

# UC Berkeley

## UC Berkeley Electronic Theses and Dissertations

### Title

Molecular mechanisms and applications of RNA targeting CRISPR endonucleases

### Permalink

<https://escholarship.org/uc/item/8jf7385p>

### Author

Seletsky, Alexandra East

### Publication Date

2017

Peer reviewed|Thesis/dissertation

# **Molecular mechanisms and applications of RNA targeting CRISPR endonucleases**

by

Alexandra Seletsky

A dissertation submitted in partial satisfaction  
of the requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jennifer Doudna, Chair

Professor Donald Rio

Professor David Schaffer

Professor Kathleen Ryan

Spring 2017

## ABSTRACT

Molecular mechanisms and applications of RNA targeting CRISPR endonucleases  
by  
Alexandra Seletsky

Doctor of Philosophy in Molecular and Cell Biology  
University of California, Berkeley

Professor Jennifer A. Doudna, Chair

Evolutionary pressure to protect against phage-induced lysis and rampant horizontal gene transfer has created a wide repertoire of defensive pathways in bacteria. CRISPR-Cas (clustered regularly interspaced short palindromic repeats, CRISPR-associated) systems are adaptive immune pathways that use RNA-guided nucleases to direct cleavage of invading nucleic acids. The programmable nature of these enzymes has enabled a revolution for DNA-targeting applications including gene editing, transcriptional control, and genomic imaging. In addition to DNA-targeting enzymes, specific subtypes of CRISPR-Cas systems recognize and degrade single stranded RNA (ssRNA) substrates. Repurposing these ssRNA-targeting enzymes into biotechnological tools is currently limited by a lack of mechanistic information. In this work, we address this issue by redirecting a well-studied DNA-targeting CRISPR nuclease, Cas9, to ssRNA targets and investigating the enzymatic mechanisms of a novel ssRNA-targeting CRISPR nuclease, Cas13a (formerly C2c2).

Typically, Cas9 ignores ssRNA while searching for dsDNA targets due to ssRNA's inherent single-stranded structure and lack of a protospacer adjacent motif (PAM). We redirected Cas9 to bind and recognize ssRNA targets through addition of a third component, a target-complementary DNA oligonucleotide or PAMmer, that provides a DNA:RNA hybrid PAM. Using primary microRNAs as a model system, we provide proof-of-concept evidence that Cas9:PAMmer complexes can mediate the isolation and subsequent mass spectrometry analysis of protein complexes bound to specific RNAs.

The complexity of redirecting Cas9 to ssRNA substrates motivated us to investigate CRISPR proteins that natively target RNA. We focused on Cas13a, a predicted ribonuclease from Type VI CRISPR-Cas systems. We discovered that Cas13a possesses two distinct catalytic activities, one for site-specific cleavage of its CRISPR RNA (crRNA) and the second for nonspecific ssRNA degradation activated by target binding. These insights allowed us to establish a revised model for ssRNA-targeting by Type VI CRISPR-Cas systems. Through biochemical characterization of the entire Cas13a protein family, we revealed hidden diversity in substrate preferences and defined orthogonal enzyme subfamilies. These diverse Cas13a homologs can be harnessed in parallel for detection of distinct RNA species within complex mixtures for both bacterial immunity and diagnostic applications. Together, this work presents two novel biotechnological applications of CRISPR-Cas nucleases for RNA isolation and RNA detection.

## ACKNOWLEDGEMENTS

Few doctoral dissertations are done in isolation, and this one is no exception. To the many people I leave out from this short note, thank you for your support, love, and comradely during my journey to a PhD at the end of my name.

Two people stand out as essential to my success as a graduate student. First thanks go to Jennifer Doudna, who welcomed a naïve rotation student with no biochemistry background into her structural biology lab, pitched a crazy sounding project (See appendix 1), provided a stimulating environment and supported me through my project-hopping path. Secondly, none of this work would have happened without the mentorship, friendship and support of Mitchell O'Connell. You shared your project when I had given up on the Doudna lab; I had been scientifically brutalized by a labmate, scoped by multiple other labs, and was inches from quitting graduate school. Beyond being a fantastic mentor, you reminded me that I did actually like science and reinforced my belief that better science comes from working together. I know any student who joins your lab at Rochester is in good hands.

To the many lab mates who have surrounded me during my years in the Doudna lab, I am grateful to your input, criticism, and dedication to science. Kevin Doxzen, my lab-dance-party-partner and Akshay Tambe, my bullshit meter for all things math and sequencing related, were the best cohort-mates I could ask for. Thanks to Megan Hochstrasser, Emine Kaya, Philip Kranzusch, Ross Wilson, Yun Bai, Steven Lin, and Stephen Floor, who answered all my questions as I learned *in vitro* biochemistry reluctantly. Graduate school would not be the same without my Yali's buddy and confidant Steven Strutt. David (Dudu) Burstein was the best for patiently answering all my bioinformatics questions and having an infectious love of science. Ben Oakes and Janice Chen, both survived rotating in Mitch and I's bay, and still talked to me afterwards?!?! To Addison Wright, and Megan for being my go-to editors, my lack of English skills required your sharp grammar eyes to fix. The Cas13a project wouldn't have happened without Spencer Knight to help kickstart it. And my sincerest apologies to Gavin Knott for convincing him that C2c2 crystallography was a viable project. Beyond the Doudna lab, I am grateful to my MCB cohort including Laura Wetzel, Shion Lim, Ethan MacSpadden, and Alli Quan for the coffee dates, qual studying and beer hour antics.

To my wonderful husband, Oleg Seletsky, who let me drag him out to the west coast, subsidized my rent for too many years, and puts up with my nonsense. Without you grounding me, I could achieve nothing. And finally, Ms. Ada (Addie) Lovelace Seletsky, your tail wailing, slobbery kisses welcome me home. No day is a bad day, with your bright eyes, puppy-smile, and fluffy cuddles washing away my self-doubts, anxieties, and concerns.



## DEDICATION

*To my parents, Bryan and Stacey East.*

*Dad, I admire your optimism and courage,  
standing tall despite losing your feet and innumerable challenges.*

*Mom, your dedication and resilience  
inspires me everyday to achieve more.*

*While neither of you studied science,  
you taught me that courage, optimism, dedication and resilience  
are the four skills required for success  
not only in science, but also in life.*

# TABLE OF CONTENTS

Abstract .....	1
Acknowledgements .....	i
Dedication .....	ii
Table of Contents .....	iii
List of Figures .....	vi
List of Tables .....	viii

## Chapter 1: Introduction to CRISPR-Cas Immune Systems

<b>1.1 Overview of bacterial defense systems.....</b>	<b>2</b>
1.1.1 Innate strategies to evade phage infection .....	2
1.1.2 Adaptive strategies to combat phage infection .....	4
<b>1.2 CRISPR-Cas system classification .....</b>	<b>4</b>
1.2.1 Class 1 systems.....	5
1.2.2 Class 2 systems.....	6
<b>1.3 Three major steps of the CRISPR-Cas pathway .....</b>	<b>8</b>
1.3.1 Acquisition of new spacers .....	9
1.3.2 Biogenesis of pathway components .....	11
1.3.3 Target search and interference of phage infection .....	13
<b>1.4 Applications of CRISPR-Cas systems .....</b>	<b>16</b>

## Chapter 2: Redirecting CRISPR Type II systems for isolating specific RNA:protein complexes

<b>2.1 Chapter Summary.....</b>	<b>18</b>
<b>2.2 Introduction.....</b>	<b>18</b>
<b>2.3 Methods .....</b>	<b>19</b>
2.3.1 Cas9 and nucleic acid preparation .....	19
2.3.2. Cleavage assays .....	22
2.3.3. RNA cleavage site mapping .....	22
2.3.4. Electrophoretic mobility shift assays .....	22
2.3.5. Cas9 Biotin Labeling.....	23
2.3.6. GAPDH mRNA pull-down .....	23
2.3.7. Design of primary microRNA targeting experiments.....	24
2.3.8 qPCR validation of pri-miRNA pulldown .....	25
2.3.9 pri-miRNA isolation and mass spectrometry of binding proteins .....	25
2.3.10 Mass spectrometry of elutions. ....	26
2.3.11 Processing of mass spectrometry data.....	27
<b>2.4 Results.....</b>	<b>27</b>
2.4.1 RNA-guided Cas9 cleaves ssRNA targets in the presence of a short PAM- presenting DNA oligonucleotide (PAMmer) .....	27
2.4.2 dCas9–gRNA binds ssRNA targets with high affinity in the presence of	

	PAMmers .....	29
2.4.3	PAMmer guided dCas9-gRNA can isolate GAPDH mRNA from total RNA and cell lysates.....	30
<b>2.5</b>	<b>Preliminary Results .....</b>	<b>33</b>
2.5.1	pri-miRNA target design and validation .....	33
2.5.2	Mass spectrometry analysis of pri-miRNA binding proteins .....	35
<b>2.6</b>	<b>Discussion.....</b>	<b>39</b>

## **Chapter 3: Discovery of RNA targeting single effector CRISPR systems**

<b>3.1</b>	<b>Summary and Introduction .....</b>	<b>42</b>
<b>3.2</b>	<b>Methods .....</b>	<b>42</b>
3.2.1	C2c2 phylogenic and candidate selection .....	42
3.2.2	C2c2 protein production and purification .....	43
3.2.3	Generation of RNA .....	45
3.2.4	Pre-crRNA processing assays .....	46
3.2.5	Product Size Mapping and 3' end moiety identification .....	46
3.2.6	Small RNA sequencing analysis .....	47
3.2.7	Target cleavage assays .....	47
3.2.8	crRNA filter-binding assays.....	47
3.2.9	Electrophoretic mobility-shift assays.....	48
3.2.10	Fluorescent RNA detection assay .....	48
3.2.11	Background cleavage in total RNA .....	49
<b>3.3</b>	<b>Results.....</b>	<b>49</b>
3.3.1	C2c2 catalyzes crRNA maturation for Type VI-A systems .....	49
3.3.2	Substrate requirements and mechanism of pre-crRNA processing by C2c2.....	50
3.3.3	C2c2:crRNA complexes degrade ssRNA upon target complementarity...	54
3.3.4	pre-crRNA processing and ssRNA targeting activities occupy distinct active sites .....	59
3.3.5	RNA detection by C2c2 harness for biotechnological applications.....	61
<b>3.4</b>	<b>Discussion.....</b>	<b>64</b>

## **Chapter 4: Delving into the Cas13a Family Tree**

<b>4.1</b>	<b>Summary .....</b>	<b>66</b>
<b>4.2</b>	<b>Introduction.....</b>	<b>66</b>
<b>4.3</b>	<b>Methods .....</b>	<b>68</b>
4.3.1	Phylogenic and repeat conservation analysis.....	68
4.3.2	Cas13a protein expression and purification.....	70
4.3.3	In-vitro RNA transcription.....	70
4.3.4	Radiolabeled ssRNA nuclease assays .....	72

4.3.5	Fluorescent ssRNA nuclease assays .....	72
4.3.6	crRNA filter-binding assays.....	73
<b>4.4</b>	<b>Results.....</b>	<b>73</b>
4.4.1	Most Cas13a homologs possess pre-crRNA processing activity .....	73
4.4.2	A conserved crRNA maturation center within most Cas13a enzymes .....	75
4.4.3	Cas13a enzymes initiate ssRNA cleavage at either uridines or adenosines .....	78
4.4.4	CRISPR repeat sequence determines non-cognate pre-crRNA processing.....	81
4.4.5	Two subfamilies of functionally orthogonal Cas13a enzymes .....	84
4.4.6	Pre-crRNA processing enhances targeting efficiencies within the context of a CRISPR array .....	86
<b>4.5</b>	<b>Discussion.....</b>	<b>90</b>
<b>Chapter 5: Fitting Type VI systems into the CRISPR toolbox</b>		
5.1	Discovery of Type VI systems .....	93
5.2	Functional studies of Type VI systems .....	93
5.3	Fitting Type VI CRISPR-Cas systems into the prokaryotic defense arsenal .....	94
5.4	Open Questions in Type VI CRISPR-Cas systems .....	95
<b>Appendix I: RNA-programmed genome editing in human cells</b>		
<b>I.1</b>	<b>Summary .....</b>	<b>98</b>
<b>I.2</b>	<b>Introduction.....</b>	<b>98</b>
<b>I.3</b>	<b>Methods .....</b>	<b>99</b>
I.3.1	Plasmid design and construction .....	99
I.3.2	Cell culture conditions and DNA transfections.....	99
I.3.3	Western blot analysis of Cas9 expression .....	99
I.3.4	Surveyor assay .....	100
I.3.5	In vitro transcription.....	100
I.3.6	Northern blot analysis .....	100
I.3.7	In vitro cleavage assay .....	100
<b>I.4</b>	<b>Results .....</b>	<b>101</b>
<b>I.5</b>	<b>Discussion.....</b>	<b>105</b>

