histogram of the number of off-targets per transcript for REPAIRv2.

C) Variant effect prediction of REPAIRv1 off-targets.

D) Distribution of REPAIRv1 off targets in cancer-related genes. TSG, tumor suppressor gene.

E) Variant effect prediction of REPAIRv2 off-targets.

F) Distribution of REPAIRv2 off targets in cancer-related genes.

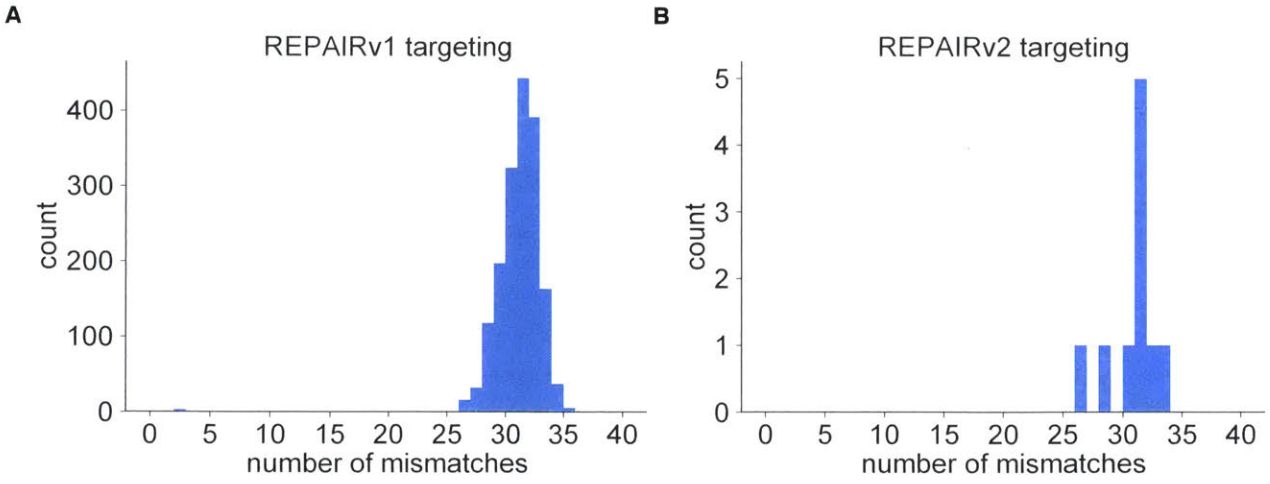

Figure S14: Evaluation of off-target sequence similarity to the guide sequence.

A) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv1 with a Cluc targeting guide.

B) Distribution of the number of mismatches (hamming distance) between the targeting guide sequence and the off-target editing sites for REPAIRv2 with a Cluc targeting guide.

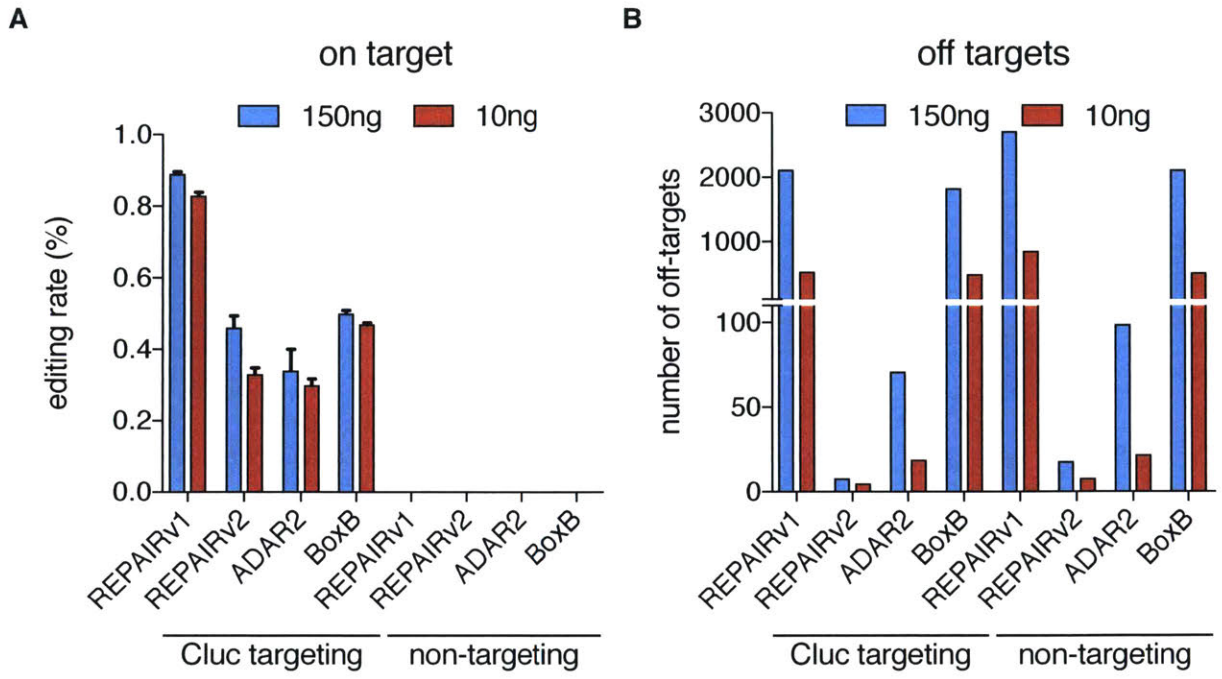

Figure S15: Comparison of REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting at two different doses of vector (150ng and 10ng effector).

A) Quantification of RNA editing activity at the *Cluc* W85X (254 A>I) on-target editing site by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting or non-targeting guide. Values shown are the mean of the three replicates.

B) Quantification of RNA editing off-targets by REPAIRv1, REPAIRv2, ADAR2 RNA targeting, and BoxB RNA targeting approaches. Each of the four methods were tested with a targeting guide for the *Cluc* W85X (254 A>I) site or non-targeting guide. For REPAIR constructs, non-targeting guide is the same as in Fig. 2C.

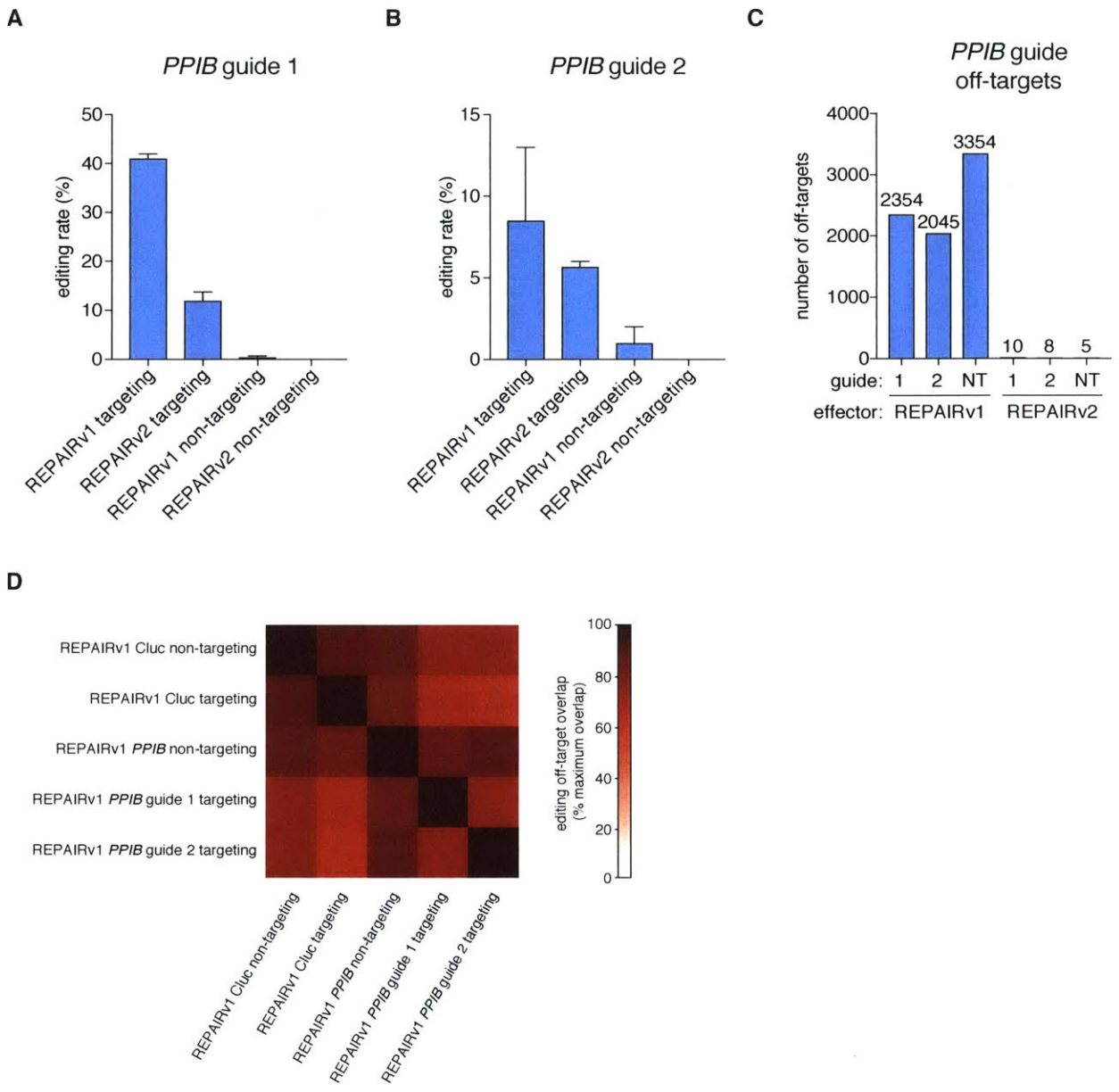

Figure S16: RNA editing efficiency and genome-wide specificity of REPAIRv1 and REPAIRv2.

A) Quantification of RNA editing activity at the *PPIB* guide 1 on-target editing site by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M.

B) Quantification of RNA editing activity at the *PPIB* guide 2 on-target editing site

by REPAIRv1, REPAIRv2 with targeting and non-targeting guides. Values represent mean +/– S.E.M.

C) Quantification of RNA editing off-targets by REPAIRv1 or REPAIRv2 with *PPIB* guide 1, *PPIB* guide 2, or non-targeting guide.

D) Overlap of off-targets between REPAIRv1 for *PPIB* targeting, Cluc targeting, and non-targeting guides. The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

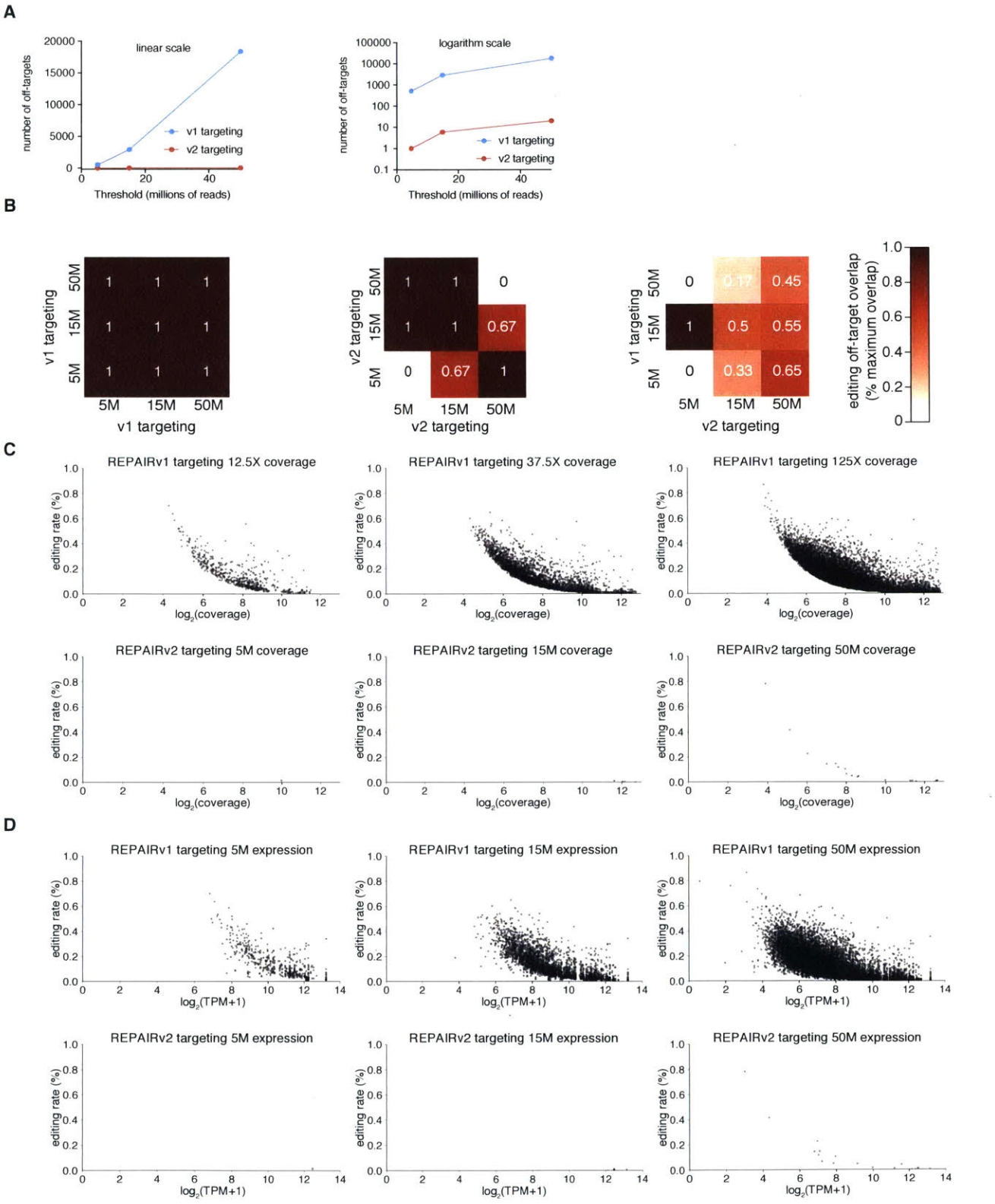

Figure S17: High coverage sequencing of REPAIRv1 and REPAIRv2 off-targets.

A) Quantitation of off-target edits for REPAIRv1 and REPAIRv2 as a function of

read depth with a total of 5 million reads (12.5x coverage), 15 million reads (37.5x coverage) and 50 million reads (125x coverage) per condition.

B) Overlap of off-target sites at different read depths of the following conditions: REPAIRv1 versus REPAIRv1 (left), REPAIRv2 versus REPAIRv2 (middle), and REPAIRv1 versus REPAIRv2 (right). The values plotted are the percent of the maximum possible intersection of the two off-target data sets.

C) Editing rate of off-target sites compared to the coverage (log2(number of reads)) of the off-target for REPAIRv1 and REPAIRv2 targeting conditions at different read depths.

D) Editing rate of off-target sites compared to the log2(TPM+1) of the off-target gene expression for REPAIRv1 and REPAIRv2 targeting conditions at different read depths.

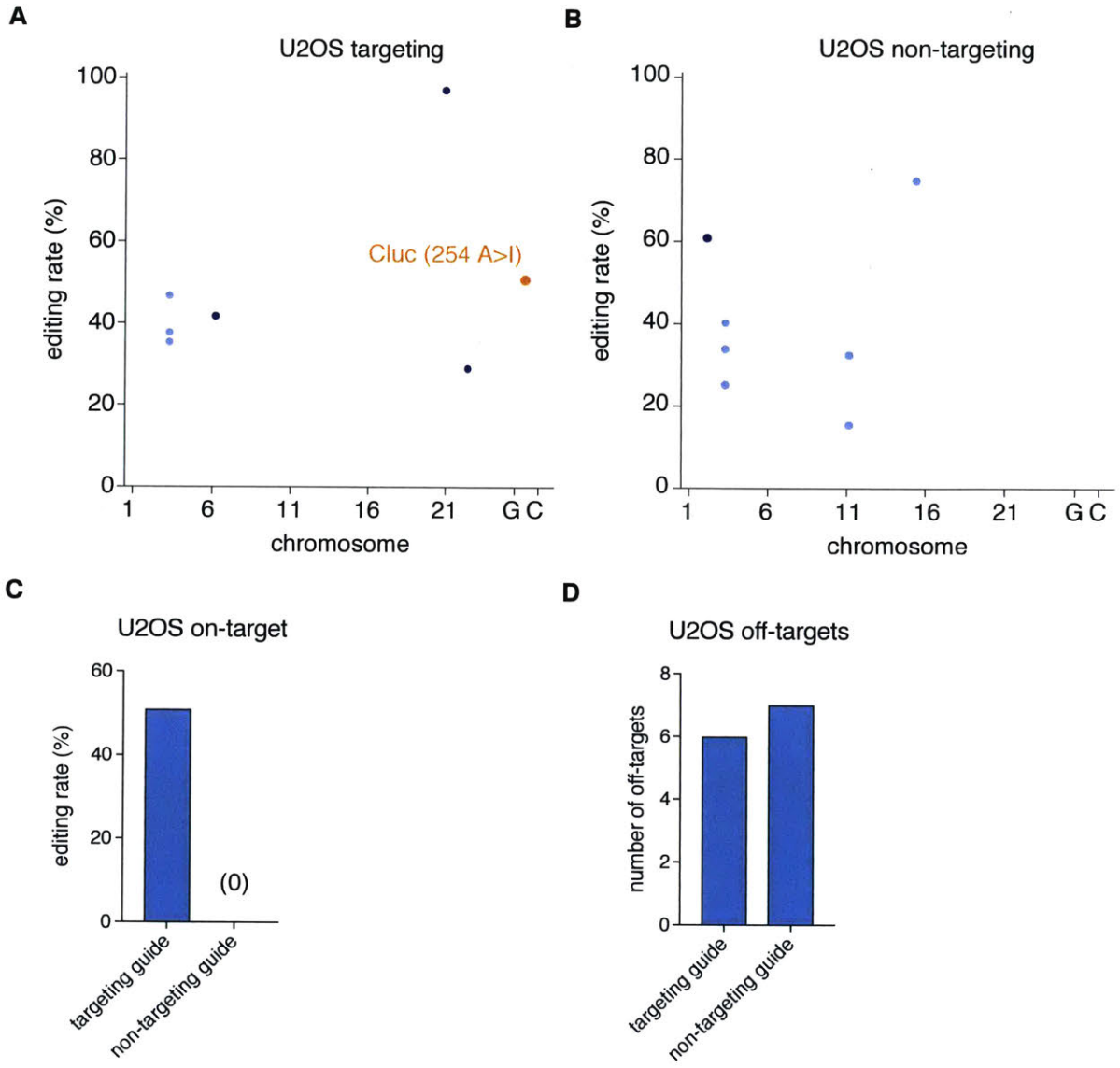

Figure S18: Quantification of REPAIRv2 activity and off-targets in the U2OS cell line.

A) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a guide targeting *Cluc* in the U2OS cell line. The on-target *Cluc* site (254 A>I) is highlighted in orange.

B) Transcriptome-wide sites of significant RNA editing by REPAIRv2 with a non-targeting guide in the U2OS cell line.

C) The on-target editing rate at the *Cluc* W85X (254 A>I) by REPAIRv2 with a

targeting guide or non-targeting guide in the U2OS cell line.

D) Quantification of off-targets by REPAIRv2 with a guide targeting *Cluc* or non-targeting guide in the U2OS cell line.

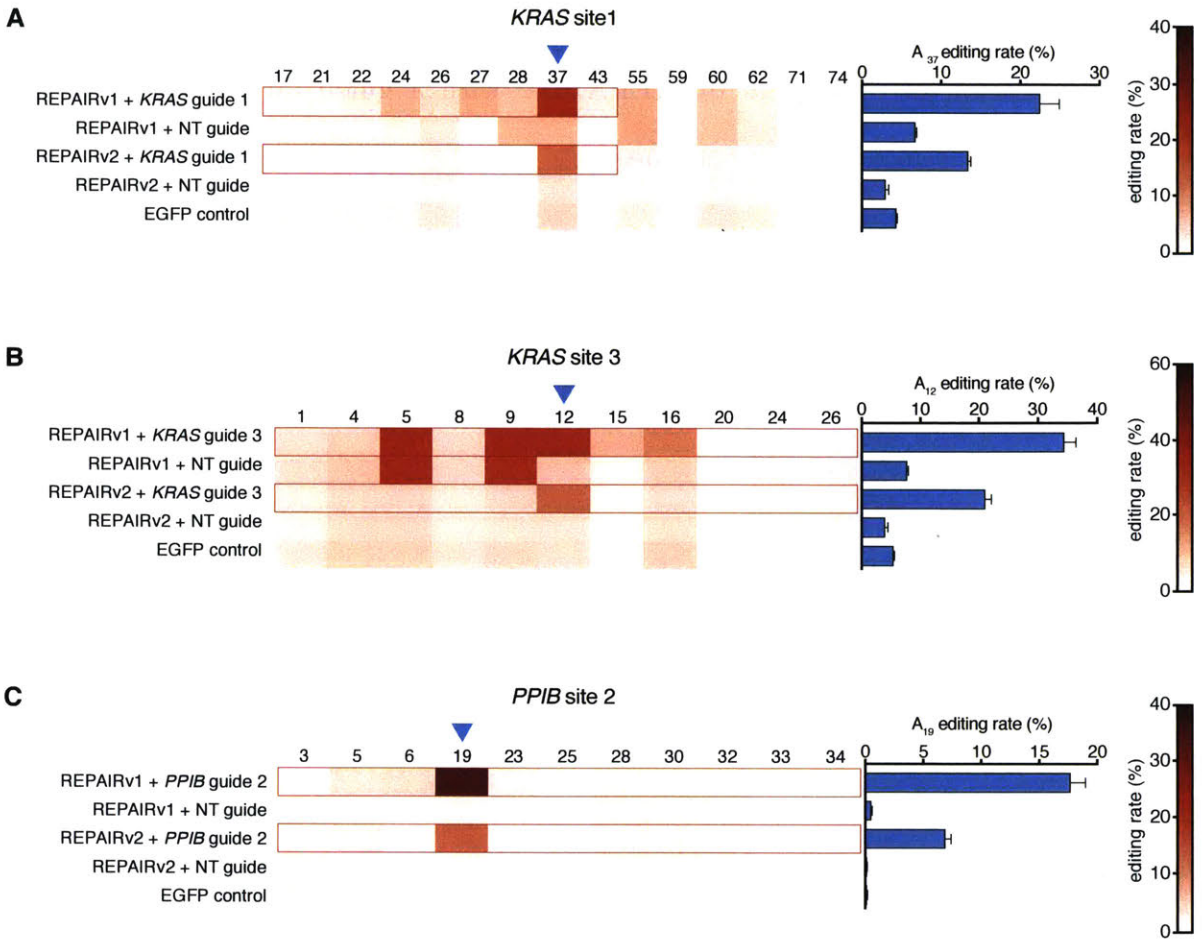

Figure S19: RNA editing efficiency and specificity of REPAIRv1 and REPAIRv2.

A) Quantification of percent editing of *KRAS* with *KRAS*-targeting guide 1 at the targeted adenosine (blue triangle) and neighboring sites for REPAIRv1 and REPAIRv2. For each guide, the region of duplex RNA is outlined in red. Values represent mean +/– S.E.M. Non-targeting guide is the same as in Fig. 2C.

B) Quantification of percent editing of *KRAS* with *KRAS*-targeting guide 3 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-targeting guide is the same as in Fig. 2C.

C) Quantification of percent editing of *PPIB* with *PPIB*-targeting guide 2 at the targeted adenosine and neighboring sites for REPAIRv1 and REPAIRv2. Non-

targeting guide is the same as in Fig. 2C.

Discussion

We show here that the RNA-guided RNA-targeting type VI-B CRISPR effector Cas13b is capable of highly efficient and specific RNA knockdown, providing the basis for improved tools for interrogating essential genes and non-coding RNA as well as controlling cellular processes at the transcript level. Catalytically inactive Cas13b (dCas13b) retains programmable RNA binding capability, which we leveraged here by fusing dCas13b to the adenosine deaminase domain of ADAR2 to achieve precise A-to-I edits, a system we term REPAIRv1 (RNA Editing for Programmable A-to-I Replacement version 1). Further engineering of the system produced REPAIRv2, which has dramatically higher specificity than previously described RNA editing platforms (*25, 29*) while maintaining high levels of on-target efficacy.

Although Cas13b exhibits high fidelity, our initial results with dCas13b-ADAR2$_{DD}$(E488Q) fusions revealed a substantial number of off-target RNA editing events. To address this, we employed a rational mutagenesis strategy to vary the ADAR2$_{DD}$ residues that contact the RNA duplex, identifying a variant, ADAR2$_{DD}$(E488Q/T375G), capable of precise, efficient, and highly specific editing when fused to dCas13b. Editing efficiency with this variant was comparable to or better than that achieved with two currently available systems, BoxB-ADAR2$_{DD}$(E488Q) or ADAR2 editing. Moreover, the REPAIRv2 system created only 20 observable off-targets in the whole transcriptome, at least an order of magnitude better than both alternative editing technologies. While it is possible that ADAR could deaminate adenosine bases on the DNA strand in RNA-DNA heteroduplexes (*20*), it is unlikely to do so in this case as Cas13b does not bind DNA efficiently and that REPAIR is cytoplasmically localized. Additionally, the lack of homology of off-target sites to the guide sequence and the strong overlap of off-targets

with the $ADAR_{DD}$(E488Q)-only condition suggest that off-targets are not mediated by off-target guide binding. Deeper sequencing and novel inosine enrichment methods could further refine our understanding of REPAIR specificity in the future.

The REPAIR system offers many advantages compared to other nucleic-acid editing tools. First, the exact target site can be encoded in the guide by placing a cytidine within the guide extension across from the desired adenosine to create a favorable A-C mismatch ideal for ADAR editing activity. Second, Cas13 has no targeting sequence constraints, such as a PFS or PAM, and no motif preference surrounding the target adenosine, allowing any adenosine in the transcriptome to be potentially targeted with the REPAIR system. The lack of motif for ADAR editing, in contrast with previous literature, is likely due to the increased local concentration of REPAIR at the target site due to dCas13b binding. We do note that DNA base editors can target either the sense or anti-sense strand, while the REPAIR system is limited to transcribed sequences, thereby constraining the total number of possible editing sites. However, due to the less constrained nature of targeting with REPAIR, this system can effect more edits within ClinVar (Fig. 4C) than Cas9-DNA base editors. Third, the REPAIR system directly deaminates target adenosines to inosines and does not rely on endogenous repair pathways, such as base-excision or mismatch repair, to generate desired editing outcomes. Therefore, REPAIR should be able to mediate efficient RNA editing even in post-mitotic cells such as neurons. Fourth, in contrast to DNA editing, RNA editing is transient and can be more easily reversed, allowing the potential for temporal control over editing outcomes. The temporary nature of REPAIR-mediated edits will likely be useful for treating diseases caused by temporary changes in cell state, such as local inflammation and could also be used to treat disease by modifying the function of proteins involved in disease-related signal transduction. For instance, REPAIR editing would allow the re-

coding of some serine, threonine and tyrosine residues that are the targets of kinases (fig. S20). Phosphorylation of these residues in disease-relevant proteins affects disease progression for many disorders including Alzheimer's disease and multiple neurodegenerative conditions (*30*). REPAIR might also be used to transiently or even chronically change the sequence of expressed, risk-modifying G-to-A variants to decrease the chance of entering a disease state for patients. For instance, REPAIR could be used to functionally mimic A-to-G alleles of *IFIH1* that protect against autoimmune disorders such as type I diabetes, immunoglobulin A deficiency, psoriasis, and systemic lupus erythematosus (*31, 32*).

The REPAIR system provides multiple opportunities for additional engineering. Cas13b possesses pre-crRNA processing activity (*13*), allowing for multiplex editing of multiple variants, any one of which alone may not affect disease, but together might have additive effects and disease-modifying potential. Extension of our rational design approach, such as combining promising mutations and directed evolution, could further increase the specificity and efficiency of the system, while unbiased screening approaches could identify additional residues for improving REPAIR activity and specificity.

Currently, the base conversions achievable by REPAIR are limited to generating inosine from adenosine; additional fusions of dCas13 with other catalytic RNA editing domains, such as APOBEC, could enable cytidine-to-uridine editing. Additionally, mutagenesis of ADAR could relax the substrate preference to target cytidine, allowing for the enhanced specificity conferred by the duplexed RNA substrate requirement to be exploited by C to U editors. Adenosine to inosine editing on DNA substrates may also be possible with catalytically inactive DNA-targeting CRISPR effectors, such as dCas9 or dCpf1, either through formation of DNA-RNA heteroduplex targets (*20*) or mutagenesis of the ADAR domain.

We have demonstrated the use of the PspCas13b enzyme as both an RNA knockdown and RNA editing tool. The dCas13b platform for programmable RNA binding has many applications, including live transcript imaging, splicing modification, targeted localization of transcripts, pull down of RNA-binding proteins, and epitranscriptomic modifications. Here, we used dCas13 to create REPAIR, adding to the existing suite of nucleic-acid editing technologies. REPAIR provides a new approach for treating genetic disease or mimicking protective alleles, and establishes RNA editing as a useful tool for modifying genetic function.

Author Contribtions:

D.B.T.C., O.O.A., J.S.G and F.Z. concieved of the study. D.B.T.C., O.O.A. and J.S.G. designed and performed all experiments, with assistance from the other authors. O.O.A. and J.S.G. designed computational pipelines for RNA editing analysis and performed statistical analyses. D.B.T.C., O.O.A., J.S.G and F.Z wrote the manuscript with input from the other authors.

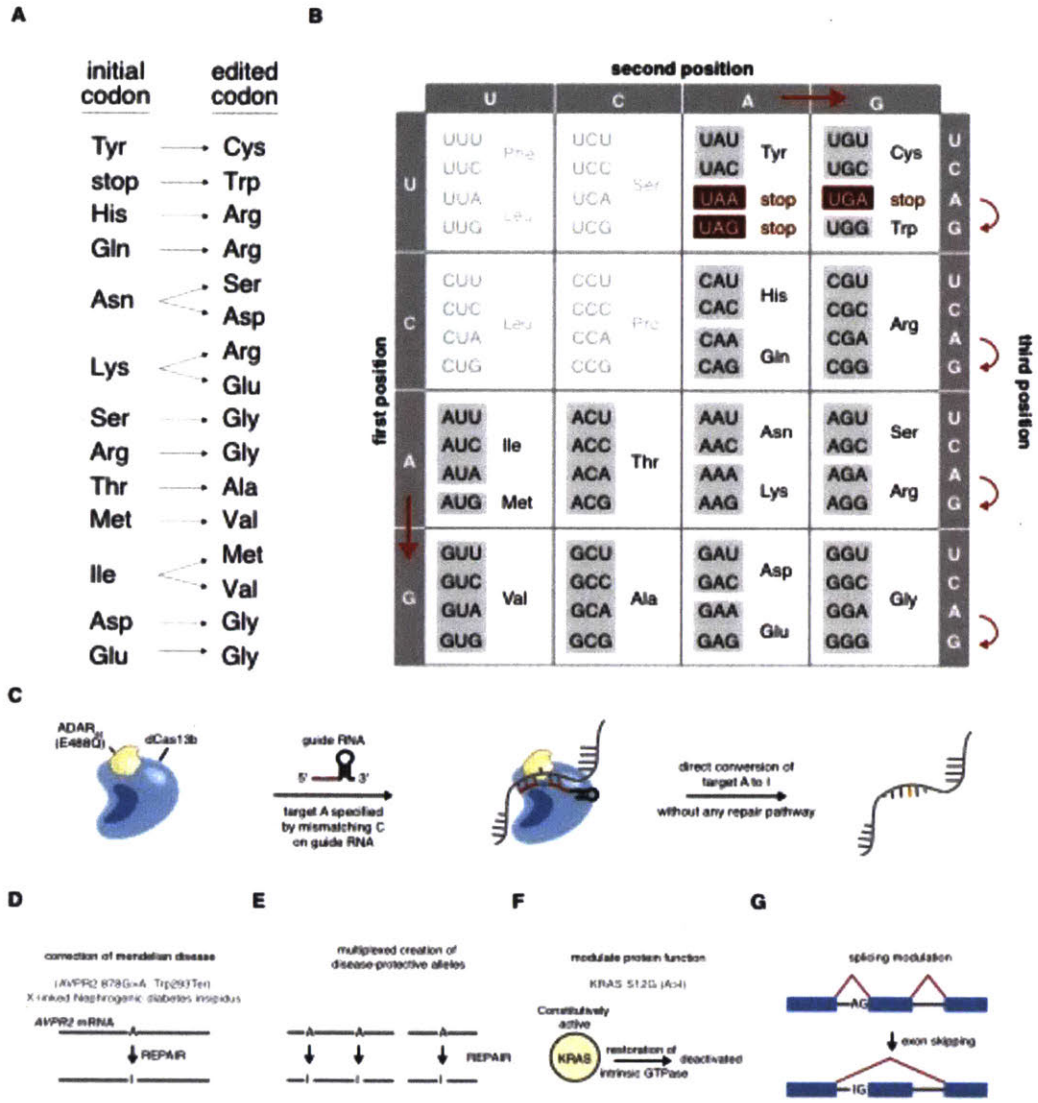

Figure S20: Demonstration of all potential codon changes with an A>I RNA editor.

A) Table of all potential codon transitions enabled by A>I editing.

B) A codon table demonstrating all the potential codon transitions enabled by A>I editing. Adapted and modified based on (*38*).