## List of Figures

Figure 1.1 Overview of bacterial defense mechanisms.....	2
Figure 1.2 Classification scheme of CRISPR-Cas systems .....	5
Figure 1.3 Structural comparison of Class 1 effector complexes .....	6
Figure 1.4 Classification of Class 2 CRISPR effector Types and Subtypes.....	7
Figure 1.5 Overview of general CRISPR-Cas pathway.....	8
Figure 1.6 Proposed mechanism of new spacer integration in <i>E. coli</i> .....	10
Figure 1.7 Overview of the crRNA biogenesis pathways .....	12
Figure 1.8 Overview of Class 2 effector interference mechanisms .....	15
Figure 2.1 Schematic of ssRNA targeting strategy using a PAMmer.....	19
Figure 2.2 RNA-guided Cas9 cleaves ssRNA targets in the presence of a short PAM-presenting DNA oligonucleotide (PAMmer). .....	28
Figure 2.3 Quantified data for cleavage of ssRNA by Cas9-gRNA in the presence of a 19-nt PAMmer. ....	29
Figure 2.4 dCas9-gRNA binds ssRNA targets with high affinity in the presence of PAMmers. ....	30
Figure 2.5 RNA-guided Cas9 can target non-PAM sites on ssRNA and isolate GAPDH mRNA from HeLa cells in a tagless manner. ....	31
Figure 2.6 RNA-guided Cas9 can utilize chemically modified PAMmers. ....	32
Figure 2.7 pri-miRNA target design.....	33
Figure 2.8 qPCR validation of pri-miRNA pulldown.....	34
Figure 2.9 Radiolabeled reporter validation of pri-miR17-92 and pri-miR7 targets. ....	35
Figure 2.10 Silver-stained SDS-PAGE of Pulldown elutions.....	36
Figure 2.11 Summary of isolated proteins by RCas9 pulldown of pri-miR7, pri- miR17-92 and GAPDH RNAs .....	36
Figure 2.12 Potential applications of RCas9 for untagged transcript analysis, detection, and manipulation. ....	40
Figure 3.1 Phylogenetic tree of C2c2 family.....	43
Figure 3.2 Alignment of protein sequences from three C2c2 homologs .....	43
Figure 3.3 Purification and Production of C2c2.....	44
Figure 3.4 C2c2 proteins process precursor crRNA transcripts to generate mature crRNAs .....	49
Figure 3.5 Mapping of pre-crRNA processing by C2c2 in vitro and in vivo. ....	51
Figure 3.6 Further investigations into the substrate requirements and mechanism of pre-crRNA processing by C2c2.....	52
Figure 3.7 LbuC2c2 mediated crRNA biogenesis depends on both structure and sequence of CRISPR repeats.....	53
Figure 3.8 LbuC2c2 catalyzes guide-dependent ssRNA degradation on cis and trans targets.....	54
Figure 3.9 LbuC2c2 ssRNA target cleavage site mapping.....	56

Figure 3.10 Dependence of RNA targeting on crRNA variants, temperature and point mutations .....	58
Figure 3.11 LbuC2c2 contains two distinct ribonuclease activities .....	59
Figure 3.12 Binding data for LbuC2c2 to mature crRNA and target ssRNA .....	60
Figure 3.13 C2c2 provides sensitive detection of transcripts in complex mixtures	62
Figure 3.14 RNase detection assay $\lambda$ 2-ssRNA time-course and background RNA cleavage .....	63
Figure 4.1 CRISPR loci architecture for Cas13a homologs used in this study .....	68
Figure 4.2 Pre-crRNA processing is broadly conserved within the Cas13a protein family .....	74
Figure 4.3 Expanded CRISPR repeat structures and HheCas13a testing .....	75
Figure 4.4 Identification of residues important for pre-crRNA cleavage by LbuCas13a .....	76
Figure 4.5 Members of the Cas13a protein family cleave ssRNA with a range of efficiencies .....	79
Figure 4.6 trans-ssRNA cleavage by Cas13a homologs .....	80
Figure 4.7 crRNA exchangeability within the Cas13a family .....	82
Figure 4.8 Functional validation of orthogonal Cas13a subfamilies for RNA detection .....	85
Figure 4.9 trans-cleavage by LbuCas13a point mutants in regions implicated in crRNA-processing .....	86
Figure 4.10 Deciphering the role of crRNA array processing for LbuCas13a .....	87
Figure 4.11 crRNA processing inactive mutant R1079A/K1080A retains similar crRNA binding affinity and does not process a pre-crRNA array ...	88
Figure 4.12 A revised model for Type VI-A CRISPR-Cas system .....	89
Figure I.1 Co-expression of Cas9 and guide RNA in human cells generates double-strand DNA breaks at the target locus .....	102
Figure I.2 Cell lysates contain active Cas9:sgRNA and support site-specific DNA cleavage .....	104
Figure I.3 3' extension of sgRNA constructs enhances site-specific NHEJ-mediated mutagenesis .....	105

## List of Tables

Table 2.1 $\lambda$ oligonucleotide sequences .....	21
Table 2.2 Oligonucleotides used in the GAPDH mRNA pull-down experiment.....	22
Table 2.3 Table of oligos used for pri-miRNA studies .....	25
Table 2.4 Enrichment Scores for positive controls in all samples .....	38
Table 2.5 Top protein binding candidates for pri-miR7 .....	38
Table 2.6 Top protein binding candidates for pri-miR17-92 .....	39
Table 2.7 Top protein binding candidates for GAPDH .....	39
Table 3.1 Oligonucleotides used in study.....	46
Table 4.1 List of Cas13a homologs used in this study .....	69
Table 4.2 List of RNA oligonucleotides used in this study.....	71
Table 4.3 Homopolymer cleavage percentages for six Cas13a homologs with associated errors. ....	79
Table 4.4 pre-crRNA processing cleavage percentages for ten Cas13a homologs with associated errors .....	83
Table 4.5 Normalized initial trans-target rates for five Cas13a homologs directed by pre-crRNAs with associated errors. ....	84
Table 4.6 Normalized initial trans-target rates for five Cas13a homologs directed by mature crRNAs with associated errors.....	85

# Chapter 1

## Introduction to CRISPR-Cas Immune Systems

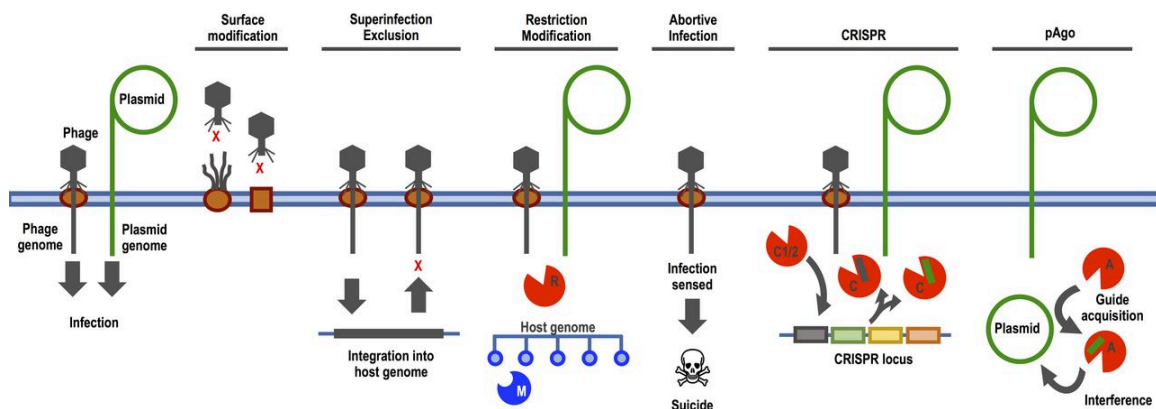


## 1.1 Overview of bacterial defense systems

The competition for survival within microbial communities drives the evolution of bacteria and archaea (Dy et al., 2014). Often overlooked, phages are a key component of these communities with a global population estimated over  $10^{30}$ , with approximately  $10^{25}$  phage infections occurring per second across the Earth (Chibani-Chennoufi et al., 2004). The rapid lifecycles of both phage and their hosts has driven the evolution of a plethora of bacterial immune pathways and conversely, interference mechanisms by the phage to counter these defenses. Beyond intriguing biochemical mechanisms, many of these systems contain robust enzymes that have been exploited by bioengineers for the past 50 years (Ghafourian et al., 2014; Snyder, 1995; Sternberg and Doudna, 2015). From restriction modification systems to the current CRISPR-craze, the field of bacterial immunology has been ripe for biotechnologically applications in addition to revealing the inherent beauty of biological systems.

### 1.1.1 Innate strategies to evade phage infection

While the bulk of this work will center on the adaptive CRISPR-Cas immune systems, first we will address other alternative mechanisms prokaryotic cells can use to combat phage infections. Host cells have many redundant pathways that can block phage infections at all stages of the phage life cycle (Fig 1.1) (Dy et al., 2014; van Houte et al., 2016). Simple measures can reduce the probability of phage infection like extracellular polymers to act as physical barriers or secreted enzymes that degrade phage proteins in the extracellular space (Labrie et al., 2010; Seed, 2015). Other more complex mechanisms can block genome injection by specific phage classes conferring immunity typically encoded within the host's incorporated prophages (Labrie et al., 2010; Seed, 2015).



**Figure 1.1 Overview of bacterial defense mechanisms**

Summary of bacterial immune pathways in relationship to the phage life cycle. Letters indicate protein components involved in the immune mechanism (M, methylase; R, restriction enzyme; C1/2, Cas1 and Cas2; C, Cas effector-nuclease complex; A, prokaryotic Argonaute enzyme). Adapted from van Houte et al., 2016.

If phage infections cannot be stopped, a variety of intracellular measures may be utilized to block phage propagation. Many of these abortive infection systems or phage genome targeting systems are very redundant leading to wide diversity, but also leading

to difficulty in determining the precise role of these pathways (Mruk and Kobayashi, 2014; Vasu and Nagaraja, 2013). Abortive infection systems are generalized suicide pathways to reduce phage propagation within a large ecosystem (Dy et al., 2014). Ideally in these systems the infected host bacteria does not survive, altruistically protecting its community. The first abortive infection system to be characterized was the *rexAB* system, which will depolarize the host membrane upon recognition of a specific phage protein-DNA complex (Snyder, 1995). The subsequent decrease in intracellular ATP leads to bacterial death and overall protection of the microbial community. Abortive infection systems are often encoded within mobile genetic elements and can be regulated by other phage defense systems including toxin-antitoxins (Fineran et al., 2009).

Toxin-Antitoxin (TA) systems are another class of altruistic innate defense pathways (Gerdes et al., 2005; Ghafourian et al., 2014). Originally determined to be a system to addict host cells to plasmids (Ogura and Hiraga, 1983), a large diversity of these systems have been subsequently studied and determined to be involved in regulation of stress response including phage defense (Ghafourian et al., 2014; Leplae et al., 2011; Yamaguchi and Inouye, 2011). Generally these pathways are two component systems where one part contains a cytotoxic activity, termed the toxin. The activity of this toxin is negated by the presence of the second antitoxin component, either protein or RNA, which is extremely labile. Under periods of stress the labile antitoxin is quickly degraded releasing the toxin protein (Yamaguchi et al., 2011). Many of the toxin proteins are nucleases, although other protein classes can be toxins including small hydrophobic peptides that puncture membranes and cytoskeletal binding proteins (Page and Peti, 2016). The nucleases of TA systems are divided into 2 general classes, ribosome-independent and ribosome-dependent mRNA interferases (Yamaguchi and Inouye, 2011). The *MazE-MazF* ribosome-independent TA system encodes an ACA-specific ribonuclease that will degrade most cellular transcripts and protects *E. coli* against P1 phage infection (Hazan and Engelberg-Kulka, 2004). Another well studied example of TA systems defending against phage is the *ToxI-ToxN* system that activates abortive infection upon degradation of the antitoxin RNA *ToxI* (Fineran et al., 2009). Through their responsiveness to cellular stressors, TA systems defend against phage with a variety of mechanisms and promote bacterial persistence (Harms et al., 2016).