C) Model of REPAIR A-to-I editing of a precisely encoded nucleotide via a mismatch in the guide sequence. The A-to-I transition is mediated by the catalytic activity of

the ADAR2 deaminase domain and will be read as a guanosine by translational machinery. The base change does not rely on endogenous repair machinery and is permanent for as long as the RNA molecule exists in the cell.

D) REPAIR can be used for correction of Mendelian disease mutations.

E) REPAIR can be used for multiplexed A-to-I editing of multiple variants for engineering pathways or modifying disease. Multiplexed guide delivery can be achieved by delivering a single CRISPR array expression cassette since the Cas13b enzyme processes its own array.

F) REPAIR can be used for modifying protein function through amino acid changes that affect enzyme domains, such as kinases.

G) REPAIR can modulate splicing of transcripts by modifying the splice acceptor site.

SUPPLEMENTARY TEXT

As the knowledge of the protospacer flanking site (PFS) may be necessary for effective Cas13 targeting, we first sought to define (PFS) preferences for the recently described Cas13b family of RNases (*12, 13*). We heterologously expressed 15 Cas13b orthologs in *E. coli* and measured interference activity using an ampicillin-resistance assay (fig. S1A, B). Sequencing of colonies revealed a mixture of PFS preferences, with Cas13b orthologs having either solely 5' PFS preferences or a dual 5' and 3' PFS (fig. S1C).

SUPPLEMENTARY METHODS

Design and cloning of bacterial constructs

Mammalian codon optimized Cas13b constructs were cloned into the chloramphenicol resistant pACYC184 vector under control of the Lac promoter. Two corresponding direct-repeat (DR) sequences separated by BsaI restriction sites were then inserted downstream of Cas13b, under control of the pJ23119 promoter. Last, oligos for targeting spacers were phosphorylated using T4 PNK (New England Biolabs), annealed and ligated into BsaI digested vectors using T7 ligase (Enzymatics) to generate targeting Cas13b vectors. Guide sequences used are in Supplementary Table 6.

Bacterial PFS screens

Ampicillin resistance plasmids for PFS screens were cloned by inserting PCR products containing Cas13b targets with two 5' randomized nucleotides and four 3' randomized nucleotides separated by a target site immediately downstream of the start codon of the

ampicillin resistance gene *bla* using NEB Gibson Assembly (New England Biolabs). 100 ng of ampicillin-resistant target plasmids were then electroporated with 65-100 ng chloramphenicol-resistant Cas13b bacterial targeting plasmids into Endura Electrocompetent Cells (Lucigen). Plasmids were added to cells, incubated for 15 minutes on ice, electroporated using the manufacturer's recommended settings, and then 950 uL of recovery media was added to cells before a one-hour outgrowth at 37° C. The outgrowth was plated onto chloramphenicol and ampicillin double selection plates. Serial dilutions of the outgrowth were used to estimate the cfu/ng DNA. 16 hours post plating, cells were scraped off plates and surviving plasmid DNA was harvested using the Qiagen Plasmid Plus Maxi Kit (Qiagen). Surviving Cas13b target sequences and their flanking regions were amplified by PCR and sequenced using an Illumina NextSeq. To assess PFS preferences, the positions containing randomized nucleotides in the original library were extracted, and sequences depleted relative to the vector only condition and that were present in both bioreplicates were extracted using custom python scripts. The $-\log_2$ of the ratio of PFS abundance in the Cas13b condition compared to the vector only control was then used to calculate preferred motifs. Specifically, all sequences having $-\log_2$(sample/vector) depletion ratios above a specific threshold were used to generate weblogos of sequence motifs (weblogo.berkeley.edu). The specific depletion ratio values used to generate weblogos for each Cas13b ortholog are listed in Supplementary table 2.

Design and cloning of mammalian constructs for RNA interference

To generate vectors for testing Cas13 orthologs in mammalian cells, mammalian codon optimized Cas13a, Cas13b, and Cas13c genes were PCR amplified and golden-gate cloned into a mammalian expression vector containing dual NLS sequences and a C-terminal msfGFP, under control of the EF1alpha promoter. For further optimization Cas13

orthologs were golden-gate cloned into destination vectors containing different C-terminal localization tags under control of the EF1alpha promoter.

The dual luciferase reporter was cloned by PCR amplifying *Gaussia* and *Cypridinia* luciferase coding DNA, the EF1alpha and CMV promoters and assembled using the NEB Gibson Assembly (New England Biolabs).

For expression of mammalian guide RNAs for Cas13a, Cas13b, or Cas13c orthologs, the corresponding direct repeat sequences were synthesized with golden-gate acceptor sites and cloned under U6 expression via restriction digest cloning. Individual guides were then cloned into the corresponding expression backbones for each ortholog by golden-gate cloning. All Cas13 plasmids are listed in Supplementary Table 5. All Cas13 guide sequences for knockdown experiments are listed in Supplementary Tables 6-8.

Measurement of Cas13 expression in mammalian cells

Dual-NLS Cas13-msfGFP constructs were transfected into HEK293FT cells with targeting and non-targeting guides. GFP fluorescence was measured 48 hours post transfection in the non-targeting guide condition using a plate reader.

Cloning of pooled mismatch libraries for Cas13 interference specificity

Pooled mismatch library target sites were created by PCR using a forward primer containing the semi-degenerate target sequences and a constant reverse primer off of a *Gluc* template. The semi-degenerate forward oligo had at each position of the Cas13

target, plus the 5' and 3' three flanking bases, a nucleotide mixture containing 94% of the correct base and 2% of each incorrect base. The mismatch library amplicon was then cloned into the dual luciferase reporter in place of wild-type *Gluc* using NEB Gibson assembly (New England Biolabs).

Design and cloning of mammalian constructs for RNA editing

PspCas13b was made catalytically inactive (dPspCas13b) via two histidine to alanine mutations (H133A/H1058A) at the catalytic site of the HEPN domains. The deaminase domains of human ADAR1 and ADAR2 were synthesized and PCR amplified for Gibson cloning into pcDNA-CMV vector backbones and were fused to dPspCas13b at the C-terminus via GS or GSGGGGS linkers. For the experiment in which we tested different linkers we cloned the following additional linkers between dPspCas13b and $ADAR2_{DD}$: GGGGSGGGGSGGGGS, EAAAK, GGSGGSGGSGGSGGSGGS, and SGSETPGTSESATPES (XTEN). Specificity mutants were generated by Gibson cloning the appropriate mutants into the dPspCas13b-GSGGGGS backbone.

The luciferase reporter vector for measuring RNA editing activity was generated by creating a W85X mutation (TGG>TAG) in the luciferase reporter plasmid used for knockdown experiments. This reporter vector expresses functional *Gluc* as a normalization control, but a defective *Cluc* due to the addition of the W85X pretermination site. To test ADAR editing motif preferences, we cloned every possible motif around the adenosine at codon 85 (XAX) of *Cluc*. All plasmids are listed in Supplementary Table 5.

Testing PFS preferences for dCas13b

For testing PFS preference of REPAIR, we cloned a pooled plasmid library containing a 6 basepair degenerate PFS sequence upstream of a target region and adenosine editing site. The library was synthesized as an ultramer from Integrated DNA Technologies (IDT) and was made double stranded via annealing a primer and using the Klenow fragment of DNA polymerase I (New England Biolabs) to fill in the sequence. This dsDNA fragment containing the degenerate sequence was then Gibson cloned into the digested reporter vector and this was then isopropanol precipitated and purified. The cloned library was then electroporated into Endura competent *E. coli* cells (Lucigen) and plated on 245mm x 245mm square bioassay plates (Nunc). After 16 hours, colonies were harvested and midiprepped using endotoxin-free MACHEREY-NAGEL midiprep kits. Cloned libraries were verified by next-generation sequencing.

Cloning pathogenic G>A mutations for assaying REPAIR activity

For cloning disease-relevant mutations for testing REPAIR activity, 34 G>A mutations related to disease pathogenesis as defined in ClinVar were selected and 200-bp regions surrounding these mutations were golden-gate cloned between mScarlett and EGFP under a CMV promoter. Two additional G>A patient mutations in *AVPR2* and *FANCC* and their cDNA sequences were synthesized and Gibson cloned under expression of EF1alpha.

Guide cloning for REPAIR

For expression of mammalian guide RNAs for REPAIR, the PspCas13b direct repeat sequences were synthesized with golden-gate acceptor sites and cloned under U6 expression via restriction digest cloning. Individual guides were then cloned into this

expression backbone by golden-gate cloning. Guide sequences for REPAIR experiments are listed in Supplementary Table 9.

Mammalian cell culture

Mammalian cell culture experiments were performed in the HEK293FT line (American Type Culture Collection (ATCC)), which was grown in Dulbecco's Modified Eagle Medium with high glucose, sodium pyruvate, and GlutaMAX (Thermo Fisher Scientific), additionally supplemented with 1× penicillin–streptomycin (Thermo Fisher Scientific) and 10% fetal bovine serum (VWR Seradigm). Cells were maintained at confluency below 80%. The U2OS specificity experiment was performed using the U2OS cell line from ATCC and cells were cultured in ATCC-formulated McCoy's 5a Medium Modified.

Unless otherwise noted, all transfections were performed with Lipofectamine 2000 (Thermo Fisher Scientific) in 96-well plates coated with poly-D-lysine (BD Biocoat). Cells were plated at approximately 20,000 cells/well 16 hours prior to transfection to ensure 90% confluency at the time of transfection. For each well on the plate, transfection plasmids were combined with Opti-MEM I Reduced Serum Medium (Thermo Fisher) to a total of 25 µl. Separately, 24.5 µl of Opti-MEM was combined with 0.5 µl of Lipofectamine 2000. Plasmid and Lipofectamine solutions were then combined and incubated for 5 minutes, after which they were pipetted onto cells. The U2OS transfections were performed using Lipofectamine 3000 according to the manufacturer's protocol.

Mammalian cell RNA knockdown assays

To assess RNA targeting in mammalian cells with reporter constructs, 150 ng of Cas13 construct was co-transfected with 300 ng of guide expression plasmid and 12.5 ng of the

knockdown reporter construct. 48 hours post-transfection, media containing secreted luciferase was removed from cells, diluted 1:5 in PBS, and measured for activity with BioLux Cypridinia and Biolux Gaussia luciferase assay kits (New England Biolabs) on a plate reader (Biotek Synergy Neo2) with an injection protocol. All replicates performed are biological replicates.

For targeting of endogenous genes, 150 ng of Cas13 construct was co-transfected with 300 ng of guide expression plasmid. 48 hours post-transfection, cells were lysed and RNA was harvested and reverse transcribed using a previously described(*33*) modification of the Cells-to-Ct kit (Thermo Fisher Scientific). cDNA expression was measured via qPCR using TaqMan qPCR probes for the *KRAS* transcript (Thermo Fisher Scientific), *GAPDH* control probes (Thermo Fisher Scientific), and Fast Advanced Master Mix (Thermo Fisher Scientific). qPCR reactions were read out on a LightCycler 480 Instrument II (Roche), with four 5 µl technical replicates in 384-well format.

Evaluation of RNA specificity using pooled libraries of mismatched targets

The ability of Cas13 to interfere with the mismatched target library was tested using HEK293FT cells seeded in 6-well plates. ~70% confluent cells were transfected using 2400 ng Cas13 vector, 4800 ng of guide, and 240 ng of mismatched target library. 48 hours post-transfection, cells were harvested and RNA was extracted using the QIAshredder (Qiagen) and the Qiagen RNeasy Mini Kit. 1 µg of extracted RNA was reverse transcribed using the qScript Flex cDNA synthesis kit (Quantabio) following the manufacturer's gene-specific priming protocol with a *Gluc* specific RT primer. cDNA was then amplified and sequenced on an Illumina NextSeq.

Sequencing was analyzed by counting reads per sequence and depletion scores were calculated by determining the $\log_2$(-read count ratio) value, where read count ratio is the ratio of read counts in the targeting guide condition versus the non-targeting guide condition. This score represents the level of Cas13 activity on the sequence, with higher values representing stronger depletion and thus higher Cas13 cleavage activity. Separate distributions for the single mismatch and double mismatch sequences were determined and plotted as heatmaps with a depletion score for each mismatch identity. For double mismatch sequences the average of all possible double mismatches at a given position were plotted.

Transcriptome-wide profiling of Cas13 in mammalian cells by RNA sequencing

For measurement of transcriptome-wide specificity, 150 ng of Cas13 construct, 300 ng of guide expression plasmid, and 15 ng of the knockdown reporter construct were co-transfected; for shRNA conditions, 300 ng of shRNA targeting plasmid, 15 ng of the knockdown reporter construct, and 150 ng of EF1-alpha driven mCherry (to balance reporter load) were co-transfected. 48 hours post-transfection, RNA was purified with the RNeasy Plus Mini kit (Qiagen), mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs), and prepared for sequencing with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs). RNA sequencing libraries were then sequenced on a NextSeq (Illumina).

To analyze transcriptome-wide sequencing data, reads were aligned to the RefSeq GRCh38 assembly using Bowtie and RSEM version 1.2.31 with default parameters(*34*). Transcript expression was quantified as $\log_2$(TPM + 1), genes were filtered for $\log_2$(TPM + 1) >2.5. For selection of differentially expressed genes, only genes with differential

changes of >2 or <.75 were considered. Statistical significance of differential expression was evaluated using a Student's t-test on three targeting replicates versus non-targeting replicates, and filtered for a false discovery rate of <0.01% by the Benjamini-Hochberg procedure.

REPAIR editing in mammalian cells

To assess REPAIR activity in mammalian cells, we transfected 150 ng of REPAIR vector, 300 ng of guide expression plasmid, and 40 ng of the RNA editing reporter. After 48 hours, RNA from cells was harvested and reverse transcribed using a method previously described(*33*) with a gene specific reverse transcription primer. The extracted cDNA was then subjected to two rounds of PCR to add Illumina adaptors and sample barcodes using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs). The library was then subjected to next generation sequencing on an Illumina NextSeq or MiSeq. RNA editing rates were then evaluated at all adenosines within the sequencing window.

In experiments where the luciferase reporter was targeted for RNA editing, we also harvested the media with secreted luciferase prior to RNA harvest. In this case, because corrected *Cluc* might be at low levels, we did not dilute the media. We measured luciferase activity with BioLux Cypridinia and Biolux Gaussia luciferase assay kits (New England Biolabs) on a plate reader (Biotek Synergy Neo2) with an injection protocol. All replicates performed are biological replicates.

PFS binding mammalian screen

To determine the contribution of the PFS to editing efficiency in mammalian cells, 625 ng of PFS target library, 4.7 μg of guide, and 2.35 μg of REPAIR were co-transfected in HEK293FT cells plated in 25 cm^2 flasks. Plasmids were mixed with 33 μl of PLUS reagent (Thermo Fisher Scientific), brought to 533 μl with Opti-MEM, incubated for 5 minutes, combined with 30 μl of Lipofectamine 2000 and 500 μl of Opti-MEM, incubated for an additional 5 minutes, and then pipetted onto cells. 48 hours post-transfection, RNA was harvested with the RNeasy Plus Mini kit (Qiagen), reverse transcribed with qScript Flex (Quantabio) using a gene specific primer, and amplified with two rounds of PCR using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs) to add Illumina adaptors and sample barcodes. The library was sequenced on an Illumina NextSeq, and RNA editing rates at the target adenosine were mapped to PFS identity. To increase coverage, the PFS was computationally collapsed to 4 nucleotides adjacent to the 5' end of the target sequence. REPAIR editing rates were calculated for each PFS, averaged over biological replicates with non-targeting rates for the corresponding PFS subtracted.

Whole-transcriptome sequencing to evaluate ADAR editing specificity

For analyzing off-target RNA editing sites across the transcriptome, we harvested total RNA from cells 48 hours post-transfection using the RNeasy Plus Miniprep kit (Qiagen). The mRNA fraction was then enriched using a NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and this RNA was then prepared for sequencing using an NEBNext Ultra RNA Library Prep Kit for Illumina (NEB). The libraries were then sequenced on an Illumina NextSeq and loaded such that there were at least 5 million reads per sample.

RNA editing analysis for targeted and transcriptome-wide experiments

Analysis of the transcriptome-wide editing RNA sequencing data was performed on the FireCloud computational framework (https://software.broadinstitute.org/firecloud/) using a custom workflow we developed: https://portal.firecloud.org/#methods/m/rna_editing_final_workflow/rna_editing_final_workflow/1. For analysis, unless otherwise denoted, sequence files were randomly downsampled to 5 million reads. For the high-coverage sequencing analysis, samples were randomly downsampled to 5 million, 15 million, or 50 million reads. An index was generated using the RefSeq GRCh38 assembly with *Gluc* and *Cluc* sequences added, and reads were aligned and quantified using Bowtie/RSEM version 1.3.0. Alignment BAMs were then sorted and analyzed for RNA editing sites using REDitools (*35, 36*) with the following parameters: -t 8 -e -d -l -U [AG or TC] -p -u -m20 -T6-0 -W -v 1 -n 0.0. Any significant edits found in untransfected or EGFP-transfected conditions were considered to be SNPs or artifacts of the transfection and filtered out from the analysis of off-targets. Off-targets were considered significant if the Fisher's exact test yielded a p-value less than 0.05 after multiple hypothesis correction by Benjamini Hochberg correction and at least 2 of 3 biological replicates identified the edit site. Overlap of edits between samples was calculated relative to the maximum possible overlap, equivalent to the fewer number of edits between the two samples. The percentage of overlapping edit sites was calculated as the number of shared edit sites divided by minimum number of edits of the two samples, multiplied by 100. For the high-coverage sequencing analysis, an additional layer of filtering for known SNP positions was performed using the Kaviar (*37*) method for identifying SNPs.

For analyzing the predicted variant effects of each off-target, the list of off-target edit sites was analyzed using the variant annotation integrator (https://genome.ucsc.edu/cgi-bin/hgVai) as part of the UCSC genome browser suite of tools using the SIFT and PolyPhen-2 annotations. To predict whether the off-target genes are oncogenic, a database of oncogenic annotations from the COSMIC catalogue of somatic mutations in cancer was used to characterize off-target genes (cancer.sanger.ac.uk).

For analyzing whether the REPAIR constructs perturbed RNA levels, the transcript per million (TPM) values output from the RSEM analysis were used for expression counts and transformed to log-space by taking the $\log_2$(TPM+1). To find differentially regulated genes, a Student's t-test was performed on three targeting guide replicates versus three non-targeting guide replicates. The statistical analysis was only performed on genes with $\log_2$(TPM+1) values greater than 2.5 and genes were only considered differentially regulated if they had a fold change greater than 2 or less than 0.8. Genes were reported if they had a false discovery rate (Benjamini Hochberg correction) of less than 0.01.

References

1. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).

2. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20-36 (2017).

3. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

4. P. Mali *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

5. B. Zetsche *et al.*, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).

6. H. Kim, J. S. Kim, A guide to genome engineering with programmable nucleases. *Nat Rev Genet* **15**, 321-334 (2014).

7. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).

8. K. Nishida *et al.*, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, (2016).

9. Y. B. Kim *et al.*, Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376 (2017).

10. O. O. Abudayyeh *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).

11. S. Shmakov *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385-397 (2015).

12. S. Shmakov *et al.*, Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182 (2017).

13. A. A. Smargon *et al.*, Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell* **65**, 618-630 e617 (2017).

14. J. S. Gootenberg *et al.*, Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442 (2017).

15. O. O. Abudayyeh *et al.*, RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284 (2017).

16. K. Nishikura, Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* **79**, 321-349 (2010).

17. M. H. Tan *et al.*, Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249-254 (2017).

18. B. L. Bass, H. Weintraub, An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**, 1089-1098 (1988).

19. M. M. Matthews *et al.*, Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol* **23**, 426-433 (2016).

20. Y. Zheng, C. Lorenzo, P. A. Beal, DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. *Nucleic Acids Res* **45**, 3369-3377 (2017).

21. A. Kuttan, B. L. Bass, Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A* **109**, E3295-3304 (2012).

22. S. K. Wong, S. Sato, D. W. Lazinski, Substrate recognition by ADAR1 and ADAR2. *RNA* **7**, 846-858 (2001).

23. M. Fukuda *et al.*, Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. *Sci Rep* **7**, 41478 (2017).

24. M. F. Montiel-Gonzalez, I. Vallecillo-Viejo, G. A. Yudowski, J. J. Rosenthal, Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci U S A* **110**, 18285-18290 (2013).

25. M. F. Montiel-Gonzalez, I. C. Vallecillo-Viejo, J. J. Rosenthal, An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* **44**, e157 (2016).

26. J. Wettengel, P. Reautschnig, S. Geisler, P. J. Kahle, T. Stafforst, Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res* **45**, 2797-2808 (2017).

27. Y. Wang, J. Havel, P. A. Beal, A Phenotypic Screen for Functional Mutants of Human Adenosine Deaminase Acting on RNA 1. *ACS Chem Biol* **10**, 2512-2519 (2015).

28. K. A. Lehmann, B. L. Bass, Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875-12884 (2000).

29. T. Stafforst, M. F. Schneider, An RNA-deaminase conjugate selectively repairs point mutations. *Angew Chem Int Ed Engl* **51**, 11166-11169 (2012).

30. C. Ballatore, V. M. Lee, J. Q. Trojanowski, Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci* **8**, 663-672 (2007).

31. Y. Li *et al.*, Carriers of rare missense variants in IFIH1 are protected from psoriasis. *J Invest Dermatol* **130**, 2768-2772 (2010).

32. R. C. Ferreira *et al.*, Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. *Nat Genet* **42**, 777-780 (2010).

33. J. Joung *et al.*, Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat Protoc* **12**, 828-863 (2017).

34. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

35. E. Picardi, A. M. D'Erchia, A. Montalvo, G. Pesole, Using REDItools to Detect RNA Editing Events in NGS Datasets. *Curr Protoc Bioinformatics* **49**, 12 12 11-15 (2015).

36. E. Picardi, G. Pesole, REDItools: high-throughput RNA editing detection made easy. *Bioinformatics* **29**, 1813-1814 (2013).

37. G. Glusman, J. Caballero, D. E. Mauldin, L. Hood, J. C. Roach, Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* **27**, 3216-3217 (2011).

38. J. D. Watson, *Molecular biology of the gene.* (Pearson, Boston, ed. Seventh edition, 2014), pp. xxxiv, 872 pages.

List of Supplementary Tables:

Supplementary Table 1: Cas13 Orthologs used in this study.

Supplementary Table 2: PFS cutoffs in bacterial screens

Supplementary Table 3: dCas13b-ADAR linker sequences used in this study for RNA editing in mammalian cells.

Supplementary Table 4: Disease information for disease-relevant mutations

Supplementary Table 5: Key plasmids used in this study

Supplementary Table 6: Guide/shRNA sequences used in this study for knockdown in mammalian cells

Supplementary Table 7: Guide sequences used for *Gluc* knockdown

Supplementary Table 8: Guide sequences used for *Cluc* knockdown

Supplementary Table 9: Guide sequences used in this study for RNA editing in mammalian cells.

Supplementary Table 1: Cas13 Orthologs used in this study.

Cas13 ID	Cas13 abbreviation	Host Organism	Protein Accession
Cas13a1	LshCas13a	*Leptotrichia shahii*	WP_018451595.1
Cas13a2	LwaCas13a	*Leptotrichia wadei* (Lw2)	WP_021746774.1
Cas13a3	LseCas13a	*Listeria seeligeri*	WP_012985477.1
Cas13a4	LbmCas13a	*Lachnospiraceae* bacterium MA2020	WP_044921188.1
Cas13a5	LbnCas13a	*Lachnospiraceae* bacterium NK4A179	WP_022785443.1
Cas13a6	CamCas13a	[*Clostridium*] *aminophilum* DSM 10710	WP_031473346.1
Cas13a7	CgaCas13a	*Carnobacterium gallinarum* DSM 4847	WP_034560163.1
Cas13a8	Cga2Cas13a	*Carnobacterium gallinarum* DSM 4847	WP_034563842.1
Cas13a9	Pprcas13a	*Paludibacter propionicigenes* WB4	WP_013443710.1
Cas13a10	LweCas13a	*Listeria weihenstephanensis* FSL R9-0317	WP_036059185.1
Cas13a11	LbfCas13a	*Listeriaceae* bacterium FSL M6-0635	WP_036091002.1

Cas13a12	Lwa2cas13a	*Leptotrichia wadei* F0279	WP_021746774.1
Cas13a13	RcsCas13a	*Rhodobacter capsulatus* SB 1003	WP_013067728.1
Cas13a14	RcrCas13a	*Rhodobacter capsulatus* R121	WP_023911507.1
Cas13a15	RcdCas13a	*Rhodobacter capsulatus* DE442	WP_023911507.1
Cas13a16	LbuCas13a	*Leptotrichia buccalis* C-1013-b	WP_015770004.1
Cas13a17	HheCas13a	*Herbinix hemicellulosilytica*	CRZ35554.1
Cas13a18	EreCas13a	[*Eubacterium*] *rectale*	WP_055061018.1
Cas13a19	EbaCas13a	*Eubacteriaceae* bacterium CHKCI004	WP_090127496.1
Cas13a20	BmaCas13a	*Blautia* sp. Marseille-P2398	WP_062808098.1
Cas13a21	LspCas13a	*Leptotrichia* sp. oral taxon 879 str. F0557	WP_021744063.1
Cas13b1	BzoCas13b	*Bergeyella zoohelcum*	WP_002664492
Cas13b2	PinCas13b	*Prevotella intermedia*	WP_036860899
Cas13b3	PbuCas13b	*Prevotella buccae*	WP_004343973
Cas13b4	AspCas13b	*Alistipes* sp. ZOR0009	WP_047447901
Cas13b5	PsmCas13b	*Prevotella* sp. MA2016	WP_036929175
Cas13b6	RanCas13b	*Riemerella anatipestifer*	WP_004919755
Cas13b7	PauCas13b	*Prevotella aurantiaca*	WP_025000926
Cas13b8	PsaCas13b	*Prevotella saccharolytica*	WP_051522484
Cas13b9	Pin2Cas13b	*Prevotella intermedia*	WP_061868553
Cas13b10	CcaCas13b	*Capnocytophaga canimorsus*	WP_013997271
Cas13b11	PguCas13b	*Porphyromonas gulae*	WP_039434803
Cas13b12	PspCas13b	*Prevotella* sp. P5-125	WP_044065294
Cas13b13	FbrCas13b	*Flavobacterium branchiophilum*	WP_014084666
Cas13b14	PgiCas13b	*Porphyromonas gingivalis*	WP_053444417
Cas13b15	Pin3Cas13b	*Prevotella intermedia*	WP_050955369
Cas13c1	FnsCas13c	*Fusobacterium necrophorum* subsp. funduliforme ATCC 51357 contig00003	WP_005959231.1
Cas13c2	FndCas13c	F*usobacterium necrophorum* DJ-2 contig0065, whole genome shotgun sequence	WP_035906563.1
Cas13c3	FnbCas13c	*Fusobacterium necrophorum* BFTR-1 contig0068	WP_035935671.1
Cas13c4	FnfCas13c	*Fusobacterium necrophorum* subsp. funduliforme 1_1_36S cont1.14	EHO19081.1
Cas13c5	FpeCas13c	*Fusobacterium perfoetens* ATCC 29250 T364DRAFT_scaffold00009.9_C	WP_027128616.1

Cas13c6	FulCas13c	*Fusobacterium ulcerans* ATCC 49185 cont2.38	WP_040490876.1
Cas13c7	AspCas13c	*Anaerosalibacter* sp. ND1 genome assembly Anaerosalibacter massiliensis ND1	WP_042678931.1

Supplementary Table 2: PFS cutoffs in bacterial screens

Cas13b ortholog	Key	-Log_2 depletion score used to generate PFS motif
Bergeyella zoohelcum	1	2
Prevotella intermedia locus 1	2	1
Prevotella buccae	3	3
Alistipes sp. ZOR0009	4	1
Prevotella sp. MA2016	5	2
Riemerella anatipestifer	6	4
Prevotella aurantiaca	7	1
Prevotella saccharolytica	8	0
Prevotella intermedia locus 2	9	0
Capnocytophaga canimorsus	10	3
Porphyromonas gulae	11	4
Prevotella sp. P5-125	12	2.1
Flavobacterium branchiophilum	13	1
Porphyromonas gingivalis	14	3
Prevotella intermedia locus 2	15	4

Supplementary Table 3: dCas13b-ADAR linker sequences used in this study for RNA editing in mammalian cells.

Figure	linker
2C	GSGGGGS
2E	GS
S3B	GSGGGGS
S3C	GS
S3D	GS
S3E: GS	GS
S3E: GSGGGGS	GSGGGGS
S3E: (GGGS)3	GGGGSGGGGSGGGGS
S3E: Rigid	EAAAK
S3E: (GGS)6	GGSGGSGGSGGSGGSGGS

S3E: XTEN	SGSETPGTSESATPES
3B	GS
S3F	GS
3C	GS
4B	GS
4D	GS
4E	GS
S5A: Δ984-1090, Δ1026-1090, Δ1053-1090	GS
S5A: Δ1-125, Δ1-88, Δ1-72	GSGGGGS
5B	GS
5C	GS
5D	GS
S6A	GS
S6C	GS
S6D	GS
S7D	GS
6A	GS
S8A	GS
6B	GS
S8B	GS
S8C	GS
S9A	GS
S9B	GS
6C	GS
6D	GS
6E	GS
6F	GS
S13A	GS
S13B	GS

Supplementary Table 4: Disease information for disease-relevant mutations

Full length candidates	**Gene**	**Disease**
NM_000054.4(AVPR2):c.878G>A (p.Trp293Ter)	*AVPR2*	Nephrogenic diabetes insipidus, X-linked
NM_000136.2(FANCC):c.1517G>A (p.Trp506Ter)	*FANCC*	Fanconi anemia, complementation group C
Additional simulated candidates		
Candidate	**Gene**	**Disease**
NM_000206.2(IL2RG):c.710G>A	*IL2RG*	X-linked severe combined immunodeficiency

(p.Trp237Ter)		
NM_000132.3(F8):c.3144G>A (p.Trp1048Ter)	*F8*	Hereditary factor VIII deficiency disease
NM_000527.4(LDLR):c.1449G>A (p.Trp483Ter)	*LDLR*	Familial hypercholesterolemia
NM_000071.2(CBS):c.162G>A (p.Trp54Ter)	*CBS*	Homocystinuria due to CBS deficiency
NM_000518.4(HBB):c.114G>A (p.Trp38Ter)	*HBB*	betaThalassemia
NM_000035.3(ALDOB):c.888G>A (p.Trp296Ter)	*ALDOB*	Hereditary fructosuria
NM_004006.2(DMD):c.3747G>A (p.Trp1249Ter)	*DMD*	Duchenne muscular dystrophy
NM_005359.5(SMAD4):c.906G>A (p.Trp302Ter)	*SMAD4*	Juvenile polyposis syndrome
NM_000059.3(BRCA2):c.582G>A (p.Trp194Ter)	*BRCA2*	Familial cancer of breast|Breast-ovarian cancer, familial 2
NM_000833.4(GRIN2A):c.3813G>A (p.Trp1271Ter)	*GRIN2A*	Epilepsy, focal, with speech disorder and with or without mental retardation
NM_002977.3(SCN9A):c.2691G>A (p.Trp897Ter)	*SCN9A*	Indifference to pain, congenital, autosomal recessive
NM_007375.3(TARDBP):c.943G>A (p.Ala315Thr)	*TARDBP*	Amyotrophic lateral sclerosis type 10
NM_000492.3(CFTR):c.3846G>A (p.Trp1282Ter)	*CFTR*	Cystic fibrosis|Hereditary pancreatitis|not provided|ataluren response - Efficacy
NM_130838.1(UBE3A):c.2304G>A (p.Trp768Ter)	*UBE3A*	Angelman syndrome
NM_000543.4(SMPD1):c.168G>A (p.Trp56Ter)	*SMPD1*	Niemann-Pick disease, type A
NM_206933.2(USH2A):c.9390G>A (p.Trp3130Ter)	*USH2A*	Usher syndrome, type 2A
NM_130799.2(MEN1):c.1269G>A (p.Trp423Ter)	*MEN1*	Hereditary cancer-predisposing syndrome
NM_177965.3(C8orf37):c.555G>A (p.Trp185Ter)	*C8orf37*	Retinitis pigmentosa 64
NM_000249.3(MLH1):c.1998G>A (p.Trp666Ter)	*MLH1*	Lynch syndrome
NM_000548.4(TSC2):c.2108G>A (p.Trp703Ter)	*TSC2*	Tuberous sclerosis 2|Tuberous sclerosis syndrome
NM_000267.3(NF1):c.7044G>A (p.Trp2348Ter)	*NF1*	Neurofibromatosis, type 1
NM_000179.2(MSH6):c.3020G>A (p.Trp1007Ter)	*MSH6*	Lynch syndrome
NM_000344.3(SMN1):c.305G>A (p.Trp102Ter)	*SMN1*	Spinal muscular atrophy, type II|Kugelberg-Welander disease

NM_024577.3(SH3TC2):c.920G>A (p.Trp307Ter)	*SH3TC2*	Charcot-Marie-Tooth disease, type 4C
NM_001369.2(DNAH5):c.8465G>A (p.Trp2822Ter)	*DNAH5*	Primary ciliary dyskinesia
NM_004992.3(MECP2):c.311G>A (p.Trp104Ter)	*MECP2*	Rett syndrome
NM_032119.3(ADGRV1):c.7406G>A (p.Trp2469Ter)	*ADGRV1*	Usher syndrome, type 2C
NM_017651.4(AHI1):c.2174G>A (p.Trp725Ter)	*AHI1*	Joubert syndrome 3
NM_004562.2(PRKN):c.1358G>A (p.Trp453Ter)	*PRKN*	Parkinson disease 2
NM_000090.3(COL3A1):c.3833G>A (p.Trp1278Ter)	*COL3A1*	Ehlers-Danlos syndrome, type 4
NM_007294.3(BRCA1):c.5511G>A (p.Trp1837Ter)	*BRCA1*	Familial cancer of breast|Breast-ovarian cancer, familial 1
NM_000256.3(MYBPC3):c.3293G>A (p.Trp1098Ter)	*MYBPC3*	Primary familial hypertrophic cardiomyopathy
NM_000038.5(APC):c.1262G>A (p.Trp421Ter)	*APC*	Familial adenomatous polyposis 1
NM_001204.6(BMPR2):c.893G>A (p.W298*)	*BMPR2*	Primary pulmonary hypertension