Beyond altruistic cell suicide pathways, restriction modification (RM) systems are another prevalent bacterial defense mechanism. RM systems are a very large, diverse family of proteins with a general mechanism that is dependent on modification of the host genome at specific sequences (Mruk and Kobayashi, 2014; Vasu and Nagaraja, 2013). They were the first bacterial immune systems to be studied in the 1960s, and enabled the field of molecular biology to develop (Arber, 1965; Snyder, 1995). Similar to TA systems, RM systems generally are composed of two enzyme components that work in tandem to protect the host (Mruk and Kobayashi, 2014; Vasu and Nagaraja, 2013). The first enzyme is a modification enzyme, that will transfer generally a methyl group to the host genomic DNA, marking it as 'self'. This process is generally coupled to genome replication ensuring proper marking of the newly synthesized strand. The second enzyme is a nuclease that will cleave DNA strands, usually at specific sequences, that do not contain the precise modification placed by the first enzyme.

These enzyme pairs have proven invaluable to the field of molecular biology, enabling scientists to precisely cut and paste nucleic acids in a predictable manner (Snyder, 1995).

### **1.1.2 Adaptive strategies to combat phage infection**

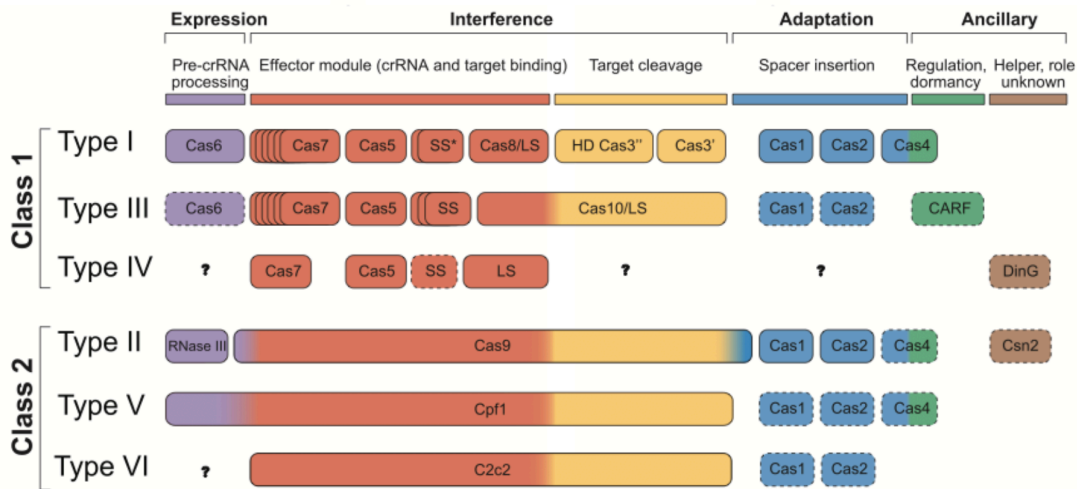
Adaptive immunity within prokaryotes is a relatively young field. The CRISPR-Cas (clustered regular interspersed short palindromic repeats - CRISPR associated) system was first demonstrated only ten years ago to protect *Streptococcus thermophilus* (*S. thermophilus*) against phage infection (Barrangou et al., 2007). This system is defined by the CRISPR genomic locus, a series of variable spacer sequences of foreign origin separated by directed repeats, and the nearby *cas* gene operon(s) encoding essential pathway proteins (van der Oost et al., 2014; Wright et al., 2016). CRISPR-Cas systems reside within a large percentage of the prokaryotic kingdom (Burstein et al., 2016; Makarova et al., 2011) and have been shown to defend against both phage and mobile genetic elements (Garneau et al., 2010; Marraffini and Sontheimer, 2008). These systems show a large amount of diversity and have been subsequently divided into two distinct classes, six types, and nineteen subtypes (Makarova et al., 2011; 2015; Shmakov et al., 2017). These systems will be covered extensively in the rest of this work.

Prokaryotic argonaute (pAgo) proteins mediate the only other proposed adaptive prokaryotic defense pathway. This fledging field originated from the discovery in 2009 that Argonaute-PIWI proteins were not only wide spread within the prokaryotic kingdom but also within genomic neighborhoods enriched for known defense pathway genes (Makarova et al., 2009). Limited experimental evidence has supported this hypothesis. *In vitro*, pAgo proteins bind both DNA and RNA guides to direct cleavage against ssDNA and ssRNA substrates (Kaya et al., 2016; Olovnikov et al., 2013; Swarts et al., 2015; 2014). *In vivo*, pAgo is reported to reduce plasmid transformation efficiency and pAgo deletion mutants retain higher plasmid expression levels (Blesa et al., 2015; Olovnikov et al., 2013; Swarts et al., 2014). Recent work indicates that some pAgos contain a second nuclease activity, enabling guide generation through a non-specific cleavage of dsDNA (Swarts et al., 2017; Zander et al., 2017). Many key mechanistic questions remain open in this field, including the mechanisms of self versus non-self discrimination and acquisition of guides from foreign nucleic acids.

## **1.2 CRISPR-Cas system classification**

Evolutionary pressure to protect against phage-induced lysis and rampant horizontal gene transfer has created a wide range of CRISPR systems (Fig 1.2) (Mohanraju et al., 2016). The gene operons and CRISPR loci have been classified through their evolutionary relationships in a 'polythetic' classification scheme to simplify this diversity (Makarova et al., 2011). These classifications generally rely on a signature or hallmark Cas ORF (open reading frame), the Cas gene composition and repeat structure. Overall, CRISPR-Cas operons are divided into 2 general classes that denote the composition of their targeting complexes either as a multi-component assembly (Class 1) or as a single polypeptide (Class 2). Within each class, there are operon types and subtypes, which correspond to the evolutionary relationship of the Cas genes. In

addition, the genes responsible for each of the distinct pathway steps are referred to as modules.



**Figure 1.2 Classification scheme of CRISPR-Cas systems**

Overview of CRISPR types is given above. General gene names are noted within the blocks representing the ORFs. Functions of the genes are aligned as noted in the columns. Adapted from Mohanraju et al., 2016.

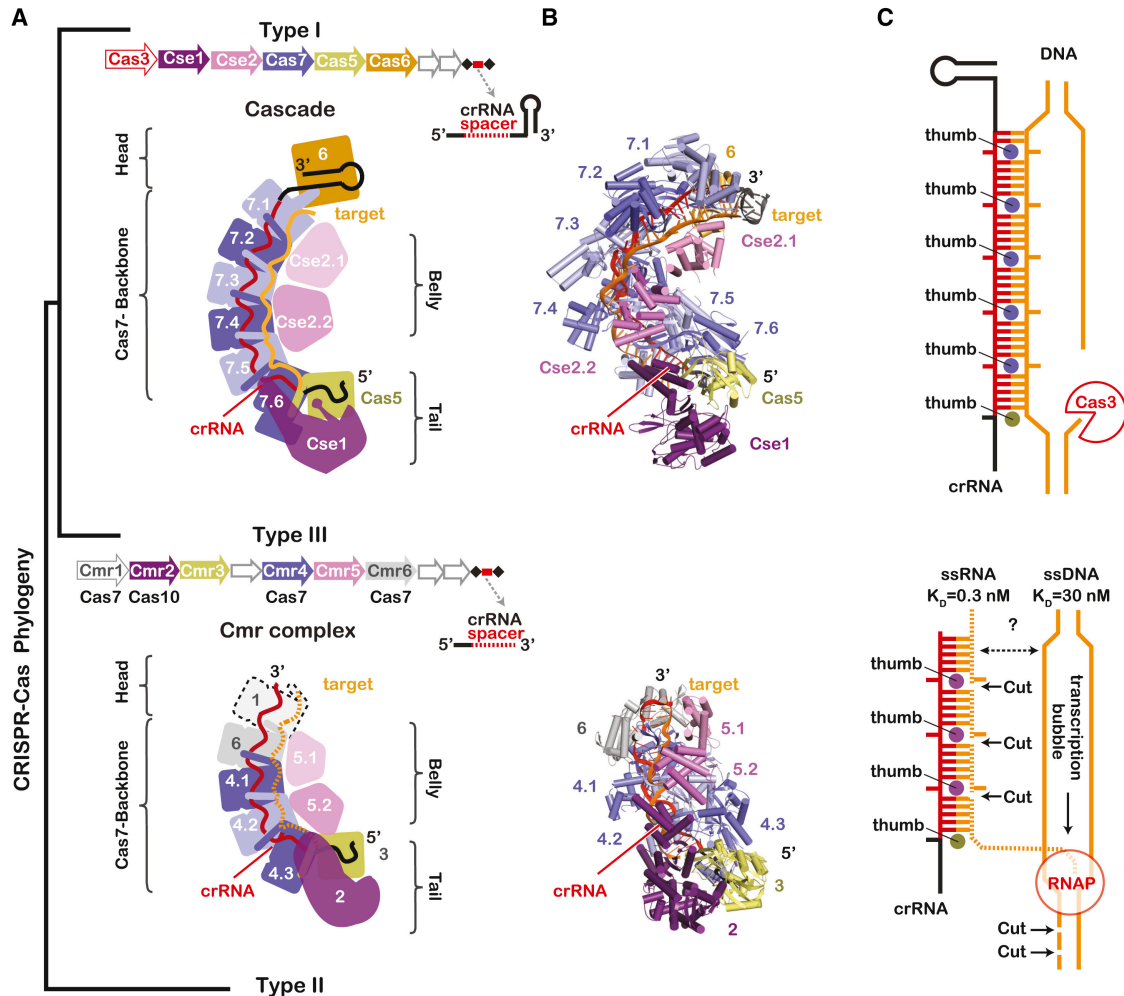
### 1.2.1 Class 1 systems

Class 1 CRISPR-Cas systems maintain a diverse set of Cas proteins, which assemble into large complexes centered around the crRNAs (van der Oost et al., 2014; Wright et al., 2016). Overall these systems are the most abundant CRISPR types, and are broken up into many subtypes depending on their exact gene composition (Burstein et al., 2016). Type I and Type III structural studies have revealed a conserved architecture of multiple repeating backbone subunits anchored generally by the processing enzyme, repeat binding proteins, and the large and small subunits (Fig 1.3) (Hochstrasser et al., 2016; Jackson et al., 2014; Mulepati et al., 2014; Staals et al., 2013; 2014; Taylor et al., 2015; van der Oost et al., 2014; Zhao et al., 2014).

The Type I targeting complex formation is initiated by specific binding and cleavage of Cas6 to the crRNA hairpin (Carte et al., 2008; Haurwitz et al., 2010). The other Cas proteins assemble onto the crRNA backbone, forming a large ~400 kDa complex referred to as Cascade (CRISPR-associated complex for antiviral defense). For *E. coli* Type I-E Cascade, the oligomerized backbone subunit is Cas7 with the crRNA ends capped with Cas5 at the 5' end and Cas6 at the 3' stem loop. Cse1 is the large subunit while a dimer of Cse2, or small subunits hug the belly of the complex (Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014).

Type III complexes form a similar structure, although the order of assembly is less established. The large subunit in Type III systems, Cas10 (Csm1 or Cmr2) is the signature protein at the base of the complex. The crRNA backbone subunits are Csm3 or Cmr4 respectively, with Csm4 or Cmr3 capping the 5' end and Csm5 or Cmr1/Cmr6 bound at the 3' end. Csm2 or Cmr5 are the small subunits stabilizing the belly of the complex (Staals et al., 2013; 2014; Taylor et al., 2015). Despite their distinct

phylogenetic relationships, Type I and Type III share a common scaffold, suggested a common shared ancestor (Jackson and Wiedenheft, 2015).



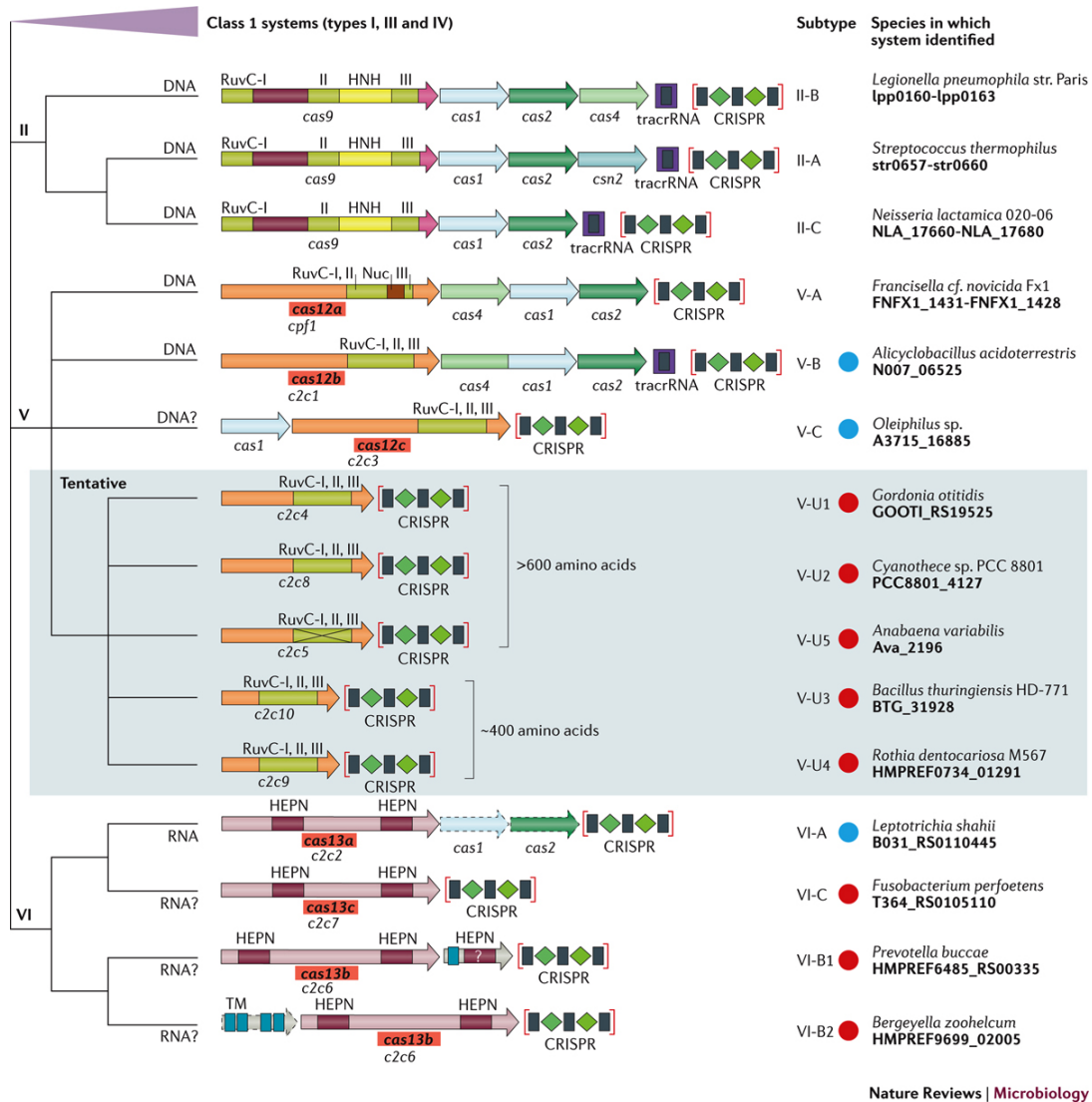
**Figure 1.3 Structural comparison of Class 1 effector complexes**

Schematics of Class 1 effector complexes is given above. Type I systems are depicted above, with type III below. (A and B) Generalized Cas gene operon is color-coded for the structural components and the multiple copies of a given gene are indicated but decimal notation in the schematics. (C) Diagrams of the crRNA:dsDNA duplex summing up the current understanding of biochemical mechanism of these complexes. Adapted from Jackson and Wiedenheft, 2015.

### 1.2.2 Class 2 systems

In contrast to Class 1 interference complexes that are composed of many smaller proteins, Class 2 effectors comprise only one large protein (Mohanraju et al., 2016). Cas9, a Cas enzyme well known for its functions beyond CRISPR biology, is an example of the Class 2, Type II CRISPR single effector. Generally Class 2 effectors are 900-1500 amino acids in length, and are located next to Cas1, Cas2 and sometimes other ancillary Cas genes (Fig 1.4). This class has been recently expanded to include

recently discovered systems from Type V and Type VI (Makarova et al., 2015; Shmakov et al., 2017; 2015; Zetsche et al., 2015). These systems are observed much less frequently than Type I and Type III CRISPR systems throughout the prokaryotic kingdom, and within Archaea there are very few examples. Even within Class 2 operons, Type II exist in only 8% of published genomes, while Type V and Type VI are more than 1 order of magnitude lower (Shmakov et al., 2017).



**Figure 1.4 Classification of Class 2 CRISPR effector Types and Subtypes.**

Generalized phylogenetic tree of Class 2 effectors with cartoon Cas operons. Predicted catalytic domains are noted for all effectors in addition to the standard gene synteny for each subtype. Previous gene names are listed in italics under the cartoon, with updated name highlighted in red. To the left, example genomes are listed for each subtype. Adapted from Shmakov et al, 2017.

Type II systems were the original CRISPR system to be studied in 2007 by scientists at Danisco, a yogurt company (Barrangou et al., 2007). Subsequently it has

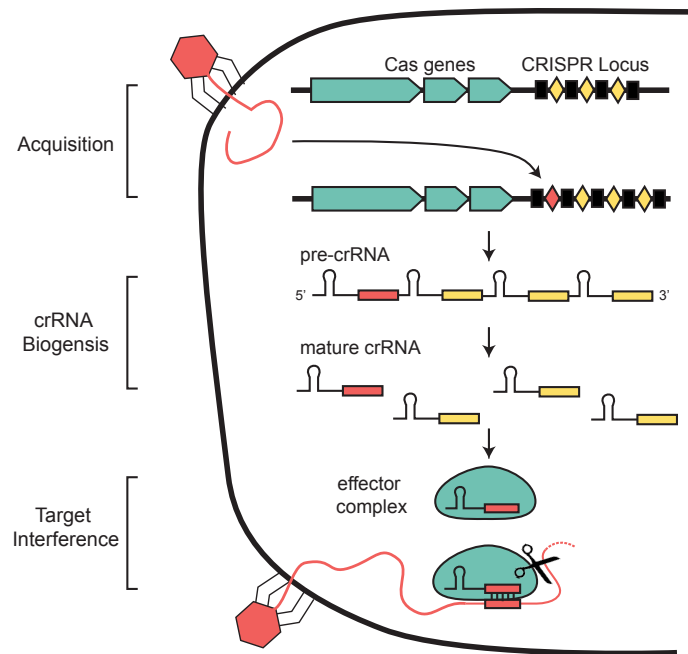


been determined that Cas9 contains two nuclease domains, a HNH domain and split RuvC that are responsible for target cleavage (Gasiunas et al., 2012; Jinek et al., 2012). Three Type II subtypes exist, primarily defined by the adjacent Cas genes, Cas2, Cas4 or the lack of both. Type II-C Cas9 proteins are the most distinct, tending to be smaller than Cas9's originating from Type II-A or II-B systems.

Similar to Type II systems, Type V systems contain the hallmark protein Cas12 (formerly Cpf1), which unlike Cas9 only contains 1 predicted nuclease domain (Shmakov et al., 2017). Since this protein cleaves both strands of its dsDNA target, a second region of the protein is suspected to have cleavage activity. Current studies are in disagreement over the likelihood and location of this potential domain (Dong et al., 2016; Gao et al., 2016; Yamano et al., 2016).

The most recent additions to the CRISPR universe are Type VI systems, which are quite distinct from other Class 2 effectors. These effector proteins were identified by searching for large ORFs next to CRISPR loci within published genomes. Cas13 (formerly C2c2) proteins were missed in previous bioinformatics searches that required the presence of Cas1 (Shmakov et al., 2015; 2017; Smargon et al., 2017). Domain mapping only finds two predicted HEPN motifs within this diverse protein family, leading to the hypothesis that these systems target RNA instead of DNA (Shmakov et al., 2015).

### 1.3 Three major steps of the CRISPR-Cas pathway



**Figure 1.5 Overview of general CRISPR-Cas pathway**

The three basic steps of CRISPR systems are diagrammed in the figure above. Cas genes noted in teal, while repeats are black squares with the spacers as yellow diamonds. A new spacer is added to the array from an invading phage (red). Pathway components are expressed during the second step, and guides are loaded into effector complexes for targeting of spacer complementary sequences.

Mechanistic studies of the CRISPR-Cas pathway have defined three major steps required for adaptive immunity: (1) integration of new spacers into the host genome at the CRISPR locus, (2) Cas protein expression and CRISPR RNA (crRNA) biogenesis, and (3) effector complex assembly and target interference (Carte et al., 2008; Kunin et al., 2007; Wright et al., 2016). These steps are diagrammed in Fig. 1.5. Other essential terminologies for CRISPR-Cas systems include protospacer and protospacer adjacent motif (PAM). Protospacers are the spacer matching sequence within the foreign DNA and are the substrates for new spacer acquisition. Self versus non-self discrimination is generally established through the PAM sequence, which is located flanking the protospacer. At the CRISPR locus within the host genome, these sequences will not appear, limiting interference against the host chromosome.

### 1.3.1 Acquisition of new spacers

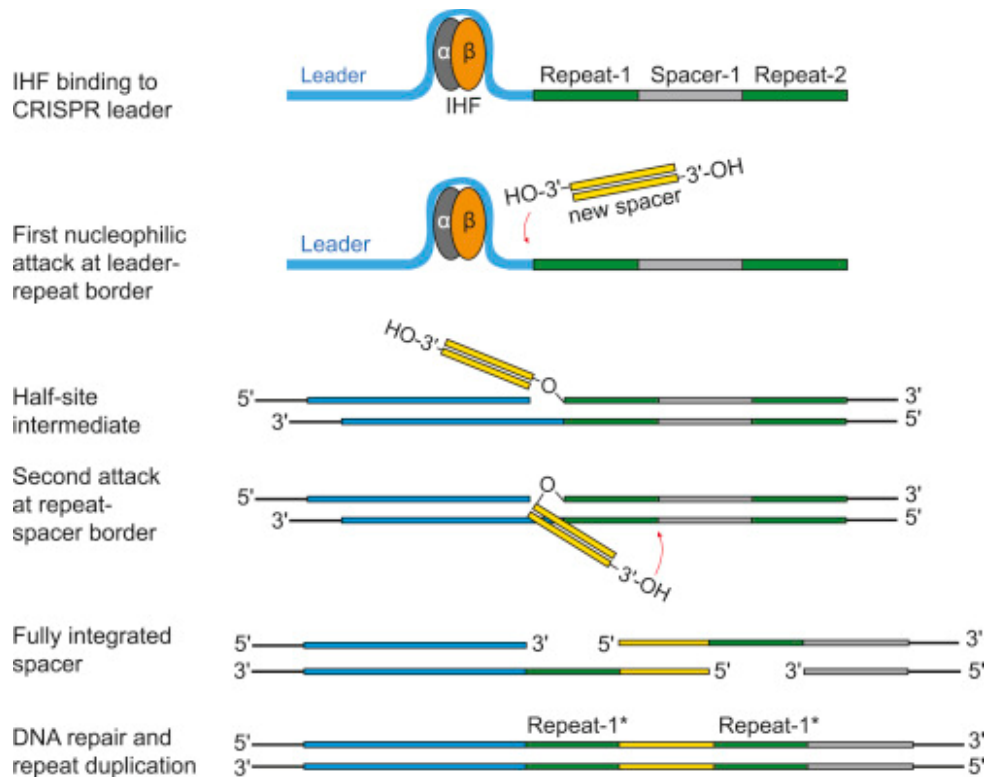
In order to protect a host during a naïve phage infection, the CRISPR-Cas system must integrate a new spacer into the CRISPR loci through a process referred to as spacer acquisition or adaptation (Amitai and Sorek, 2016; Heler et al., 2014; Sternberg et al., 2016). Expansion of the CRISPR array in response to phage infection was established by the first landmark study that established the field of CRISPR biology in 2007 (Barrangou et al., 2007). But, it was not until 2012 that the mechanism of these steps began to be unraveled by multiple groups working on the *E. coli* CRISPR system (Datsenko et al., 2012; Swarts et al., 2012; Yosef et al., 2012). CRISPR acquisition can be triggered in a laboratory setting in response to either transformed plasmid (Diez-Villaseñor et al., 2013; Erdmann and Garrett, 2012; Swarts et al., 2012; Yosef et al., 2012) or bacteriophage infections (Barrangou et al., 2007; Datsenko et al., 2012). Bacteriophage infection mimics the native context of the CRISPR pathway and has been successfully used in many species including *Streptococcus thermophilus* (Barrangou et al., 2007; Horvath et al., 2008; Paez-Espino et al., 2013), *Sulfolobus solfataricus* (Erdmann et al., 2013a), *Pseudomonas aeruginosa* (Cady et al., 2012), and *E. coli* (Datsenko et al., 2012; Sashital et al., 2012).