Supplementary Table 5: Key plasmids used in this study

Plasmid	Description	Benchling link
pC0037	CMV-Cluciferase-polyA EF1a-G-luciferase-polyA	https://benchling.com/s/seq-GMa3RAbt0JkjT8kX9aRa
pC0038	CMV-Cluciferase(W85X)-polyA EF1a-G-luciferase-polyA	https://benchling.com/s/seq-W2n4wX4vSUuslGzYgYO5
pC0039	CMV-dCas13b12-GS-ADAR2DD(E488Q)	https://benchling.com/s/seq-arzpsupZEzGu3ghBDhtv
pC0040	LwaCas13a crRNA backbone	https://benchling.com/s/seq-0SqKieU2CWyd3RRawuKp
pC0041	RanCas13b crRNA backbone	https://benchling.com/s/seq-yKHvxw5C84w9inEx3XaU
pC0042	PguCas13b crRNA backbone	https://benchling.com/s/seq-ZLKtRrNkhNw0BOzcgdW5
pC0043	PspCas13b crRNA backbone	https://benchling.com/s/seq-OH6nMmZCZn930BWqcFNa
pC0044	EF1a-BsiWI-Cas13b6-NES-mapk	https://benchling.com/s/seq-hxOBIW6sDZE1o4DMz6lZ
pC0045	EF1a-BsiWI-Cas13b11-NES-HIV	https://benchling.com/s/seq-GYuyzloHGlD8CNO4TCSy
pC0046	EF1a-BsiWI-Cas13b12-NES-HIV	https://benchling.com/s/seq-g62SIhluOlRdD8aArJaC
pC0047	CMV-dCas13b12-ADAR1DD(E1008Q)	https://benchling.com/s/seq-R3zRpb4whgEiZBoTvpgM
pC0048	CMV-dCas13b12-longlinker-ADAR2DD(E488Q)	https://benchling.com/s/seq-Y92Xyc1WxOlZDLMNv8K8

pC0049	EF1a-BsiWI-Cas13-B12-NES-HIV, H133A/H1058A	https://benchling.com/s/seq-lK5ZoHDkOCTPV0SwG7VD
pC0050	CMV-dCas13b12-longlinker-ADAR2DD(wt)	https://benchling.com/s/seq-YuFM6m06znFKA9txLrrw
pC0051	W85X REPAIR targeting guide	https://benchling.com/s/seq-pJjKdbYG6YdpAMKAyXE0
pC0052	REPAIR non-targeting guide	https://benchling.com/s/seq-U9gHnOW41C1DVUBGQypw
pC0053	CMV-dCas13b12-GS-ADAR2DD(E488Q)-delta-984-1090	https://benchling.com/s/seq-HASFia3255bkdC9iUtxu
pC0054	T375G specificity mutant	https://benchling.com/s/seq-lWXqpjFVHeqkLlHVFZ4t
pC0055	T375G Cas13b C-term delta 984-1090	https://benchling.com/s/seq-1KNBN52nxWXZgwekbbiO

Supplementary Table 6: Guide/shRNA sequences used in this study for knockdown in mammalian cells

Name	**Spacer sequence**	**Interference Mechanism**	**Notes**	**First figure**
Bacterial PFS guide	GCCAGCUUUCCGGGCAUUGGCUUCCAUC	Cas13b	Used for all orthologs	
Cas13a-Gluc guide 1	GCCAGCUUUCCGGGCAUUGGCUUCCAUC	Cas13a	Used for all Cas13a orthologs	Figure 1B
Cas13a-Gluc guide 2	ACCCAGGAAUCUCAGGAAUGUCGACGAU	Cas13a	Used for all Cas13a orthologs	Figure 1B
Cas13a-non-targeting guide (LacZ)	AGGGUUUUCCCAGUCACGACGUUGUAAA	Cas13a	Used for all Cas13a orthologs	Figure 1B
Cas13b-Gluc guide 1.1	GGGCAUUGGCUUCCAUCUCUUUGAGCACCU	Cas13b	Used for orthologs 1-3, 6, 7, 10, 11, 12, 14, 15	Figure 1B
Cas13b-Gluc guide 1.2	GUGCAGCCAGCUUUCCGGGCAUUGGCUUCC	Cas13b	Used for ortholog 4	Figure 1B
Cas13b-Gluc guide 1.3	GCAGCCAGCUUUCCGGGCAUUGGCUUCCAU	Cas13b	Used for ortholog 5	Figure 1B
Cas13b-Gluc guide 1.4	GGCUUCCAUCUCUUUGAGCACCUCCAGCGG	Cas13b	Used for ortholog 8, 9	Figure 1B
Cas13b-Gluc guide 1.5	GGAAUGUCGACGAUCGCCUCGCCUAUGCCG	Cas13b	Used for ortholog 13	Figure 1B

Cas13b-Gluc guide 2.1	GAAUGUCGACGAUCGCCUCG CCUAUGCCGC	Cas13b	Used for orthologs 1-3, 6, 7, 10, 11, 14, 15	Figure 1B
Cas13b-Gluc guide 2.2	GACCUGUGCGAUGAACUGCU CCAUGGGCUC	Cas13b	Used for ortholog 12	Figure 1B
Cas13b-Gluc guide 2.2	GUGUGGCAGCGUCCUGGGAU GAACUUCUUC	Cas13b	Used for ortholog 4	Figure 1B
Cas13b-Gluc guide 2.3	GUGGCAGCGUCCUGGGAUGA ACUUCUUCAU	Cas13b	Used for ortholog 5	Figure 1B
Cas13b-Gluc guide 2.4	GCUUCUUGCCGGGCAACUUC CCGCGGUCAG	Cas13b	Used for ortholog 8, 9	Figure 1B
Cas13b-Gluc guide 2.6	GCAGGGUUUUCCCAGUCACG ACGUUGUAAAA	Cas13b	Used for ortholog 13	Figure 1B
Cas13b-non targeting guide	GCAGGGUUUUCCCAGUCACG ACGUUGUAAAA	Cas13b	Used for all orthologs	Figure 1B
Cas13a-Gluc guide-RNASeq	ACCCAGGAAUCUCAGGAAUG UCGACGAU	Cas13a		Figure 1E
shRNA-Gluc guide	CAGCUUUCCGGGCAUUGGCU U	shRNA		Figure 1F
Cas13b-Gluc guide-RNASeq	CCGCUGGAGGUGCUCAAAGA GAUGGAAGCC	Cas13b		Figure 1F
Cas13a-Gluc-guide-1	GCCAGCUUUCCGGGCAUUGG CUUCCAUC	Cas13a		Figure S2A
Cas13a-Gluc-guide-2	ACCCAGGAAUCUCAGGAAUG UCGACGAU	Cas13a		Figure S2A
Cas13b-Gluc-opt-guide-1	GGGCAUUGGCUUCCAUCUCU UUGAGCACCU	Cas13b		Figure S2A
Cas13b-Gluc-opt-guide-2	GAAUGUCGACGAUCGCCUCG CCUAUGCCGC	Cas13b		Figure S2A
Cas13a KRAS guide 1	CAAGGCACUCUUGCCUACGC CACCAGCU	Cas13a		Figure S2B
Cas13a KRAS guide 2	UCAUAUUCGUCCACAAAAUG AUUCUGAA	Cas13a		Figure S2B
Cas13a KRAS guide 3	AUUAUUUAUGGCAAAUACAC AAAGAAAG	Cas13a		Figure S2B
Cas13a KRAS guide	GAAUAUCUUCAAAUGAUUUA GUAUUAUU	Cas13a		Figure S2B

4				
Cas13a KRAS guide 5	ACCAUAGGUACAUCUUCAGA GUCCUUAA	Cas13a		Figure S2B
Cas13b KRAS guide 1	GUCAAGGCACUCUUGCCUAC GCCACCAGCU	Cas13b		Figure S2B
Cas13b KRAS guide 2	GAUCAUAUUCGUCCACAAAA UGAUUCUGAA	Cas13b		Figure S2B
Cas13b KRAS guide 3	GUAUUAUUUAUGGCAAAUAC ACAAAGAAAG	Cas13b		Figure S2B
Cas13b KRAS guide 4	GUGAAUAUCUUCAAAUGAUU UAGUAUUAUU	Cas13b		Figure S2B
Cas13b KRAS guide 5	GGACCAUAGGUACAUCUUCA GAGUCCUUAA	Cas13b		Figure S2B
shRNA KRAS guide 1	aagagugccuugacgauacagcCUCGA Ggcuguaucgucaaggcacucuu	shRNA		Figure S2B
shRNA KRAS guide 2	aaucauuuuguggacgaauauCUCGA Gauauucguccacaaaaugauu	shRNA		Figure S2B
shRNA KRAS guide 3	aaauaauacuaaaucauuugaCUCGAG ucaaaugauuuaguauuauuu	shRNA		Figure S2B
shRNA KRAS guide 4	aauaauacuaaaucauuugaaCUCGAG uucaaaugauuuaguauuauu	shRNA		Figure S2B
shRNA KRAS non-targeting guide	aaggacucugaagauguaccuCUCGAG agguacaucuucagaguccuu	shRNA		Figure S2B

Supplementary Table 7: Guide sequences used for *Gluc* knockdown

Name	Spacer sequence	Position	Notes	First figure
Gluc tiling guide 1	GAGAUCAGGGCAAACAGAACUUUGACUCCC	2	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 2	GGAUGCAGAUCAGGGCAAACAGAACUUUGA	7	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 3	GCACAGCGAUGCAGAUCAGGGCAAACAGAA	13	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 4	GCUCGGCCACAGCGAUGCAGAUCAGGGCAA	19	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 5	GGGGCUUGGCCUCGGCCACAGCGAUGCAGA	28	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 6	GUGGGCUUGGCCUCGGCCACAGCGAUGCAG	29	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 7	GUCUCGGUGGGCUUGGCCUCGGCCACAGCG	35	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 8	GUUCGUUGUUCUCGGUGGGCUUGGCCUCGG	43	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 9	GGAAGUCUUCGUUGUUCUCGGUGGGCUUGG	49	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 10	GAUGUUGAAGUCUUCGUUGUUCUCGGUGGG	54	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 11	GCGGCCACGAUGUUGAAGUCUUCGUUGUUC	62	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 12	GUGGCCACGGCCACGAUGUUGAAGUCUUCG	68	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 13	GGUUGCUGGCCACGGCCACGAUGUUGAAGU	73	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 14	GUCGCGAAGUUGCUGGCCACGGCCACGAUG	80	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 15	GCCGUGGUCGCGAAGUUGCUGGCCACGGCC	86	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 16	GCGAGAUCCGUGGUCGCGAAGUUGCUGGCC	92	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 17	GCAGCAUCGAGAUCCGUGGUCGCGAAGUUG	98	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 18	GGGUCAGCAUCGAGAUCCGUGGUCGCGAAG	101	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 19	GCUUCCCGCGGUCAGCAUCGAGAUCCGUGG	109	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 20	GGGGCAACUUCCCGCGGUCAGCAUCGAGAU	115	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 21	GUCUUGCCGGGCAACUUCCCGCGGUCAGCA	122	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 22	GGCAGCUUCUUGCCGGGCAACUUCCCGCGG	128	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 23	GCCAGCGGCAGCUUCUUGCCGGGCAACUUC	134	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 24	GCACCUCCAGCGGCAGCUUCUUGCCGGGCA	139	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 25	GCUUUGAGCACCUCCAGCGGCAGCUUCUUG	146	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 26	GCAUCUCUUUGAGCACCUCCAGCGGCAGCU	151	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 27	GUCCAUCUCUUUGAGCACCUCCAGCGGCAG	153	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 28	GGGCAUUGGCUUCCAUCUCUUUGAGCACCU	163	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 29	GUCCGGGCAUUGGCUUCCAUCUCUUUGAGC	167	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 30	GGCCAGCUUUCCGGGCAUUGGCUUCCAUCU	175	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 31	GGGUGCAGCCAGCUUUCCGGGCAUUGGCUU	181	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 32	GAGCCCCUGGUGCAGCCAGCUUUCCGGGCA	188	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 33	GAUCAGACAGCCCCUGGUGCAGCCAGCUUU	195	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 34	GGCAGAUCAGACAGCCCCUGGUGCAGCCAG	199	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 35	GACAGGCAGAUCAGACAGCCCCUGGUGCAG	203	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 36	GUGAUGUGGGACAGGCAGAUCAGACAGCCC	212	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 37	GACUUGAUGUGGGACAGGCAGAUCAGACAG	215	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 38	GGGGCGUGCACUUGAUGUGGGACAGGCAGA	223	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 39	GCUUCAUCUUGGGCGUGCACUUGAUGUGGG	232	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 40	GUGAACUUCUUCAUCUUGGGCGUGCACUUG	239	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 41	GGGAUGAACUUCUUCAUCUUGGGCGUGCAC	242	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 42	GUGGGAUGAACUUCUUCAUCUUGGGCGUGC	244	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 43	GGGCAGCGUCCUGGGAUGAACUUCUUCAUC	254	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 44	GGGUGUGGCAGCGUCCUGGGAUGAACUUCU	259	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 45	GUUCGUAGGUGUGGCAGCGUCCUGGGAUGA	265	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 46	GCGCCUUCGUAGGUGUGGCAGCGUCCUGGG	269	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 47	GUCUUUGUCGCCUUCGUAGGUGUGGCAGCG	276	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 48	GCUUUGUCGCCUUCGUAGGUGUGGCAGCGU	275	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 49	GUGCCGCCCUGUGCGGACUCUUUGUCGCCU	293	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 50	GUAUGCCGCCCUGUGCGGACUCUUUGUCGC	295	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 51	GCCUCGCCUAUGCCGCCCUGUGCGGACUCU	302	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 52	GGAUCGCCUCGCCUAUGCCGCCCUGUGCGG	307	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 53	GAUGUCGACGAUCGCCUCGCCUAUGCCGCC	315	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 54	GCAGGAAUGUCGACGAUCGCCUCGCCUAUG	320	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 55	GAAUCUCAGGAAUGUCGACGAUCGCCUCGC	325	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 56	GCCCAGGAAUCUCAGGAAUGUCGACGAUCG	331	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 57	GCCUUGAACCCAGGAAUCUCAGGAAUGUCG	338	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 58	GCCAAGUCCUUGAACCCAGGAAUCUCAGGA	344	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 59	GUGGGCUCCAAGUCCUUGAACCCAGGAAUC	350	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 60	GCCAUGGGCUCCAAGUCCUUGAACCCAGGA	353	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 61	GGAACUGCUCCAUGGGCUCCAAGUCCUUGA	361	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 62	GUGCGAUGAACUGCUCCAUGGGCUCCAAGU	367	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 63	GGACCUGUGCGAUGAACUGCUCCAUGGGCU	373	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 64	GACAGAUCGACCUGUGCGAUGAACUGCUCC	380	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 65	GACACACAGAUCGACCUGUGCGAUGAACUG	384	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 66	GUGCAGUCCACACACAGAUCGACCUGUGCG	392	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 67	GCCAGUUGUGCAGUCCACACACAGAUCGAC	399	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 68	GGGCAGCCAGUUGUGCAGUCCACACACAGA	404	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 69	GUUUGAGGCAGCCAGUUGUGCAGUCCACAC	409	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 70	GAAGCCCUUUGAGGCAGCCAGUUGUGCAGU	415	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 71	GCACGUUGGCAAGCCCUUUGAGGCAGCCAG	424	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 72	GACUGCACGUUGGCAAGCCCUUUGAGGCAG	428	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 73	GGGUCAGAACACUGCACGUUGGCAAGCCCU	437	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 74	GCAGGUCAGAACACUGCACGUUGGCAAGCC	439	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 75	GAGCAGGUCAGAACACUGCACGUUGGCAAG	441	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 76	GGCCACUUCUUGAGCAGGUCAGAACACUGC	452	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 77	GCGGCAGCCACUUCUUGAGCAGGUCAGAAC	457	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 78	GUGCGGCAGCCACUUCUUGAGCAGGUCAGA	459	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 79	GAGCGUUGCGGCAGCCACUUCUUGAGCAGG	464	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 80	GAAAGGUCGCACAGCGUUGCGGCAGCCACU	475	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 81	GCUGGCAAAGGUCGCACAGCGUUGCGGCAG	480	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 82	GGGCAAAGGUCGCACAGCGUUGCGGCAGCC	478	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 83	GUGGAUCUUGCUGGCAAAGGUCGCACAGCG	489	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 84	GCACCUGGCCCUGGAUCUUGCUGGCAAAGG	499	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 85	GUGGCCCUGGAUCUUGCUGGCAAAGGUCGC	495	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 86	GUGAUCUUGUCCACCUGGCCCUGGAUCUUG	509	Note that the Cas13a spacers are truncated by	1C

			two nucleotides at the 5' end	
Gluc tiling guide 87	GCCCCUUGAUCUUGUCCACCUGGCCCUGGA	514	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 88	GCCCUUGAUCUUGUCCACCUGGCCCUGGAU	513	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 89	GCCUUGAUCUUGUCCACCUGGCCCUGGAUC	512	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 90	GGCAAAGGUCGCACAGCGUUGCGGCAGCCA	477	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 91	GCAAAGGUCGCACAGCGUUGCGGCAGCCAC	476	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 92	GAAGGUCGCACAGCGUUGCGGCAGCCACUU	474	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Gluc tiling guide 93	GAGGUCGCACAGCGUUGCGGCAGCCACUUC	473	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Non-targeting guide 1	GGUAAUGCCUGGCUUGUCGACGCAUAGUCUG	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Non-targeting guide 2	GGGAACCUUGGCCGUUAUAAAGUCUGACCAG	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C
Non-targeting guide 3	GGAGGGUGAGAAUUUAGAACCAAGAUUGUUG	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1C

Supplementary Table 8: Guide sequences used for *Cluc* knockdown

Name	**Spacer sequence**	**Position**	**Notes**	**First figure**
Cluc tiling guide 1	GAGUCCUGGCAAUGAACAGUGGCGCAGUAG	32	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 2	GGGUGCCACAGCUGCUAUCAAUACAUUCUC	118	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D

Cluc tiling guide 3	GUUACAUACUGACACAUUCGGCAACAUGUU	197	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 4	GUAUGUACCAGGUUCCUGGAACUGGAAUCU	276	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 5	GCCUUGGUUCCAUCCAGGUUCUCCAGGGUG	350	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 6	GCAGUGAUGGGAUUCUCAGUAGCUUGAGCG	431	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 7	GAGCCUGGCAUCUCAACAACAGCGAUGGUG	512	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 8	GUGUCUGGGGCGAUUCUUACAGAUCUUCCU	593	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 9	GCUGGAUCUGAAGUGAAGUCUGUAUCUUCC	671	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 10	GGCAACGUCAUCAGGAUUUCCAUAGAGUGG	747	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 11	GAGGCGCAGGAGAUGGUGUAGUAGUAGAAG	830	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 13	GAGGGACCCUGGAAUUGGUAUCUUGCUUUG	986	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 14	GGUAAGAGUCAACAUUCCUGUGUGAAACCU	1066	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 15	GACCAGAAUCUGUUUUCCAUCAACAAUGAG	1143	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 16	GAUGGCUGUAGUCAGUAUGUCACCAUCUUG	1227	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 17	GUACCAUCGAAUGGAUCUCUAAUAUGUACG	1304	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D

Cluc tiling guide 18	GAGAUCACAGGCUCCUUCAGCAUCAAAAGA	1380	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 19	GCUUUGACCGGCGAAGAGACUAUUGCAGAG	1461	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 20	GCCCCUCAGGCAAUACUCGUACAUGCAUCG	1539	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Cluc tiling guide 21	GCUGGUACUUCUAGGGUGUCUCCAUGCUUU	1619	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Non-targeting guide 1	GGUAAUGCCUGGCUUGUCGACGCAUAGUCUG	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Non-targeting guide 2	GGGAACCUUGGCCGUUAUAAAGUCUGACCAG	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D
Non-targeting guide 3	GGAGGGUGAGAAUUUAGAACCAAGAUUGUUG	N/A	Note that the Cas13a spacers are truncated by two nucleotides at the 5' end	1D

Supplementary Table 9: Guide sequences used in this study for RNA editing in mammalian cells. Mismatched base flips are capitalized

Name	**Spacer sequence**	**Notes**	**First figure**
Tiling 30 nt 30 mismatch distance	gCauccugcggccucuacucugcauucaauu	Has a 5' G for U6 expression	2C
Tiling 30 nt 28 mismatch distance	gacCauccugcggccucuacucugcauucaa	Has a 5' G for U6 expression	2C
Tiling 30 nt 26 mismatch distance	gaaacCauccugcggccucuacucugcauuc	Has a 5' G for U6 expression	2C
Tiling 30 nt 24 mismatch distance	gcuaaacCauccugcggccucuacucugcau	Has a 5' G for U6 expression	2C
Tiling 30 nt 22 mismatch distance	guucuaaacCauccugcggccucuacucugc	Has a 5' G for U6 expression	2C
Tiling 30 nt 20 mismatch distance	guguucuaaacCauccugcggccucuacucu	Has a 5' G for U6 expression	2C
Tiling 30 nt 18 mismatch distance	gaauguucuaaacCauccugcggccucuacu	Has a 5' G for U6 expression	2C
Tiling 30 nt 16 mismatch distance	gagaauguucuaaacCauccugcggccucua	Has a 5' G for U6 expression	2C
Tiling 30 nt 14 mismatch distance	gauagaauguucuaaacCauccugcggccuc	Has a 5' G for U6 expression	2C
Tiling 30 nt 12 mismatch distance	gccauagaauguucuaaacCauccugcggcc	Has a 5' G for U6 expression	2C

Tiling 30 nt 10 mismatch distance	guuccauagaauguucuaaacCauccugcgg	Has a 5' G for U6 expression	2C
Tiling 30 nt 8 mismatch distance	gcuuuccauagaauguucuaaacCauccugc	Has a 5' G for U6 expression	2C
Tiling 30 nt 6 mismatch distance	gcucuuuccauagaauguucuaaacCauccu	Has a 5' G for U6 expression	2C
Tiling 30 nt 4 mismatch distance	gaucucuuuccauagaauguucuaaacCauc	Has a 5' G for U6 expression	2C
Tiling 30 nt 2 mismatch distance	ggaaucucuuuccauagaauguucuaaacCa	Has a 5' G for U6 expression	2C
Tiling 50 nt 50 mismatch distance	gCauccugcggccucuacucugcauucaauuacaua cugacacauucggca	Has a 5' G for U6 expression	2C
Tiling 50 nt 48 mismatch distance	gacCauccugcggccucuacucugcauucaauuaca uacugacacauucgg	Has a 5' G for U6 expression	2C
Tiling 50 nt 46 mismatch distance	gaaacCauccugcggccucuacucugcauucaauua cauacugacacauuc	Has a 5' G for U6 expression	2C
Tiling 50 nt 44 mismatch distance	gcuaaacCauccugcggccucuacucugcauucaau uacauacugacacau	Has a 5' G for U6 expression	2C
Tiling 50 nt 42 mismatch distance	guucuaaacCauccugcggccucuacucugcauuca auuacauacugacac	Has a 5' G for U6 expression	2C
Tiling 50 nt 40 mismatch distance	guguucuaaacCauccugcggccucuacucugcauu caauuacauacugac	Has a 5' G for U6 expression	2C
Tiling 50 nt 38 mismatch distance	gaauguucuaaacCauccugcggccucuacucugca uucaauuacauacug	Has a 5' G for U6 expression	2C
Tiling 50 nt 36 mismatch distance	gagaauguucuaaacCauccugcggccucuacucug cauucaauuacauac	Has a 5' G for U6 expression	2C
Tiling 50 nt 34 mismatch distance	gauagaauguucuaaacCauccugcggccucuacuc ugcauucaauuacau	Has a 5' G for U6 expression	2C
Tiling 50 nt 32 mismatch distance	gccauagaauguucuaaacCauccugcggccucuac ucugcauucaauuac	Has a 5' G for U6 expression	2C
Tiling 50 nt 30 mismatch distance	guuccauagaauguucuaaacCauccugcggccucu acucugcauucaauu	Has a 5' G for U6 expression	2C
Tiling 50 nt 28 mismatch distance	gcuuuccauagaauguucuaaacCauccugcggccu cuacucugcauucaa	Has a 5' G for U6 expression	2C
Tiling 50 nt 26 mismatch distance	gcucuuuccauagaauguucuaaacCauccugcggc cucuacucugcauuc	Has a 5' G for U6 expression	2C
Tiling 50 nt 24 mismatch distance	gaucucuuuccauagaauguucuaaacCauccugcg gccucuacucugcau	Has a 5' G for U6 expression	2C
Tiling 50 nt 22 mismatch distance	ggaaucucuuuccauagaauguucuaaacCauccug cggccucuacucugc	Has a 5' G for U6 expression	2C
Tiling 50 nt 20 mismatch distance	guggaaucucuuuccauagaauguucuaaacCaucc ugcggccucuacucu	Has a 5' G for U6 expression	2C
Tiling 50 nt 18 mismatch distance	gacuggaaucucuuuccauagaauguucuaaacCau ccugcggccucuacu	Has a 5' G for U6 expression	2C
Tiling 50 nt 16 mismatch distance	ggaacuggaaucucuuuccauagaauguucuaaacC auccugcggccucua	Has a 5' G for U6 expression	2C
Tiling 50 nt 14 mismatch	guggaacuggaaucucuuuccauagaauguucuaaa	Has a 5' G for U6 expression	2C

distance	cCauccugcggccuc		
Tiling 50 nt 12 mismatch distance	gccuggaacuggaaucucuuuccauagaauguucuaaacCauccugcggcc	Has a 5' G for U6 expression	2C
Tiling 50 nt 10 mismatch distance	guuccuggaacuggaaucucuuuccauagaauguucuaaacCauccugcgg	Has a 5' G for U6 expression	2C
Tiling 50 nt 8 mismatch distance	ggguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccugc	Has a 5' G for U6 expression	2C
Tiling 50 nt 6 mismatch distance	gcagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauccu	Has a 5' G for U6 expression	2C
Tiling 50 nt 4 mismatch distance	gaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCauc	Has a 5' G for U6 expression	2C
Tiling 50 nt 2 mismatch distance	gguaccagguuccuggaacuggaaucucuuuccauagaauguucuaaacCa	Has a 5' G for U6 expression	2C
Tiling 70 nt 70 mismatch distance	gCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggguuuau	Has a 5' G for U6 expression	2C
Tiling 70 nt 68 mismatch distance	gacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggguuu	Has a 5' G for U6 expression	2C
Tiling 70 nt 66 mismatch distance	gaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggu	Has a 5' G for U6 expression	2C
Tiling 70 nt 64 mismatch distance	gcuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccug	Has a 5' G for U6 expression	2C
Tiling 70 nt 62 mismatch distance	guucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuucc	Has a 5' G for U6 expression	2C
Tiling 70 nt 60 mismatch distance	guguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuu	Has a 5' G for U6 expression	2C
Tiling 70 nt 58 mismatch distance	gaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuu	Has a 5' G for U6 expression	2C
Tiling 70 nt 56 mismatch distance	gagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacaugu	Has a 5' G for U6 expression	2C
Tiling 70 nt 54 mismatch distance	gauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacau	Has a 5' G for U6 expression	2C
Tiling 70 nt 52 mismatch distance	gccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaac	Has a 5' G for U6 expression	2C
Tiling 70 nt 50 mismatch distance	guuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggca	Has a 5' G for U6 expression	2C
Tiling 70 nt 48 mismatch distance	gcuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucgg	Has a 5' G for U6 expression	2C
Tiling 70 nt 46 mismatch distance	gcucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauuc	Has a 5' G for U6 expression	2C
Tiling 70 nt 44 mismatch distance	gaucucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacau	Has a 5' G for U6 expression	2C
Tiling 70 nt 42 mismatch distance	ggaaucucuuuccauagaauguucuaaacCauccuggccucuacucugcauucaauuacauacugacac	Has a 5' G for U6 expression	2C
Tiling 70 nt 40 mismatch distance	guggaaucucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugac	Has a 5' G for U6 expression	2C

Tiling 70 nt 38 mismatch distance	gacuggaaucucuuuccauagaauguucuaaacCau ccugcggccucuacucugcauucaauuacauacug	Has a 5' G for U6 expression	2C
Tiling 70 nt 36 mismatch distance	ggaacuggaaucucuuuccauagaauguucuaaacC auccugcggccucuacucugcauucaauuacauac	Has a 5' G for U6 expression	2C
Tiling 70 nt 34 mismatch distance	guggaacuggaaucucuuuccauagaauguucuaaa cCauccugcggccucuacucugcauucaauuacau	Has a 5' G for U6 expression	2C
Tiling 70 nt 32 mismatch distance	gccuggaacuggaaucucuuuccauagaauguucua aacCauccugcggccucuacucugcauucaauuac	Has a 5' G for U6 expression	2C
Tiling 70 nt 30 mismatch distance	guuccuggaacuggaaucucuuuccauagaauguuc uaaacCauccugcggccucuacucugcauucaauu	Has a 5' G for U6 expression	2C
Tiling 70 nt 28 mismatch distance	ggguuccuggaacuggaaucucuuuccauagaaugu ucuaaacCauccugcggccucuacucugcauucaa	Has a 5' G for U6 expression	2C
Tiling 70 nt 26 mismatch distance	gcagguuccuggaacuggaaucucuuuccauagaau guucuaaacCauccugcggccucuacucugcauuc	Has a 5' G for U6 expression	2C
Tiling 70 nt 24 mismatch distance	gaccagguuccuggaacuggaaucucuuuccauaga auguucuaaacCauccugcggccucuacucugcau	Has a 5' G for U6 expression	2C
Tiling 70 nt 22 mismatch distance	gguaccagguuccuggaacuggaaucucuuuccaua gaauguucuaaacCauccugcggccucuacucugc	Has a 5' G for U6 expression	2C
Tiling 70 nt 20 mismatch distance	gauguaccagguuccuggaacuggaaucucuuucca uagaauguucuaaacCauccugcggccucuacucu	Has a 5' G for U6 expression	2C
Tiling 70 nt 18 mismatch distance	gguauguaccagguuccuggaacuggaaucucuuuc cauagaauguucuaaacCauccugcggccucuacu	Has a 5' G for U6 expression	2C
Tiling 70 nt 16 mismatch distance	gacguauguaccagguuccuggaacuggaaucucuu uccauagaauguucuaaacCauccugcggccucua	Has a 5' G for U6 expression	2C
Tiling 70 nt 14 mismatch distance	gacacguauguaccagguuccuggaacuggaaucuc uuuccauagaauguucuaaacCauccugcggccuc	Has a 5' G for U6 expression	2C
Tiling 70 nt 12 mismatch distance	gcaacacguauguaccagguuccuggaacuggaaucu cuuuccauagaauguucuaaacCauccugcggcc	Has a 5' G for U6 expression	2C
Tiling 70 nt 10 mismatch distance	gcccaacacguauguaccagguuccuggaacuggaau cucuuuccauagaauguucuaaacCauccugcgg	Has a 5' G for U6 expression	2C
Tiling 70 nt 8 mismatch distance	ggacccaacacguauguaccagguuccuggaacugga aucucuuuccauagaauguucuaaacCauccugc	Has a 5' G for U6 expression	2C
Tiling 70 nt 6 mismatch distance	guugacccaacacguauguaccagguuccuggaacug gaaucucuuuccauagaauguucuaaacCauccu	Has a 5' G for U6 expression	2C
Tiling 70 nt 4 mismatch distance	gccuugacccaacacguauguaccagguuccuggaac uggaaucucuuuccauagaauguucuaaacCauc	Has a 5' G for U6 expression	2C
Tiling 70 nt 2 mismatch distance	guuccuugacccaacacguauguaccagguuccugga acuggaaucucuuuccauagaauguucuaaacCa	Has a 5' G for U6 expression	2C
Tiling 84 nt 84 mismatch distance	gCauccugcggccucuacucugcauucaauuacaua cugacacauucggcaacauguuuuuccuggguuuauu uucacacagucca	Has a 5' G for U6 expression	2C
Tiling 84 nt 82 mismatch distance	gacCauccugcggccucuacucugcauucaauuaca uacugacacauucggcaacauguuuuuccuggguuua uuuucacacaguc	Has a 5' G for U6 expression	2C
Tiling 84 nt 80 mismatch	gaaacCauccugcggccucuacucugcauucaauua	Has a 5' G for U6 expression	2C