The two Cas ORFs found commonly across all CRISPR-Cas subtypes, Cas1 and Cas2, were established to be necessary for new spacer acquisition (Yosef et al., 2012). Cas1/Cas2 form a stable complex and complex formation is necessary for acquisition *in vivo* (Nunez et al., 2014). Initial *in vitro* biochemical studies of Cas1 and Cas2 have characterized both as metal-dependent endonucleases, with activity against a variety of substrates include ssDNA, dsDNA, ssRNA and Holiday junctions (Babu et al., 2011; Beloglazova et al., 2008; Nam et al., 2012a). Recent clarity in the field was gained with the establishment of *in vitro* integration assays suggesting the preferred substrate of the Cas1/Cas2 complex is dsDNA with 3' overhangs (Nunez et al., 2015a; 2015b; Rollie et al., 2015; Wright and Doudna, 2016). The 3' OH of the incoming spacer is assumed to engage in nucleophilic attack on one strand of the CRISPR repeat as diagrammed in Fig 1.6. These *in vitro* mechanisms are supported by the capture of integration intermediates from *in vivo* cultures (Arslan et al., 2014).

While the biochemical mechanism of spacer integration has been well studied, how specific protospacers are selected for integration into the CRISPR array is less established. Few sequence preferences are observed across the protospacer length,



with constant a GC content for both Type I and Type II systems (Díez-Villaseñor et al., 2013; Paez-Espino et al., 2013; Savitskaya et al., 2013; Shah et al., 2013; Swarts et al., 2012). Some systems display a preference for a flanking PAM sequence, although the proposed mechanisms vary. For Type I systems PAM preference during naïve acquisition is dictated by Cas1/Cas2 (Wang et al., 2015; Yosef et al., 2013). In contrast, for Type II systems the interference protein Cas9 influences PAM and protospacer choice (Heler et al., 2015; 2016; Wei et al., 2015b) and structural evidence suggests a super complex between Csn2, Cas1, and Cas2 may also play an important role in Type II acquisition (Ka et al., 2016).



**Figure 1.6 Proposed mechanism of new spacer integration in *E. coli***

IHF (gray and orange) binds to a specific site in the leader sequence and induces bending of the leader sequence. Cas1-Cas2 bound to a protospacer recognizes the bent leader-repeat border sequence and catalyzes the first nucleophilic integration attack. The second integration event occurs via a ruler mechanism 28 bp downstream at the repeat-spacer-1 border. Following DNA repair, the new spacer is integrated and duplicates repeat-1\*. Adapted from Nunez et al., 2016.

In addition to Cas1 and Cas2, *cis*-acting sequences within the CRISPR locus are vital for spacer acquisition. Upstream of the repeat array is a conserved region, termed the leader, which determines the polarity of the locus (Fineran and Charpentier, 2012; Yosef et al., 2012). Incoming spacers are consistently integrated at the leader-proximal end of the CRISPR array (Horvath et al., 2008). Truncation analysis of CRISPR loci from Type I and Type II systems established the *cis*-acting regions as essential for *in vivo* spacer acquisition (Díez-Villaseñor et al., 2013; Moch et al., 2016; Wang et al.,

2016; Wei et al., 2015a; Yosef et al., 2012). It has recently been reported that a host protein, integration host factor (IHF) binds to this region of the leader, remodeling the DNA topology to enable Cas1/Cas2 to catalyze the nucleophilic attack (Nunez et al., 2016; Yoganand et al., 2017).

The mechanism for discriminating between self and non-self protospacers during spacer acquisition is still an open question. Analysis of self-targeting spacers in Cas1/Cas2 overexpression strains suggests that the double strand break repair complex RecBCD may play a role in generating protospacer fragments (Levy et al., 2015). Host chromosomal sequences have a higher density of Chi sites that pause RecBCD-mediated degradation, limiting the protospacer fragment production for Cas1/Cas2. Target interference complexes have also been implicated in generating potential protospacers (Künne et al., 2016; Redding et al., 2015) potentially increasing the frequency of new spacers originating from foreign DNA targets.

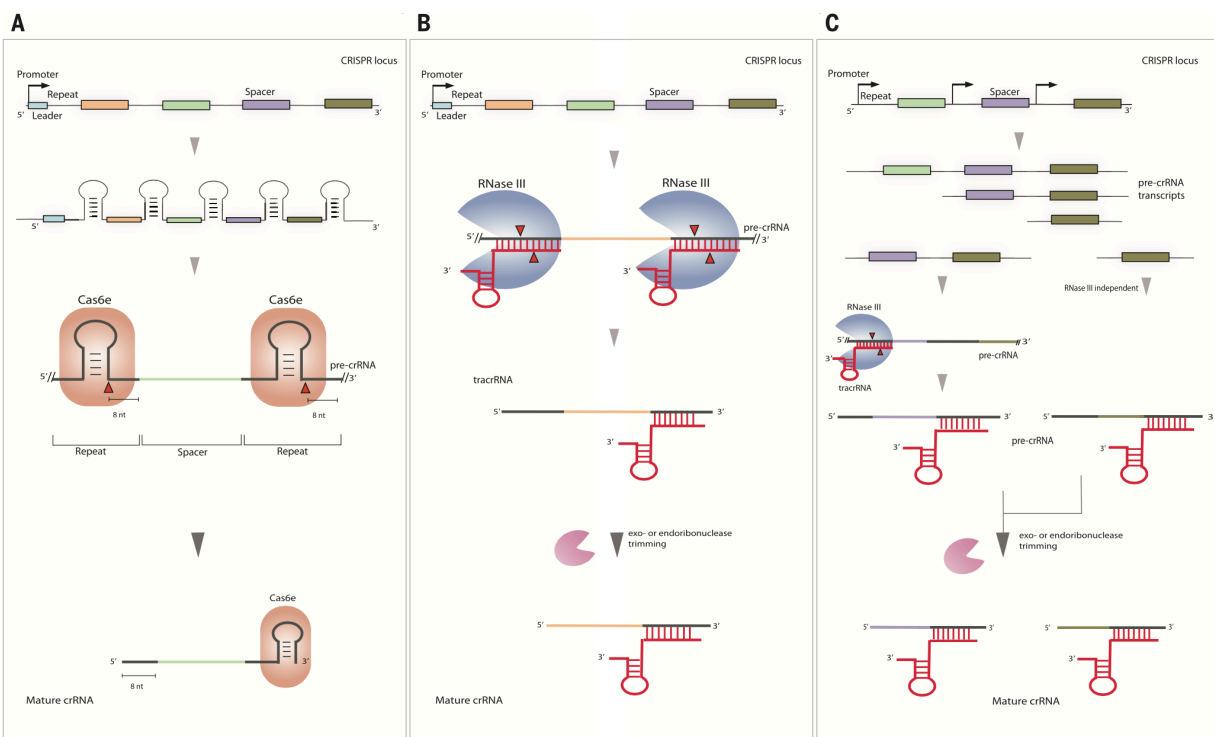
Another paradigm of CRISPR-Cas acquisition is the feedback between stalled interference complexes and upregulation of new spacer integration termed primed acquisition (Amitai and Sorek, 2016; Heler et al., 2014; Sternberg et al., 2016). Priming was first observed during phage infection of *E. coli* and the propensity for multiple targeting spacers to be integrated in response to escapee phage mutations (Datsenko et al., 2012). Further mechanistic studies revealed that stalled Cascade complexes bound to degenerate target sites trigger enhanced spacer acquisition (Blosser et al., 2015; Fineran et al., 2014; Redding et al., 2015; Xue et al., 2015). Priming has been observed in Type I-B (Li et al., 2014), Type I-E (Datsenko et al., 2012; Shmakov et al., 2014), and Type I-F (Richter et al., 2014) suggesting this may be a widespread mechanism in Type I systems. Type II systems also have a feedback loop between interference machinery and acquisition, but interaction is required for all *in vivo* acquisition events (Heler et al., 2015).

Little is known about acquisition for Type III systems, as the systems tend to be genetically intractable, limiting the field to observational acquisition studies. Initial studies on *S. solfataricus* (Erdmann and Garrett, 2012; Erdmann et al., 2013b; Shah et al., 2013) suggest that Type III CRISPR arrays are dynamic, but the coexistence of Type III with Type I operons has muddied the conclusions. Crosstalk by these arrays has been observed at both the crRNA biogenesis (Deng et al., 2013) and interferences (Zhang et al., 2012) steps, raising the question if crosstalk also occurs during acquisition. Due to the dearth of mechanistic studies on Type III acquisition, it is unclear what the precise protospacer substrate of these ssRNA-targeting systems should be. Some loci contain Cas1 fusions to reverse transcriptases (RT) (Makarova et al., 2011; Shmakov et al., 2015) suggesting a role for RNA during spacer acquisition. One of these Cas1-RT fusions has been shown to ligate RNA segments into the CRISPR array and mediate *in vivo* spacer acquisition, but it remains unclear if this is a widespread mechanism of RNA spacer acquisition (Silas et al., 2016).

### **1.3.2 Biogenesis of pathway components**

Transcriptional control and regulation of these systems is relatively understudied within the CRISPR systems. In Type II systems higher transcription and protein production is observed upon phage infection of *S. thermophilus* (Agari et al., 2010; Young et al., 2012) although the mechanism of induction is unknown. For Type I

systems, the regulation appears to be highly variable. In *E. coli* Type I-E, the general transcriptional silencers H-NS, and CRP suppresses Cas gene expression (Medina-Aparicio et al., 2011; Pul et al., 2010; Westra et al., 2010; Yang et al., 2014) while cell envelop stress can trigger expression (Perez-Rodriguez et al., 2010). For Type I-F systems in *Pectobacterium atrosepticum* Cas gene expression is linked to glucose usage through CRP-cAMP based regulation (Patterson et al., 2015). More recently quorum sensing in *Pseudomonas aeruginosa* has been implicated in transcriptional control of their Type I-F CRISPR-Cas system (Høyland-Kroghsbo et al., 2017; Patterson et al., 2016). Studies in *S. solfataricus* have observed that the CRISPR locus is constitutively expressed, perhaps reflecting the continual exposure of these hosts to phage within their native environment (Deng et al., 2012; Lillestøl et al., 2009). Overall most species demonstrate system specific transcriptional regulation to activate their CRISPR-Cas systems.



### Figure 1.7 Overview of the crRNA biogenesis pathways

(A) Type I and Type III systems generally use an endoribonuclease Cas6 for array processing. Cleavage events are noted by red triangles. (B) Type II processing generally occurs through an RNase III mediated pathway that cleaves the dsRNA duplex between the crRNA and tracrRNA. (C) For some Type II systems, individual promoters exist within the repeat sequences enabling mature crRNA production from the initial transcription product. Adapted from Mohanraju et al., 2016.

Upon expression, the CRISPR array is generally expressed as a long non-coding transcript that must be processed into mature guides (Brouns et al., 2008). For Type I and Type III systems, the Cas6 gene family encodes RNA endonucleases that are

responsible for cleaving within the repeat sequence to generate the mature guide (Fig 1.7) (Carte et al., 2008; Charpentier et al., 2015; Hochstrasser and Doudna, 2015). This enzyme binds the repeat hairpin in a sequence and structure dependent manner with very high affinity (Haurwitz et al., 2010; 2012; Sternberg et al., 2012). After cleavage, the remaining repeat sequences are referred to as the handles. The 5' handle is generally 8 nts in Type I and Type III systems (Hochstrasser and Doudna, 2015). In contrast, the 3' handle displays much more heterogeneity across the different subtypes. In some subtypes, other host nucleases non-specifically trim the 3' end, presumably using the protection of the interference complex as a template (Hale et al., 2008; Scholz et al., 2013; Walker et al., 2016).

For Type II systems, a distinct mechanism is used to liberate mature crRNA's from the array transcript. A second RNA is encoded by the locus, the trans-activating RNA (tracrRNA) which contains significant complementarity to the CRISPR repeat sequence (Deltcheva et al., 2011). This tracrRNA binds to the crRNA, creating a long dsRNA duplex that is recognized and cleaved by RNase III. Cas9 is also required for proper guide processing to occur, most likely through protection and stabilization of the dsRNA (Deltcheva et al., 2011). The best studied exception to this model is the Type II-C systems within *Neisseria meningitidis* which uses individual promoters encoded by the CRISPR repeats to express mature guides independent of a processing enzyme (Zhang et al., 2013).

Recently it has been shown that the single effector protein from Type V systems, Cas12a (previously Cpf1) can self-process its own guide (Fonfara et al., 2016). The crRNAs within Cas12a containing systems, contain a single hairpin, and no tracrRNA. The processing nuclease domain of Cas12a is still debated. The original findings suggested a set of residues may be involved but the *in vivo* and *in vitro* data for processing activity was not consistent (Fonfara et al., 2016). Structural insights into the residues responsible for processing LbCpf1 suggest that 3 of the implicated residues are co-planar surrounding the 5' nucleotide of the guide (Dong et al., 2016; Gao et al., 2016).

Another common feature of CRISPR immune systems is the presence of multiple CRISPR-Cas loci, whether from redundant or distinct subtypes, within one host genome. Previous work on Type I and III systems has shown that the crRNA-processing enzyme, Cas6, can be shared across multiple effector complexes within some host organisms (Nickel et al., 2013; Niewoehner et al., 2014), while other organisms require distinct enzymes for each crRNA repeat species (Scholz et al., 2013; Sokolowski et al., 2014). Type I and Type III repeat species have also been shown to be processed by the same Cas6, suggesting cross talk between these two distinct CRISPR types (Nickel et al., 2013).

### **1.3.3 Target search and interference of phage infection**

Each of the different CRISPR Types targets a specific class of foreign nucleic acid substrate to protect their host. Best studied of these systems are the dsDNA targeting Type I and Type II complexes (van der Oost et al., 2014; Wright et al., 2016). Both of these systems use a protospacer adjacent motif (PAM) mechanism to protect the host genome at the CRISPR locus from self-targeting (Mojica et al., 2009; Sashital et al., 2012; Westra et al., 2013). RNA targeting by Type III and Type VI demonstrate a

wider variety of target interference mechanisms (Abudayyeh et al., 2016; Tamulaitis et al., 2017). For each of the CRISPR-Cas Types, we will address the specifics of their targeting mechanisms and outstanding questions.

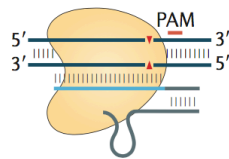
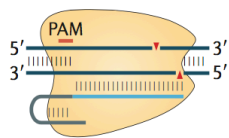
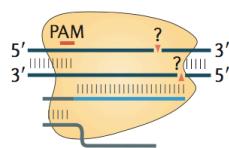
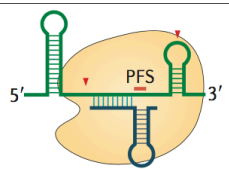
Type I interference complexes initially survey DNA substrates for the presence of a PAM (Redding et al., 2015). Upon PAM recognition Cascade initiates target binding through the seed region at the PAM proximal end of the target (Semenova et al., 2011). dsDNA unwinding generates an R loop and triggers a conformational change within Cascade to recruit Cas3, the system nuclease (Blosser et al., 2015; Hochstrasser et al., 2014; Semenova et al., 2011). Cas3 is loaded onto the open R loop, and will efficiently degrade the DNA through a presumed sliding mechanism (Redding et al., 2015; Sinkunas et al., 2013; Westra et al., 2012).

While they are structurally very similar, Type III interference complexes contain two different targeting modes, cleaving both ssRNA and ssDNA (Tamulaitis et al., 2017). Divided into four major subtypes, these interference complexes are generally referred to Csm (III-A/D) or Cmr (III-B/C) based on their subunit composition. Originally it was thought that Csm and Cmr complexes were distinguished by targeting either ssDNA (Marraffini and Sontheimer, 2008; 2010) or ssRNA respectively (Hale et al., 2009). Recent work unified the field, establishing that Type III complexes generally bind ssRNA targets, potentially dependent on nascent mRNAs within transcription bubbles (Samai et al., 2015). This activates Cas10 (Csm1 or Cmr2) nuclease for non-specific ssDNA cleavage, presumably on the R loop generated by the transcription bubble nearby (Elmore et al., 2016; Estrella et al., 2016; Han et al., 2016; Kazlauskienė et al., 2016; Liu et al., 2017c). The non-specific DNase activity is controlled by cleavage of the bound ssRNA target by the backbone subunits (Csm3 or Cmr4) at 6 nucleotide intervals (Kazlauskienė et al., 2016). Self-targeting in Type III is inhibited through a PAM-independent mechanism that relies on detection of base pairing between the crRNA handle and the ssRNA target sequence (Elmore et al., 2016; Estrella et al., 2016; Kazlauskienė et al., 2016).

Despite this unified mechanism of Type III, many open questions remain for these systems. Specifically, there are ancillary nucleases, Csx1/Csm6 proteins encoded within or nearby Type III operons, whose role in foreign DNA defense is unknown (Makarova et al., 2015). Both proteins have been shown to have non-specific nuclease activity (Niewoehner and Jinek, 2016; Sheppard et al., 2016) and are necessary for immune function in *S. solfataricus* (Type III-B, Cmr complex, Csx1) (Deng et al., 2013) and *S. epidermidis* (Type III-A, Csm complex, Csm6) (Jiang et al., 2016). How they are coordinated with the large interference complexes, along with their overall regulation is unknown.

Like Type I systems, Type II CRISPR systems target dsDNA, but use a single effector protein for activity (Jinek et al., 2012; Mohanraju et al., 2016). Cas9 utilizes two RNAs, the crRNA and trans-acting RNA (tracrRNA) that contain a long stretch of complementarity (Deltcheva et al., 2011). This crRNA: tracrRNA duplex can be joined to form a single guide RNA with minimal loss of activity, reducing the complexity of the system (Jinek et al., 2012). Target search is mediated through the PAM, similar to Type I interference complex Cascade, with target-guide duplex propagating from the PAM proximal to distal end (Sternberg et al., 2014). Target duplex formation triggers a further proofreading mechanism at the PAM distal end to serve as a final checkpoint for

nuclease activation (Sternberg et al., 2015). Cleavage mediated by Cas9 produces a double cut, nicking both strands of the dsDNA to generate a blunt end (Gasiunas et al., 2012; Jinek et al., 2012).

		Nuclease domains	tracrRNA	PAM	Substrate	Cleavage pattern
<b>Type II</b> Cas9		RuvC and HNH	Yes	3', GC-rich	dsDNA	Blunt ends
<b>Type V-A</b> Cas12a (Cpf1)		RuvC and Nuc	No	5', AT-rich	dsDNA	Staggered ends, 5' overhangs
<b>Type V-B</b> Cas12b (C2c1)		RuvC	Yes	5', AT-rich	dsDNA	Staggered seven-nucleotide cut of target DNA
<b>Type VI-A</b> Cas13a (C2c2)		2 HEPN domains	No	5', non-G PFS	ssRNA	Cleaves ssRNA near uracil and collateral activity

### Figure 1.8 Overview of Class 2 effector interference mechanisms

Graphic table of the major Class 2 CRISPR effectors. Proteins are depicted in yellow with guide:target base pairing noted by lines. Cut sites are noted as red triangles and notable features regarding the mechanisms are listed to the right. Adapted from Shmakov et al, 2017.

With the recent expansion of Class 2 CRISPR systems, many more variations on the same mechanistic theme have emerged (Shmakov et al., 2015; 2017). The first of these systems to be studied and applied towards genome manipulations was Cas12a (formerly Cpf1) from Type V systems (Zetsche et al., 2015). This Class 2 effector uses a single guide RNA to direct cleavage, and like Cas9 generates a double stranded break within the target dsDNA. Surprisingly, Cas12a generates a staggered cut, at the PAM-distal end of the target through a mechanism that is still debated (Dong et al., 2016; Zetsche et al., 2015). Crystal structures of the protein complexes reveal a mouth shaped architecture, very distinct from bilobed Cas9 (Dong et al., 2016; Gao et al., 2016; Yamano et al., 2016). Other groups have characterized the specificity of Cas12a within human cells and *in vitro* cleavage assays, concluding that Cas12a demonstrates a similar level of specificity and activity as Cas9 (Kim et al., 2017; Kleinstiver et al., 2016). Cas12b (formerly C2c1) is another Type V effector protein. Unlike Cas12a, it

uses two guide RNAs and appears to generate staggered dsDNA breaks extending beyond the spacer-target complementarity region (Shmakov et al., 2015; Yang et al., 2016). Little biochemically work has been done to determine the mechanism of cleavage for Cas12b beyond the basic requirements of the guide and protein (Shmakov et al., 2015; Yang et al., 2016). Structural work on Cas12a suggest a single activity site within the RuvC domain sequentially cleaves the two DNA strands (Liu et al., 2017a; Yang et al., 2016).

#### **1.4 Applications of CRISPR-Cas systems**

The programmable nature of CRISPR-Cas enzymes has enabled a revolution for DNA-targeting applications including gene editing, transcriptional control, and genomic imaging (Sternberg and Doudna, 2015; Wright et al., 2016). In contrast, repurposing the ssRNA-targeting Cas enzymes into biotechnological tools has been limited by a lack of mechanistic information. Type III systems are not amenable for facile engineering due to their multi-component nature and muddled mechanism (Tamulaitis et al., 2017). No single effector system prior to my studies had been shown to specifically target ssRNA. Yet, the allure of a programmable RNA binding-proteins has attracted interest for many years (Mackay et al., 2011). While RNAi systems had demonstrated a wide range of versatility problems exist in their propensity for off-targets and penetrance. In this work, we address this issue by redirecting a well-studied DNA-targeting CRISPR nuclease, Cas9, to ssRNA targets and investigating the enzymatic mechanisms of a novel ssRNA-targeting CRISPR nuclease, Cas13a (formerly C2c2).

# Chapter 2

## Redirecting CRISPR Type II systems for isolating specific RNA:protein complexes

A portion of the content presented in this chapter has been previously published as part of the following research article: Mitchell R. O'Connell, Benjamin L. Oakes, Samuel H. Sternberg, Alexandra East-Seletsky, Matias Kaplan & Jennifer A. Doudna. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516, 263–266.

For published studies: Mitchell R. O'Connell and Samuel H. Sternberg conceived the project. Mitchell R. O'Connell, Benjamin L. Oakes, and Samuel H. Sternberg performed biochemical investigations. Mitchell R. O'Connell, Benjamin L. Oakes, Alexandra Seletsky, and Matias Kaplan developed pulldown methods.

For preliminary results on mass spectrometry application development: Mitchell R. O'Connell and Alexandra Seletsky conceived the project, and Alexandra Seletsky performed all experimental work with assistance from Mitchell R. O'Connell.