distance	cauacugacacauucggcaacauguuuuuccugguuuauuuucacacag		
Tiling 84 nt 78 mismatch distance	gcuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccugguuuauuuucacac	Has a 5' G for U6 expression	2C
Tiling 84 nt 76 mismatch distance	guucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggguuuauuuucac	Has a 5' G for U6 expression	2C
Tiling 84 nt 74 mismatch distance	guguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccugguuuauuuuc	Has a 5' G for U6 expression	2C
Tiling 84 nt 72 mismatch distance	gaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggguuuauuu	Has a 5' G for U6 expression	2C
Tiling 84 nt 70 mismatch distance	gagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggguuuau	Has a 5' G for U6 expression	2C
Tiling 84 nt 68 mismatch distance	gauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggguuu	Has a 5' G for U6 expression	2C
Tiling 84 nt 66 mismatch distance	gccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccuggu	Has a 5' G for U6 expression	2C
Tiling 84 nt 64 mismatch distance	guuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuuccug	Has a 5' G for U6 expression	2C
Tiling 84 nt 62 mismatch distance	gcuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcacauguuuuucc	Has a 5' G for U6 expression	2C
Tiling 84 nt 60 mismatch distance	gcucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuuuu	Has a 5' G for U6 expression	2C
Tiling 84 nt 58 mismatch distance	gaucucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacauguuu	Has a 5' G for U6 expression	2C
Tiling 84 nt 56 mismatch distance	ggaaucucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacaugu	Has a 5' G for U6 expression	2C
Tiling 84 nt 54 mismatch distance	guggaaucucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaacau	Has a 5' G for U6 expression	2C
Tiling 84 nt 52 mismatch distance	gacuggaaucucuuuccauagaauguucuaaacCauccugcggccucuacucugcauucaauuacauacugacacauucggcaac	Has a 5' G for U6 expression	2C
Tiling 84 nt 50 mismatch	ggaacuggaaucucuuuccauagaauguucuaaacC	Has a 5' G for U6 expression	2C

distance	auccugcggccucuacucugcauucaauuacauacug acacauucggca		
Tiling 84 nt 48 mismatch distance	guggaacuggaaucucuuuccauagaauguucuaaa cCauccugcggccucuacucugcauucaauuacaua cugacacauucgg	Has a 5' G for U6 expression	2C
Tiling 84 nt 46 mismatch distance	gccuggaacuggaaucucuuuccauagaauguucua aacCauccugcggccucuacucugcauucaauuaca uacugacacauuc	Has a 5' G for U6 expression	2C
Tiling 84 nt 44 mismatch distance	guuccuggaacuggaaucucuuuccauagaauguuc uaaacCauccugcggccucuacucugcauucaauua cauacugacacau	Has a 5' G for U6 expression	2C
Tiling 84 nt 42 mismatch distance	ggguuccuggaacuggaaucucuuuccauagaaugu ucuaaacCauccugcggccucuacucugcauucaau uacauacugacac	Has a 5' G for U6 expression	2C
Tiling 84 nt 40 mismatch distance	gcagguuccuggaacuggaaucucuuuccauagaau guucuaaacCauccugcggccucuacucugcauuca auuacauacugac	Has a 5' G for U6 expression	2C
Tiling 84 nt 38 mismatch distance	gaccagguuccuggaacuggaaucucuuuccauaga auguucuaaacCauccugcggccucuacucugcauu caauuacauacug	Has a 5' G for U6 expression	2C
Tiling 84 nt 36 mismatch distance	gguaccagguuccuggaacuggaaucucuuuccaua gaauguucuaaacCauccugcggccucuacucugca uucaauuacauac	Has a 5' G for U6 expression	2C
Tiling 84 nt 34 mismatch distance	gauguaccagguuccuggaacuggaaucucuuucca uagaauguucuaaacCauccugcggccucuacucug cauucaauuacau	Has a 5' G for U6 expression	2C
Tiling 84 nt 32 mismatch distance	gguauguaccagguuccuggaacuggaaucucuuuc cauagaauguucuaaacCauccugcggccucuacuc ugcauucaauuac	Has a 5' G for U6 expression	2C
Tiling 84 nt 30 mismatch distance	gacguauguaccagguuccuggaacuggaaucucuu uccauagaauguucuaaacCauccugcggccucuac ucugcauucaauu	Has a 5' G for U6 expression	2C
Tiling 84 nt 28 mismatch distance	gacacguauguaccagguuccuggaacuggaaucuc uuuccauagaauguucuaaacCauccugcggccucu acucugcauucaa	Has a 5' G for U6 expression	2C
Tiling 84 nt 26 mismatch distance	gcaacacguauguaccagguuccuggaacuggaaucu cuuuccauagaauguucuaaacCauccugcggccuc uacucugcauuc	Has a 5' G for U6 expression	2C
Tiling 84 nt 24 mismatch distance	gcccaacacguauguaccagguuccuggaacuggaau cucuuuccauagaauguucuaaacCauccugcggcc ucuacucugcau	Has a 5' G for U6 expression	2C
Tiling 84 nt 22 mismatch distance	ggacccaacacguauguaccagguuccuggaacugga aucucuuuccauagaauguucuaaacCauccugcgg ccucuacucugc	Has a 5' G for U6 expression	2C
Tiling 84 nt 20 mismatch	guugacccaacacguauguaccagguuccuggaacug	Has a 5' G for U6 expression	2C

distance	gaaucucuuuccauagaauguucuaaacCauccugc ggccucuacucu		
Tiling 84 nt 18 mismatch distance	gccuugacccaacacguauguaccagguuccuggaac uggaaucucuuuccauagaauguucuaaacCauccu gcggccucuacu	Has a 5' G for U6 expression	2C
Tiling 84 nt 16 mismatch distance	guuccuugacccaacacguauguaccagguuccugga acuggaaucucuuuccauagaauguucuaaacCauc cugcggccucua	Has a 5' G for U6 expression	2C
Tiling 84 nt 14 mismatch distance	ggguuccuugacccaacacguauguaccagguuccug gaacuggaaucucuuuccauagaauguucuaaacCa uccugcggccuc	Has a 5' G for U6 expression	2C
Tiling 84 nt 12 mismatch distance	guugguuccuugacccaacacguauguaccagguucc uggaacuggaaucucuuuccauagaauguucuaaac Cauccugcggcc	Has a 5' G for U6 expression	2C
Tiling 84 nt 10 mismatch distance	gccuugguuccuugacccaacacguauguaccagguu ccuggaacuggaaucucuuuccauagaauguucuaa acCauccugcgg	Has a 5' G for U6 expression	2C
Tiling 84 nt 8 mismatch distance	ggcccuugguuccuugacccaacacguauguaccagg uuccuggaacuggaaucucuuuccauagaauguucu aaacCauccugc	Has a 5' G for U6 expression	2C
Tiling 84 nt 6 mismatch distance	gccgcccuugguuccuugacccaacacguauguacca gguuccuggaacuggaaucucuuuccauagaauguu cuaaacCauccu	Has a 5' G for U6 expression	2C
Tiling 84 nt 4 mismatch distance	gcgccgcccuugguuccuugacccaacacguauguac cagguuccuggaacuggaaucucuuuccauagaaug uucuaaacCauc	Has a 5' G for U6 expression	2C
Tiling 84 nt 2 mismatch distance	ggucgccgcccuugguuccuugacccaacacguaugu accagguuccuggaacuggaaucucuuuccauagaa uguucuaaacCa	Has a 5' G for U6 expression	2C
ADAR non-targeting guide	GUAAUGCCUGGCUUGUCGACGCAU AGUCUG	Has a 5' G for U6 expression	2C
PFS binding screen guide for TAG motif	gaaaacgcagguuccucCaguuucgggagcagcgcac gucucccuguaguc	Has a 5' G for U6 expression	3B
PFS binding screen guide for AAC motif	gacgcagguuccucuagCuucgggagcagcgcacguc ucccuguagucaag	Has a 5' G for U6 expression	3B
PFS binding screen non-targeting	GUAAUGCCUGGCUUGUCGACGCAU AGUCUG	Has a 5' G for U6 expression	3B
Motif preference targeting guide	gauagaauguucuaaacCauccugcggccucuacuc ugcauucaauuacau	Has a 5' G for U6 expression	3C
Motif preference non-targeting guide	GUAAUGCCUGGCUUGUCGACGCAU AGUCUG	Has a 5' G for U6 expression	3C
PPIB tiling guide 50 mismatch distance	gCaaggccacaaaauuauccacuguuuuuggaacag ucuuuccgaagagac	Has a 5' G for U6 expression	S3D
PPIB tiling guide 42 mismatch distance	gccuguagcCaaggccacaaaauuauccacuguuuu uggaacagucuuucc	Has a 5' G for U6 expression	S3D

PPIB tiling guide 34 mismatch distance	gcuuucucuccuguagcCaaggccacaaaauuaucc acuguuuuuggaaca	Has a 5' G for U6 expression	S3D
PPIB tiling guide 26 mismatch distance	ggccaaauccuuucucuccuguagcCaaggccacaaa auuauccacuguuu	Has a 5' G for U6 expression	S3D
PPIB tiling guide 18 mismatch distance	guuuuuguagccaaauccuuucucuccuguagcCaa ggccacaaaauuauc	Has a 5' G for U6 expression	S3D
PPIB tiling guide 10 mismatch distance	gauuugcuguuuuuguagccaaauccuuucucuccu guagcCaaggccaca	Has a 5' G for U6 expression	S3D
PPIB tiling guide 2 mismatch distance	gacgauggaauuugcuguuuuuguagccaaauccuu ucucuccuguagcCa	Has a 5' G for U6 expression	S3D
Targeting guide, opposite base G	gauagaauguucuaaacGauccugcggccucuacuc ugcauucaauuacau	Has a 5' G for U6 expression	S3D
Targeting guide, opposite base A	gauagaauguucuaaacAauccugcggccucuacuc ugcauucaauuacau	Has a 5' G for U6 expression	S3D
Targeting guide, opposite base C	gauagaauguucuaaacUauccugcggccucuacuc ugcauucaauuacau	Has a 5' G for U6 expression	S3D
AVPR2 guide 37 mismatch distance	ggucccacgcggccCacagcugcaccaggaagaaggg ugcccagcacagca	Has a 5' G for U6 expression	4A
AVPR2 guide 35 mismatch distance	ggggucccacgcggccCacagcugcaccaggaagaag ggugcccagcacag	Has a 5' G for U6 expression	4A
AVPR2 guide 33 mismatch distance	gccgggucccacgcggccCacagcugcaccaggaaga agggugcccagcac	Has a 5' G for U6 expression	4A
FANCC guide 37 mismatch distance	gggugaugacauccCaggcgaucguguggccuccag gagcccagagcagga	Has a 5' G for U6 expression	4B
FANCC guide 35 mismatch distance	gagggugaugacauccCaggcgaucguguggccucc aggagcccagagcag	Has a 5' G for U6 expression	4B
FANCC guide 32 mismatch distance	gaucagggugaugacauccCaggcgaucguguggcc uccaggagcccagag	Has a 5' G for U6 expression	4B
Synthetic disease gene target IL2RG	gguggcuccauucacucCaaugcugagcacuuccac agaguggguuaaagc	Has a 5' G for U6 expression	4E
Synthetic disease gene target F8	guuucuaauauauuuugCcagacugauggacuauu cucaauuaauaaugau	Has a 5' G for U6 expression	4E
Synthetic disease gene target LDLR	gagauguugcuguggauCcaguccacagccagcccg ucgggggccuggaug	Has a 5' G for U6 expression	4E
Synthetic disease gene target CBS	gcaggccggcccagcugCcaggugcaccugcucggag caucgggccggauc	Has a 5' G for U6 expression	4E
Synthetic disease gene target HBB	gcaaagaaccucuggguCcaaggguagaccaccagca gccugcccagggcc	Has a 5' G for U6 expression	4E
Synthetic disease gene target ALDOB	gaagagaaacuuaguuuCcagggcuuugguagagg gcaaagguugauagca	Has a 5' G for U6 expression	4E
Synthetic disease gene target DMD	gucagccuagugcagagCcacugguaguuggugguu agaguuucaaguucc	Has a 5' G for U6 expression	4E
Synthetic disease gene target SMAD4	ggcucauugugaacaggCcaguaaguccgggaugg ggcggcauaggcggg	Has a 5' G for U6 expression	4E
Synthetic disease gene	guagcuaaagaacuugaCcaagacauaucaggaucc	Has a 5' G for U6 expression	4E

target BRCA2	acccucagcuccuaga		
Synthetic disease gene target GRIN2A	ggggcauuguucugugcCcaguccugcugguagaccugcucccgguggcu	Has a 5' G for U6 expression	4E
Synthetic disease gene target SCN9A	gagaagucguucaugugCcaccgugggagcguacagucaucauugaucuug	Has a 5' G for U6 expression	4E
Synthetic disease gene target TARDBP	gggauuaaugcugaacgCaccaaaguucaucccaccacccauauuacuacc	Has a 5' G for U6 expression	4E
Synthetic disease gene target CFTR	gcuccaaaggcuuuccuCcacuguugcaaaguuauugaaucccaagacaca	Has a 5' G for U6 expression	4E
Synthetic disease gene target UBE3A	gaugaaugaacgauuucCcagaacucccuaaucagaacagagucccuggua	Has a 5' G for U6 expression	4E
Synthetic disease gene target SMPD1	ggagccucugccggagcCcagagaacccgagagucagacagagccagcgcc	Has a 5' G for U6 expression	4E
Synthetic disease gene target USH2A	ggcuuccguggagacacCcaaucaauuugaagagaucuugaagugaugcca	Has a 5' G for U6 expression	4E
Synthetic disease gene target MEN1	gugggacugcccuccucCcauuugcagaugccgucguagaaucgcagcagg	Has a 5' G for U6 expression	4E
Synthetic disease gene target C8orf37	gcuucuucaauaguucuCcagcuacacuggcaggcauaugcccguguuccu	Has a 5' G for U6 expression	4E
Synthetic disease gene target MLH1	gauuccuuuucuucgucCcaauucaccucaguggcuagucgaagaaugaag	Has a 5' G for U6 expression	4E
Synthetic disease gene target TSC2	gcagcuucagcaccuucCagucagacuccugcuucaagcacugcagcagga	Has a 5' G for U6 expression	4E
Synthetic disease gene target NF1	gccauuugcuugcagugCcacuccagaggauuccggauugccauaaaauacu	Has a 5' G for U6 expression	4E
Synthetic disease gene target MSH6	guucaauaguuuuggucCaguaucguuuacagcccuucuugguagauuuca	Has a 5' G for U6 expression	4E
Synthetic disease gene target SMN1	ggcaaccgucuucugacCaaauggcagaacauuuguccccaacuuuccacu	Has a 5' G for U6 expression	4E
Synthetic disease gene target SH3TC2	gcgacuuuccaaugaacCacugaagcccagguaugacaaagccgaugaucu	Has a 5' G for U6 expression	4E
Synthetic disease gene target DNAH5	guuuacacucaugcuucCacagcuuuaacagaucauuugguuccuugauga	Has a 5' G for U6 expression	4E
Synthetic disease gene target MECP2	gcuuaagcuuccgugucCagccuucaggcagggugggucaucauacaugg	Has a 5' G for U6 expression	4E
Synthetic disease gene target ADGRV1	ggacagcugggcugaucCaugaugucauccagaaacacuggggacccucag	Has a 5' G for U6 expression	4E
Synthetic disease gene target AHI1	gucucaucucaacuuucCauauccguaucauggaaucauagcauccuguaa	Has a 5' G for U6 expression	4E
Synthetic disease gene target PRKN	gcaugcagacgcgguucCacucgcagccacaguuccagcaccacucgagcc	Has a 5' G for U6 expression	4E
Synthetic disease gene target COL3A1	guugguuagggucaaccCaguauucuccacucuugaguucaggauggcaga	Has a 5' G for U6 expression	4E
Synthetic disease gene target BRCA1	gcuacacuguccaacacCcacucucgggucaccacaggugccucacacauc	Has a 5' G for U6 expression	4E