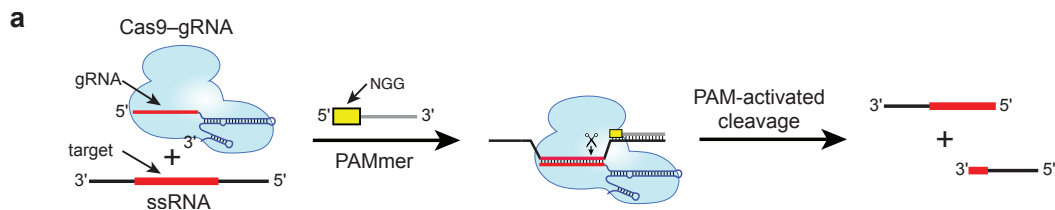


## 2.1 Chapter Summary

The CRISPR-associated protein Cas9 is an RNA-guided DNA endonuclease that uses RNA:DNA complementarity to identify target sites for sequence-specific double-stranded DNA (dsDNA) cleavage (Barrangou et al., 2007; Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012; Wiedenheft et al., 2012). In its native context, Cas9 acts on DNA substrates exclusively because both binding and catalysis require recognition of a short DNA sequence, the protospacer adjacent motif (PAM), next to and on the strand opposite the 20-nucleotide target site in dsDNA (Gasiunas et al., 2012; Jinek et al., 2012; Mojica et al., 2009; Sternberg et al., 2014). Cas9 has proven to be a versatile tool for genome engineering and gene regulation in many cell types and organisms (Mali et al., 2013), but it has been thought to be incapable of targeting RNA (Gasiunas et al., 2012). Here we show that Cas9 binds with high affinity to single-stranded RNA (ssRNA) targets matching the Cas9-associated guide RNA sequence when the PAM is presented *in trans* as a separate DNA oligonucleotide. Furthermore, PAM-presenting oligonucleotides (PAMmers) stimulate site-specific endonucleolytic cleavage of ssRNA targets, similar to PAM-mediated stimulation of Cas9-catalyzed DNA cleavage (Sternberg et al., 2014). Using specially designed PAMmers, Cas9 can be specifically directed to bind or cut RNA targets while avoiding corresponding DNA sequences, and we demonstrate that this strategy enables the isolation of a specific endogenous mRNA from cells. These results reveal a fundamental connection between PAM binding and substrate selection by Cas9, and highlight the utility of Cas9 for programmable and tagless transcript recognition.

## 2.2 Introduction

CRISPR–Cas immune systems must discriminate between self and non-self to avoid an autoimmune response (Marraffini and Sontheimer, 2010). In type I and II systems, foreign DNA targets which contain adjacent PAM sequences are targeted for degradation, whereas potential targets in CRISPR loci of the host do not contain PAMs and are avoided by RNA-guided interference complexes (Garneau et al., 2010; Gasiunas et al., 2012; Mojica et al., 2009; Sashital et al., 2012). Single-molecule and bulk biochemical experiments showed that PAMs act both to recruit Cas9–guide RNA complexes (Cas9–gRNA) to potential target sites and to trigger nuclease domain activation (Sternberg et al., 2014). Cas9 from *Streptococcus pyogenes* recognizes a 5'-NGG-3' PAM on the non-target (displaced) DNA strand (Jinek et al., 2012; Mojica et al., 2009), suggesting that PAM recognition may stimulate catalysis through allosteric regulation. Moreover, the HNH nuclease domain of Cas9, which mediates target strand cleavage (Gasiunas et al., 2012; Jinek et al., 2012), is homologous to other HNH domains that cleave RNA substrates (Hsia et al., 2004; Pommer et al., 2001). Based on the observation that single-stranded DNA (ssDNA) targets can be activated for cleavage by a separate PAMmer oligonucleotide (Sternberg et al., 2014), and that similar HNH domains can cleave RNA, we wondered whether a similar strategy would enable Cas9 to cleave ssRNA targets in a programmable fashion (Fig. 2.1).



**Figure 2.1 Schematic of ssRNA targeting strategy using a PAMmer.**

**a**, Schematic depicting the approach used to target ssRNA for programmable, sequence-specific cleavage.

Following on from these initial experiments that demonstrated we can redirect Cas9 for ssRNA target recognition and binding, we aimed to adapt this strategy into a robust technology for interrogating various aspects of RNA biology. While other technologies exist for isolating specific RNA species, most of these techniques have strong disadvantages that have limited their adoption by the scientific community. The competing technologies include capture hybridization analysis of RNA targets (CHART) (Simon et al., 2011), Chromatin Isolation by RNA Purification (ChIRP) (Chu et al., 2011) and RNA Antisense Purification (RAP) (Engreitz et al., 2013). All of these methods rely on multiple biotinylated oligonucleotide probes tiling across the RNA of interest. Through stringent, non-native washes, the specific RNAs can be isolated for analysis of associated nucleic acids and proteins. While these methods have been successful for isolating high abundance, long non-coding RNAs like HOTAIR (Chu et al., 2011) and XIST (McHugh et al., 2015), they have had limited success on low abundance transcripts and have a limited ability to isolate highly similar gene isoforms. As an alternative, we set out to develop a Cas9:PAMmer mediated method of ssRNA isolation.

## 2.3 Methods

### 2.3.1 Cas9 and nucleic acid preparation

Wild-type Cas9 and catalytically inactive dCas9 (D10A/H840A) from *S. pyogenes* were purified as previously described (Jinek et al., 2012). crRNAs (42 nt) were either ordered synthetically (Integrated DNA Technologies) or transcribed *in vitro* with T7 polymerase using single-stranded DNA templates, as described (Sternberg et al., 2012). tracrRNA was transcribed *in vitro* and contained nucleotides 15–87 following the numbering scheme used previously (Jinek et al., 2012).  $\lambda$ -targeting sgRNAs were *in vitro* transcribed from linearized plasmids and contain full-length crRNA and tracrRNA connected via a GAAA tetraloop insertion. *GAPDH* mRNA-targeting sgRNAs were *in vitro* transcribed from dsDNA PCR products based on an optimized sgRNA design (Chen et al., 2013). Target ssRNAs (55–56 nt) were *in vitro* transcribed using single-stranded DNA templates. Sequences of all nucleic acid substrates used in this study can be found in (Table 2.1 and Table 2.2).

All RNAs were purified using 10–15% denaturing polyacrylamide gel electrophoresis (PAGE). crRNA–tracrRNA duplexes were prepared by mixing equimolar

concentrations of each RNA in hybridization buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>), heating to 95 °C for 30 s and slow cooling. Fully double-stranded DNA/RNA substrates (substrates 1, 8–10 in Fig 2.3 and substrates 1–2 in Fig. 2.4) were prepared by mixing equimolar concentrations of each nucleic acid strand in hybridization buffer, heating to 95 °C for 30 s, and slow cooling. RNA, DNA, and chemically modified PAMmers were synthesized commercially (Integrated DNA Technologies). DNA and RNA substrates were 5'-radiolabeled using [ $\gamma$ -<sup>32</sup>P]-ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs). dsDNA and dsRNA substrates (Fig. 2.3) were 5'-radiolabeled on both strands, whereas only the target ssRNA was 5'-radiolabeled in other experiments.

Oligo Name	Sequence
$\lambda$ 2-targeting crRNA	5'-GUGAUAAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUUG-3'
$\lambda$ 3-targeting crRNA	5'-CUGGUGAACUCCGAUAGUGGUUUUAGAGCUAUGCUGUUUUUG-3'
$\lambda$ 4-targeting crRNA	5'-CAGATATAGCCTGGTGGTTCGUUUUAGAGCUAUGCUGUUUUUG-3'
ssDNA T7 template	5'-AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTA GCCTTATTTTAACCTGCTATGCTG TCCTATAGTGAGTCGTATTA -3'
tracrRNA (nt 15-87)	5'GGACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUUUUUU -3'
$\lambda$ 2-targeting sgRNA T7 template	5'TAATACGACTCACTATAGGTGATAAGTGGAATGCCATGGTTTTAGAGC TATGCTGTTTTGAAACAAAACAGCATAGCAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCCTTTTTT-3'
$\lambda$ 2-targeting sgRNA	5'GGUGAUAAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUUGAAAC AAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCUUUUUUUU-3'
$\lambda$ 2 target dsDNA duplex	5'-GAGTGGAAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAA TTGAGC-3' 3'-CTCACCTTCTACGGTCACTATTCACCTTACGGTACACCCGACAGTTTTA ACTCG-5'
$\lambda$ 2 ssDNA target strand	3'-CTCACCTTCTACGGTCACTATTCACCTTACGGTACACCCGACAGTTTTA ACTCG-5'
$\lambda$ 2 ssDNA nontarget strand	5'-GAGTGGAAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAA TTGAGC-3'
$\lambda$ 2 ssRNA target strand	3'-CUCACCUUCCUACGGUCACUAUUCACCUUACGGUACACCCGACAGUUU UAACUCGG-5'
$\lambda$ 2 ssRNA nontarget strand T7 template	5'-GCTCAATTTTGACAGCCACATGGCATTCCACTTATCACTGGCATCCTTC CACTCCTATAGTGAGTCGTA TTA-3'
$\lambda$ 2 ssRNA nontarget strand	5'-GGAGTGGAAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAA ATTGAGC-3'
19 nt $\lambda$ 2 DNA PAMmer	5'-TGGGCTGTCAAAATTGAGC-3'
18 nt $\lambda$ 2 "GG" PAMmer	5'-GGGCTGTCAAAATTGAGC-3'
19 nt $\lambda$ 2 DNA mutated PAMmer	5'-ACCGCTGTCAAAATTGAGC-3'
16 nt $\lambda$ 2 DNA "PAM-less" PAMmer	5'-GCTGTCAAAATTGAGC-3'
18 nt $\lambda$ 2 RNA PAMmer	5'-GGGCUUGUCAAAAUUGAGC-3'
5 nt $\lambda$ 2 DNA PAMmer	5'-TGGGC-3'
10 nt $\lambda$ 2 DNA PAMmer	5'-TGGGCTGTCA-3'
15 nt $\lambda$ 2 DNA PAMmer	5'-TGGGCTGTCAAAATT-3'
$\lambda$ 3 ssRNA target strand T7 template	5'-AACGTGCTGCGGCTGGCTGGTGAACCTCCGATAGTGCGGGTGTGAATG ATTTCCCTATAGTGAGTCGTAT TA-3'
$\lambda$ 3 ssRNA target strand	3'-UUGCACGACGCCGACCGACCACUUGAAGGCUAUCAGCCCAACUUA CUAAAGG-5'
$\lambda$ 4 ssRNA target strand T7 template	5'-TCACAACAATGAGTGGCAGATATAGCCTGGTGGTTCAGGCGGCGCATTT TTATTGCTATAGTGAGTCGT ATTA-3'
$\lambda$ 4 ssRNA target strand	3'-AGUGUUUGUACUCACCGUCUAUAUCGGACCACCAAGUCCGCCGCGUAA AAUAACGG-5'
147 $\lambda$ 3 ssDNA nontarget strand	5'-AACGTGCTGCGGCTGGCTGGTGAACCTCCGATAGTGCGGGTGTGAATG ATTTCC-3'
$\lambda$ 4 ssDNA nontarget strand	5'-TCACAACAATGAGTGGCAGATATAGCCTGGTGGTTCAGGCGGCGCATTT TTATTG-3'
19 nt $\lambda$ 3 DNA PAMmer	5'-CGGGTGTGAATGATTTCC-3'
19 nt $\lambda$ 4 DNA PAMmer	5'-AGGCGGCGCATTTTTATTG-3'

21 nt λ2 5'- extended DNA PAMmer	5'-TGTGGGCTGTCAAAATTGAGC-3'
21 nt λ3 5'- extended DNA PAMmer	5'-TGCGGGTGTGAATGATTTCC-3'
24 nt λ2 5'- extended DNA PAMmer	5'-CCATGTGGGCTGTCAAAATTGAGC-3'
24 nt λ3 5'- extended DNA PAMmer	5'-TAGTGC GG GTGTTGAATGATTTCC-3'
27 nt λ2 5'- extended DNA PAMmer	5'-ATGCCATGTGGGCTGTCAAAATTGAGC-3'
27 nt λ3 5'- extended DNA PAMmer	5'-CGATAGTGC GG GTGTTGAATGATTTCC-3'
30 nt λ2 5'- extended DNA PAMmer	5'-GGAATGCCATGTGGGCTGTCAAAATTGAGC-3'
30 nt λ3 5'- extended DNA PAMmer	5'-TTCCGATAGTGC GG GTGTTGAATGATTTCC-3'
33 nt λ2 5'- extended DNA PAMmer	5'-AGTGAATGCCATGTGGGCTGTCAAAATTGAGC-3'
33 nt λ3 5'- extended DNA PAMmer	5'-AACTCCGATAGTGC GG GTGTTGAATGATTTCC-3'
36 nt λ2 5'- extended DNA PAMmer	5'-ATAAGTGAATGCCATGTGGGCTGTCAAAATTGAGC-3'
39 nt λ2 5'- extended DNA PAMmer	5'-GTGATAAGTGAATGCCATGTGGGCTGTCAAAATTGAGC-3'
148 39 nt λ3 5'- extended DNA PAMmer	5'-CTGGTGAACCTCCGATAGTGC GG GTGTTGAATGATTTCC-3'
non-PAM λ2 dsDNA	5'-GAGTGG AAGGATGCCAGTGATAAGTGAATGCCATGACCGCTGTCAAAA TTGAGC-3'
	3'-CTCACCTTCTACGGTCACTATTCACCTTACGGTACTGGCGACAGTTTAA ACTCG-5'
non-PAM λ2 ssRNA target strand T7 template	5'-GAGTGG AAGGATGCCAGTGATAAGTGAATGCCATGACCGCTGTCAAAA TTGAGCCTATAGTGAGTCGTA TTA-3'
non-PAM λ2 ssRNA target strand	3'-CUCACCUUCCUACGGUCACUUAUUCACCUUACGGUACTGGCGACAGUUU UAACUCGG-5'
λ2 2'OMe capped PAMmerd	5'-*UGGGCTGTCAAAATTGAG*C-3'
λ2 PS capped PAMmerd	5'-*GGGCTGTCAAAATTGAG*C-3'
λ2 2'F capped PAMmerd	5'-*UGGGCTGTCAAAATTGAG*C-3'
λ2 LNA capped PAMmerd	5'-*TGGGCTGTCAAAATTGAG*C-3'
λ2 19 nt 2'OMe interspersed PAMmerd	5'-*UGGC*UGTCA*AAATT*GAG*C-3'