Synthetic disease gene target MYBPC3	gcugcacuguguaccccCagagcuccguguugccgac auccugggguggcu	Has a 5' G for U6 expression	4E
Synthetic disease gene target APC	gagcuuccugccacuccCaacagguuucacaguaagc gcguaucuguucca	Has a 5' G for U6 expression	4E
Synthetic disease gene target BMPR2	gacggcaagagcuuaccCagucacuuguguggagac uuaaauacuugcaua	Has a 5' G for U6 expression	4E
KRAS tiling guide 50 mismatch distance	gCaaggccacaaaauuauccacuguuuuuggaacag ucuuuccgaagagac	Has a 5' G for U6 expression	5A
KRAS tiling guide 42 mismatch distance	gccuguagcCaaggccacaaaauuauccacuguuuu uggaacagucuuucc	Has a 5' G for U6 expression	5A
KRAS tiling guide 34 mismatch distance	gcuuucucuccuguagcCaaggccacaaaauuaucc acuguuuuuggaaca	Has a 5' G for U6 expression	5A
KRAS tiling guide 26 mismatch distance	ggccaaauccuuucucuccuguagcCaaggccacaaa auuauccacuguuu	Has a 5' G for U6 expression	5A
KRAS tiling guide 18 mismatch distance	guuuuuguagccaaauccuuucucuccuguagcCaa ggccacaaaauuauc	Has a 5' G for U6 expression	5A
KRAS tiling guide 10 mismatch distance	gauuugcuguuuuuguagccaaauccuuucucuccu guagcCaaggccaca	Has a 5' G for U6 expression	5A
KRAS tiling guide 2 mismatch distance	gacgauggaauuugcuguuuuuguagccaaauccuu ucucuccuguagcCa	Has a 5' G for U6 expression	5A
KRAS tiling non-targeting guide	GUAAUGCCUGGCUUGUCGACGCAU AGUCUG	Has a 5' G for U6 expression	5A
Luciferase W85X targeting guide for transcriptome specificity	gauagaauguucuaaacCauccugcggccucuacuc ugcauucaauuacau	Has a 5' G for U6 expression	5B
Non-targeting guide for transcriptome specificity	GCAGGGUUUUCCCAGUCACGACGU UGUAAAGUUG	Has a 5' G for U6 expression	5C
endogenous KRAS guide 2	gucaaggcacucuugccCacgccaccagcuccaacua ccacaaguuuauau	Has a 5' G for U6 expression	6F
endogenous PPIB guide 1	gcaaagaucacccggccCacaucuucaucuccaauuc guaggucaaaauac	Has a 5' G for U6 expression	6G
endogenous KRAS guide 1	GcgccaccagcuccaacCaccacaaguuuauauuca gucauuuucagcagg	Has a 5' G for U6 expression	S13A
endogenous KRAS guide 3	GuuucuccaucaauuacCacuugcuuccuguagga auccucuauuGUugga	Has a 5' G for U6 expression	S13B
endogenous PPIB guide 2	GcuuucucuccuguagcCaaggccacaaaauuaucc acuguuuuuggaaca	Has a 5' G for U6 expression	S13C
endogenous non-targeting guide	GUAAUGCCUGGCUUGUCGACGCAU AGUCUG	Has a 5' G for U6 expression	6F
BoxB Cluc guide	ucuuuccauaGGCCCUGAAAAAGGGCCu guucuaaacCauccugcggccucuacucGGCCC UGAAAAAGGGCCauucaauuac	Has a 5' G for U6 expression	S8B
BoxB non-targeting guide	cagcuggcgaGGCCCUGAAAAAGGGCCg gggaugugcCgcaaggcgauuaaguuggGGCCC UGAAAAAGGGCCacgccagggu	Has a 5' G for U6 expression	S8B

Stafforst full length ADAR2 guide 1	GUGGAAUAGUAUAACAAUAUGCUAAAUGUUGUUAUAGUAUCCCACucuaaaCCAuccugcgGGGCCCUCUUCAGGGCCC	Has a 5' G for U6 expression	S8C
Stafforst full length ADAR2 non-targeting guide	GUGGAAUAGUAUAACAAUAUGCUAAAUGUUGUUAUAGUAUCCCACacccuggcguuacccaGGGCCCUCUUCAGGGCCC	Has a 5' G for U6 expression	S8C

Chapter 4

Discussion and future directions

The work presented in this thesis describes the discovery, characterization and application of type VI CRISPR systems encoding Cas13 RNA-guided RNases, with a special emphasis on type VI-B CRISPR loci and the Cas13b RNase they encode. However, questions remain relating to both the molecular mechanism of type VI CRISPR function and their applications for transcriptome editing.

How do type VI-B systems perform CRISPR adaptation?

Work presented in this thesis describes crRNA biogenesis and interference of type VI-B CRISPR systems. However, adaptation of type VI-B loci in response to mobile genetic elements has not been demonstrated (*1*). VI-B loci lack *cas1* and *cas2*, which typically mediate insertion of new spacers into the CRISPR array, sometimes with the help of additional Cas proteins derived from the same locus (*2-4*). Although there are examples of individual CRISPR loci that lack *cas1* and *cas2*, the existence of an entire subtype of CRISPR that lacks these genes suggests a different mechanism operates for adaptation of VI-B loci to exogenous nucleic acids.

One possibility is that Cas1 and Cas2 are supplied in *trans* from other CRISPR loci present in the genome. Expressing VI-B and other naturally co-occurring CRISPR loci together in *E. coli* lacking endogenous CRISPR systems, challenging these cells with bacteriophage and sequencing the VI-B CRISPR array to see if adaptation has occurred, could test this possibility. This hypothesis is appealing due to the high co-occurrence of VI-B loci with distant *cas1* and/or *cas2* genes and the strict genetic requirement of *cas1* and *cas2* for adaptation across multiple different types of CRISPR loci (*2, 5, 6*).

It is also possible that type VI-B systems are sufficient for adaptation and don't require *cas1* or *cas2*. Reconstituting VI-B loci in *E. coli* that lack other CRISPR systems and testing their ability to adapt through challenge with bacteriophage could address this possibility. Adaptation independent of *cas1/cas2* has not been reported, and these genes have been shown to be genetically required for adaptation across diverse types of CRISPR loci, making this possibility unlikely.

What is the molecular basis for Csx27 and Csx28 modulation of Cas13b activity?

The RNA-interference activity of Cas13b is repressed and enhanced by the genetic expression of cognate *csx27* and *csx28* genes, respectively. What is mechanism by which the products of these genes modulate Cas13b activity? Providing a detailed answer to this question will likely require biochemical studies. *csx* gene products could mediate their effects by binding to Cas13b, the crRNA, or both to change interference levels. It is also possible that *csx* gene products modulate RNA-interference activity without physically interacting with Cas13b proteins.

By testing tagged versions of Cas13b and a cognate Csx protein for their ability to co-purify it should be possible to assess their direct interaction *in vivo*. Such an interaction may also be crRNA-dependent; by including both crRNA expressing and crRNA negative controls in co-purification studies, the contribution of this element to Csx function can be assessed. The findings of physical interactions from these experiments could be further supported by super-shift assays using purified Csx, Cas13b and crRNA. If shown to interact physically, then time-course cleavage assays supplemented with Csx proteins could help to

understand if this interaction alters the kinetics of either specific or non-specific RNase activities of Cas13b.

Bacterial genetics could be used to address the possibility that *csx* gene products influence RNA interference independent of direct interaction with Cas13b. By comparing resistance to MS2 bacteriophage of bacteria harboring only *csx* genes, only *cas13b*, *csx* and *cas13b*, and vector controls, it should be possible to determine if *csx* effects on RNA interference are dependent on the presence of *cas13b*. Another possibility is that Csx proteins regulate expression or translation of Cas13b *in vivo* to modulate interference activity. By measuring the levels of tagged Cas13b by western blot with and without expressed *csx* genes, it should be possible to assess the validity of this hypothesis.

A genetic approach could also address the function of the putative HEPN domain in *csx28* and whether it mediates increased interference. Mutating catalytic residues in the *csx28* HEPN domain and comparing MS2 resistance levels to constructs with the wild type gene can clarify the involvement of this putative nuclease domain in RNA interference activity.

What is the molecular basis for non-specific RNase activity by Cas13 enzymes?

This thesis and other studies suggest that following on-target cleavage of RNA substrates, both Cas13a and Cas13b exhibit subsequent non-specific RNase activity termed the collateral effect (*1, 7, 8*). Collateral activity has been observed *in vitro* for both enzymes and is dependent on initial cleavage of a target with complementarity to the crRNA by Cas13a/b (*1, 7*). Importantly, binding to target sequences with HEPN mutant versions of Cas13a/b does not lead to non-specific

cleavage, implicating this domain in both on and off-target RNase activity (*8*). Insight into the mechanism of off-target activity by Cas13 enzymes has been guided by structural studies of Cas13a; whether similar structures underlie Cas13b activity remains to be seen.

Initial studies of Cas13a in complex with a crRNA suggested that conformational changes underlie nuclease activation, showing that the HEPN domains are positioned on the external surface of the protein, facing away from the site of crRNA binding (*9, 10*). Recently, a structure of the ternary of Cas13a from *Leptotrichia buccalis* with a crRNA and a short target RNA has been solved, confirming that targeting binding induces a significant conformational change of the enzyme (*11*). However, the HEPN domains in the target bound structure still face externally and do not directly interact with the short ssRNA target, suggesting that a long ssRNA target is needed to reach the HEPN domains for *cis* cleavage of targets. It is tempting to hypothesize that the conformational change induced by target binding activates the HEPN domains, which, still facing externally, are not restricted to interacting with the crRNA target, allowing for cleavage of targets in *trans*, leading to collateral activity.

Further structural studies will likely yield additional insights into the non-specific RNase activity of Cas13 enzymes. It still remains possible that the recently crystalized ternary complex of Cas13a is not the nuclease active conformation, but rather a stable, non-active conformation of the enzyme. Solving the ternary structures of Cas13s with a crRNA and a longer target sequence that can reach the HEPN domains would likely shed light on whether this structure is indeed a nuclease competent conformation that could mediate the collateral effect.

Fully explaining collateral activity might require structural studies of a quaternary structure of Cas13, a crRNA, target RNA and a 'collateral' RNA, that could directly capture HEPN association with off-target substrates.

Why do Cas13 enzymes fail to exhibit off-target RNase activity in mammalian cells?

The absence of non-specific RNase activity for Cas13 enzymes in human cell is supported by RNA-sequencing data and the absence of a growth suppression phenotype that is thought to be the result of such activity in prokaryotes (*7, 12, 13*). This observation is especially puzzling in light of the observation of on-target cleavage activity of transcripts by both Cas13a and Cas13b enzymes in human cell lines (*12, 13*). A satisfying explanation will likely require an understanding of the mechanistic details of such behavior. Nevertheless, it is still possible to generate some hypotheses by making broad assumptions about the nature of collateral cleavage.

Assuming that the same Cas13 molecule is responsible for on and off-target RNase cleavage, one possibility is that the half-life of Cas13 molecules is significantly reduced in eukaryotic cells relative to prokaryotic or *in vitro* contexts, allowing for the initial on-target cleavage event, but undetectable off-target activity that would normally follow. Decreased Cas13 stability could be due to the chemical composition of the eukaryotic cytoplasm or protein degradation pathways that differ from *in vitro* and prokaryotic environments. Comparing the *in vivo* half-life of Cas13 using pulse-chase analysis under activating and non-activating conditions in prokaryotic and eukaryotic cells could evaluate this hypothesis. Additionally, testing a wide range of *in vitro* buffer compositions for specific and

non-specific Cas13 RNase activity may provide additional insight as to whether these activities can be separated chemically, aiding in the explanation of this phenomenon.

Because the nuclease domains of Cas13a face externally from the crRNA:target binding channel, it has been proposed that HEPN nuclease activity supplied in *trans* from another Cas13 molecule may be required for cleavage and that Cas13 enzymes may function as dimers, to execute RNase activity(*9*). If the nuclease active state of Cas13 is a dimer, it could make collateral activity more sensitive to enzyme concentration levels—which likely differ between *in vitro*, prokaryotic and eukaryotic contexts. Titrating Cas13 enzyme levels for *in vitro* cleavage assays could assess the hypothesis that collateral activity is Cas13-concentration dependent. If concentration of Cas13 does control the balance of RNase activities, an enzyme concentration should exist for which only specific RNase activity is observed.

Ultimately, designing experiments to address the lack of non-specific RNase activity by Cas13 enzymes in mammalian cells will be challenging in the absence of a detailed molecular understanding of how such activity arises. Once a mechanistic understanding of collateral activity is established, precise experiments utilizing this information can help to clarify the curious lack of non-specific RNase cleavage by Cas13 in mammalian cells.

Improving CRISPR-Cas13 RNA base editing

Chapter 3 describes programmable adenosine to inosine editing using catalytically inactive Cas13b (dCas13b) fused to the catalytic domain of human

adenosine deaminase acting on RNA 2 ($ADAR2_{DD}$) (*13*). There are a number of ways that this technology can be improved.

In the work presented, increased specificity of editing was accompanied by a concomitant decrease in on-target editing rates. Ideally, high levels of on-target editing would attend specific editing by dCas13b-$ADAR2_{DD}$ fusions. By screening additional $ADAR2_{DD}$ RNA binding mutations using a reporter that can estimate on and off-target editing rates, as described in chapter 3, it should be possible to evaluate if novel mutants exist that exhibit both high specificity and on-target editing.

An attractive feature of RNA base-editing technologies is that they are likely to be robust to cell state, as they rely on direct chemical conversion of bases to recode transcripts. Testing the validity of this assertion by expressing RNA base editing machinery in post mitotic cells, which are unable to utilize HDR to achieve precise editing outcomes, could be an exciting area of future research.

In contrast to HDR, RNA base editing can only encode specific sequence changes at present, limiting its utility in research and therapeutic settings. Expanding base-editing capabilities to encode additional sequence changes would enable its application in a wider range of contexts. This could be achieved through fusion of dCas13 enzymes to other, naturally occurring RNA-editing enzymes with different substrate specificity, such as cytidine deaminases that convert cytidine to uridine (*14*). Using catalytically inactive Cas9 fusions to cytidine deaminases has already shown utility for DNA base editing applications, suggesting that this strategy might work to expand RNA editing to other bases (*15*). Alternatively, it

may be possible to evolve existing RNA-editing domains to accept new substrates to expand base-editing capabilities.

Transcriptome editing applications of Cas13 in mammalian cells

Transcriptome editing can be broadly defined as the ability to remove or modify the function of endogenous transcripts. Powerful technologies exist that can suppress transcript function in mammalian cells (*16-19*). The description of a dCas13b enzyme with RNA-binding activity in mammalian cells opens the possibility for a simple method to modify mammalian transcript function. By fusing RNA modifying enzymes to dCas13b, it should therefore be possible to reconstitute enzymatic function on endogenous transcripts, similar to the work presented in chapter 3 with dCas13b-ADAR$_{DD}$ fusions.

Future studies will likely explore the extent to which enzymatic functions distinct from adenosine deamination can be reconstituted on transcripts using dCas13b fusions. A range of therapeutic and research applications for programmable RNA binding proteins has been previously proposed, but exploration was hindered by the low re-programmability of earlier technologies (*20*). Therapeutic applications of Cas13 could include development as a tool to regulate alternative splicing patterns, modulation of translation, or allele specific transcriptional targeting.

Cas13 could also be used in a research context to explore the function of RNA modifications whose basic or contextual function is unclear by allowing control over the timing and location of such marks. Of particular interest for this type of application is the developing field of epitranscriptomics, which seeks to

understand how novel RNA modifications lead to post-transcriptional gene regulation (*21*). With the recent development of techniques to detect such modifications, technologies that can be used to modify their existence, such as dCas13b fusion to RNA modifying enzymes, will become increasingly important to understand their function.

Conclusion

CRISPR-Cas systems utilize diverse mechanisms to achieve cleavage of target nucleic acids to provide immunity to mobile genetic elements. DNA nucleases from class 2 CRISPR systems have improved genome editing significantly by providing a simple method for targeting genomic loci for cleavage or recruitment of DNA-modifying domains. I presented work in this thesis describing the discovery and characterization of the type VI-B class 2 CRISPR system that encodes an RNA-guided RNase, Cas13b. I show that Cas13b orthologs exhibit RNase activity in both prokaryotic and eukaryotic cells and that catalytically inactive versions of this enzyme can be used to direct enzymatic activity to endogenous transcripts. The sum of this work has been to extend the simplicity and functionality of CRISPR genome editing to the level of RNA. By allowing RNA-targeting in diverse contexts, Cas13b has the potential to be utilized for a broad range of therapeutic and research applications.

References

1. A. A. Smargon *et al.*, Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular cell* **65**, 618 (Feb 16, 2017).
2. R. Heler *et al.*, Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199 (Mar 12, 2015).
3. S. Shmakov *et al.*, Diversity and evolution of class 2 CRISPR-Cas systems. *Nature reviews. Microbiology* **15**, 169 (Mar, 2017).
4. R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709 (Mar 23, 2007).
5. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic acids research* **40**, 5569 (Jul, 2012).
6. Y. Wei, R. M. Terns, M. P. Terns, Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes & development* **29**, 356 (Feb 15, 2015).
7. O. O. Abudayyeh *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (Aug 05, 2016).
8. A. East-Seletsky *et al.*, Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270 (Oct 13, 2016).
9. L. Liu *et al.*, Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168**, 121 (Jan 12, 2017).
10. G. J. Knott *et al.*, Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nature structural & molecular biology* **24**, 825 (Oct, 2017).
11. L. Liu *et al.*, The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170**, 714 (Aug 10, 2017).
12. O. O. Abudayyeh *et al.*, RNA targeting with CRISPR-Cas13. *Nature* **550**, 280 (Oct 12, 2017).
13. D. B. T. Cox *et al.*, RNA editing with CRISPR-Cas13. *Science* **358**, 1019 (Nov 24, 2017).
14. S. G. Conticello, The AID/APOBEC family of nucleic acid mutators. *Genome biology* **9**, 229 (2008).

15. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420 (May 19, 2016).
16. S. M. Elbashir *et al.*, Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494 (May 24, 2001).
17. A. Fire *et al.*, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806 (Feb 19, 1998).
18. M. L. Stephenson, P. C. Zamecnik, Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 285 (Jan, 1978).
19. P. C. Zamecnik, M. L. Stephenson, Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 280 (Jan, 1978).
20. J. P. Mackay, J. Font, D. J. Segal, The prospects for designer single-stranded RNA-binding proteins. *Nature structural & molecular biology* **18**, 256 (Mar, 2011).
21. M. Helm, Y. Motorin, Detecting RNA modifications in the epitranscriptome: predict and validate. *Nature reviews. Genetics* **18**, 275 (May, 2017).