**Table 2.1 λ oligonucleotide sequences**

§ Positions of modifications depicted with asterisks preceding each modified nucleotide in each case (except for PS linkages which are depicted between bases)

PS: phosphorothioate bond

LNA: locked nucleic acid

Oligo Name	Sequence
GAPDH-targeting sgRNA 1 T7 template	5'TAATACGACTCACTATAGGGGCAGAGATGATGACCCTGTTTAAGAGCTATGCTGGAA ACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCTTTTTTT-3'
GAPDH-targeting sgRNA 1	5'GGGGCAGAGAUUGACCCUGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUU UAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUU UU-3'
GAPDH-targeting sgRNA 2 T7 template	5'TAATACGACTCACTATAGGCCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCTTTTTTT-3'
GAPDH-targeting sgRNA 2	5'GGCCAAAGUUGUCAUGGAUGACGUUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAG UUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUU UUUU-3'
GAPDH-targeting sgRNA 3 T7 template	5'TAATACGACTCACTATAGGCCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCTTTTTTT-3'
GAPDH-targeting sgRNA 3	5'GGAUGUCAUCAUUUUGGCAGGGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAG UUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUU UUUU-3'
GAPDH-targeting sgRNA 4 T7 template	5'TAATACGACTCACTATAGGATGTCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGA AACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTG CTTTTTTT-3'
GAPDH-targeting sgRNA 4	5'GGATGTCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTT AAAT AAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCCGGTGCTTTTTTT-3'
GAPDH PAMmer 1	5'-ATGACCCTTGGGGCTCCCCCTGCAAA-3'
GAPDH PAMmer 2	5'-TGGATGACCGGGGGCCAGGGGTGCTAAG-3'
GAPDH PAMmer 3	5'-TTGGCAGGTGGTTCTAGACGGCAGGTC-3'

GAPDH PAMmer 4	5'-CCCCAGCGTGGAAGGTGGAGGAGTGGG-3'
GAPDH PAMmer 1 2'OMe v1	5'-A*UGACC*CTAGG*GGCTC*CCCC*UGCAA*A-3'
GAPDH PAMmer 1 2'OMe v2	5'-*ATG*ACCC*UAGG*GGCT*CCCC*CCTG*CAA*A-3'
GAPDH PAMmer 1 2'OMe v3	5'-*ATG*ACC*CU*AGG*GGC*UCC*CCC*CTG*CAA*A-3'
GAPDH PAMmer 1 2'OMe v4	5'-*AT*GA*CC*CT*AGG*GG*CT*CC*CC*CC*UG*CA*AA-3'
GAPDH PAMmer 1 2'OMe v5	5'-*AT*GA*CC*CT*AG*GG*GC*TC*CC*CC*CU*GC*AA*A-3'
GAPDH cDNA primer Fwd	5'-CTCACTGTTCTCTCCCTCCGC-3'
GAPDH cDNA primer Rev	5'-AGGGGTCTACATGGCAACTG-3'
$\beta$ -actin cDNA primer Fwd	5'-AGAAAATCTGGCACCACACC-3'
$\beta$ -actin cDNA primer Rev	5'-GGAGTACTTGCGCTCAGGAG-3'

**Table 2.2 Oligonucleotides used in the *GAPDH* mRNA pull-down experiment**

sgRNAs for *GAPDH* were designed according to Chen, B. *et al.* (Chen et al., 2013)

‡ Positions of 2'OMe modifications are depicted with asterisks preceding each modified nucleotide.

### 2.3.2. Cleavage assays

Cas9–gRNA complexes were reconstituted before cleavage experiments by incubating Cas9 and the crRNA–tracrRNA duplex for 10 min at 37 °C in reaction buffer (20 mM Tris-HCl, pH 7.5, 75 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM dithiothreitol (DTT), 5% glycerol). Cleavage reactions were conducted at 37 °C and contained ~1 nM 5'-radiolabeled target substrate, 100 nM Cas9–RNA, and 100 nM PAMmer, where indicated. Aliquots were removed at each time point and quenched by the addition of RNA gel loading buffer (95% deionized formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 50 mM EDTA (pH 8.0), 0.025% (w/v) SDS). Samples were boiled for 10 min at 95 °C prior to being resolved by 12% denaturing PAGE. Reaction products were visualized by phosphorimaging and quantified with ImageQuant (GE Healthcare).

### 2.3.3. RNA cleavage site mapping

A hydrolysis ladder (OH<sup>-</sup>) was obtained by incubating ~25 nM 5'-radiolabeled  $\lambda$ 2 target ssRNA in hydrolysis buffer (25 mM CAPS (*N*-cyclohexyl-3-aminopropanesulfonic acid), pH 10.0, 0.25 mM EDTA) at 95 °C for 10 min, before quenching on ice. An RNase T1 ladder was obtained by incubating ~25 nM 5'-radiolabeled  $\lambda$ 2 target ssRNA with 1 Unit RNase T1 (NEB) for 5 min at 37 °C in RNase T1 buffer (20 mM sodium citrate, pH 5.0, 1 mM EDTA, 2 M urea, 0.1 mg mL<sup>-1</sup> yeast tRNA). The reaction was quenched by phenol/chloroform extraction before adding RNA gel loading buffer. All products were resolved by 15% denaturing PAGE.

### 2.3.4. Electrophoretic mobility shift assays

In order to avoid dissociation of the Cas9–gRNA complex at low concentrations during target ssRNA binding experiments, binding reactions contained a constant excess of dCas9 (300 nM), increasing concentrations of sgRNA, and 0.1–1 nM of target ssRNA. The reaction buffer was supplemented with 10  $\mu$ g mL<sup>-1</sup> heparin in order to avoid non-specific association of apo-dCas9 with target substrates (Sternberg et al., 2014). Reactions were incubated at 37 °C for 45 min before being resolved by 8% native PAGE at 4 °C (0.5 $\times$  TBE buffer with 5 mM MgCl<sub>2</sub>). RNA and DNA were visualized by phosphorimaging, quantified with ImageQuant (GE Healthcare), and analyzed with Kaleidagraph (Synergy Software).

### 2.3.5. Cas9 Biotin Labeling

To ensure specific labeling at a single residue on Cas9, two naturally occurring cysteine residues were mutated to serine (C80S and C574S) and a cysteine point mutant was introduced at residue M1. To attach the biotin moiety, 10  $\mu$ M WT Cas9 or dCas9 was reacted with a 50-fold molar excess of EZ-Link<sup>®</sup> Maleimide-PEG2-Biotin (Thermo Scientific) at 25 °C for 2 h. The reaction was quenched by the addition of 10 mM DTT, and unreacted Maleimide-PEG2-Biotin was removed using a Bio-Gel<sup>®</sup> P-6 column (Bio-Rad). Labeling was verified using a streptavidin bead binding assay, where 8.5 pmol of biotinylated Cas9 or non-biotinylated Cas9 was mixed with either 25  $\mu$ L streptavidin-agarose (Pierce Avidin Agarose; Thermo Scientific) or 25  $\mu$ L streptavidin magnetic beads (Dynabeads MyOne Streptavidin C1; Life Technologies). Samples were incubated in Cas9 reaction buffer at RT for 30 min, followed by three washes with Cas9 reaction buffer and elution in boiling SDS-PAGE loading buffer. Elutions were analysed using SDS-PAGE. Cas9 M1C biotinylation was also confirmed using mass spectroscopy performed in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley. Samples of intact Cas9 proteins were analyzed using an Agilent 1200 liquid chromatograph equipped with a Viva C8 (100 mm  $\times$  1.0 mm, 5  $\mu$ m particles, Restek) analytical column and connected in-line with an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). Mass spectra were recorded in the positive ion mode. Mass spectral deconvolution was performed using ProMass software (Novatia).

### 2.3.6. GAPDH mRNA pull-down

HeLa-S3 cell lysates were prepared as previously described (Lee et al., 2013). Total RNA was isolated from HeLa-S3 cells using Trizol reagent according to the manufacturer's instructions (Life Technologies). Cas9–sgRNA complexes were reconstituted before pull-down experiments by incubating a two-fold molar excess of Cas9 with sgRNA for 10 min at 37 °C in reaction buffer. HeLa total RNA (40  $\mu$ g) or HeLa lysate ( $\sim 5 \times 10^6$  cells) was added to reaction buffer with 40 U RNasin (Promega), PAMmer (5  $\mu$ M) and the biotin-dCas9 (50 nM):sgRNA (25 nM) in a total volume of 100  $\mu$ L and incubated at 37 °C for 1 h. This mixture was then added to 25  $\mu$ L magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1; Life Technologies) pre-equilibrated in reaction buffer and agitated at 4 °C for 2 h. Beads were then washed six times with 300  $\mu$ L wash buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5mM MgCl<sub>2</sub>, 0.1% Triton X-100, 5% glycerol, 1mM DTT, 10  $\mu$ g ml<sup>-1</sup> heparin). Immobilized RNA was eluted by heating beads at 70 °C in the presence of DEPC-treated water and a phenol/chloroform mixture. Eluates were then treated with an equal volume of glyoxal loading dye (Life Technologies) and heated at 50 °C for 1 h before separation via 1% BPTe agarose gel (30 mM Bis-Tris, 10 mM PIPES, 10 mM EDTA, pH 6.5). Northern blot transfers were carried out according to Chomczynski *et al.* (Chomczynski, 1992). Following transfer, membranes were crosslinked using UV radiation and incubated in pre-hybridization buffer (UltraHYB<sup>®</sup> Ultrasensitive Hybridization Buffer; Life Technologies) for 1 h at 46 °C prior to hybridization. Radioactive Northern probes were synthesized using random priming of *GAPDH* and  *$\beta$ -actin* partial cDNAs (for cDNA primers, see Table 2.2) in the presence of [ $\alpha$ -<sup>32</sup>P]-dATP (PerkinElmer), using a Prime-It II Random Primer Labeling kit (Agilent Technologies). Hybridization was carried out for 3 h in pre-hybridization buffer at 46 °C followed by two washes with 2 $\times$ SSC (300 mM

NaCl, 30 mM trisodium citrate, pH 7, 0.5% (w/v) SDS) for 15 min at 46 °C. Membranes were imaged using a phosphorscreen.

### 2.3.7. Design of primary microRNA targeting experiments

Target sites on pri-miRNAs were chosen using two microarray design tools, OligoWiz and PICKY to reduce potential off targets, control for nucleotide diversity, and control the desired PAMmer binding site melting temperature (Chou, 2010; Wernersson et al., 2007). Other design considerations were the lack of a PAM site to ensure no dsDNA targeting and close proximity to microprocessor cleavage site on the pri-miRNA. PAMmers were ordered from IDT with 2'-OMe modification pattern determined by Mitchell O'Connell to protect against RNaseH degradation while retaining RCas9 activity (unpublished work). sgRNA templates were PCR amplified using previously established method (Lin et al., 2014).

Name	Function	Sequence
MOC-234_sgRNA_univ_Olap_R	guidePCR	AAA AAA AGC ACC GAC TCG GTG CCA CTT TTT CAA GTT GAT AAC GGA CTA GCC TTA TTT AAA CTT GCT ATG CTG TTT CCA GC
MOC-235_sgRNA_univ_end_F	guidePCR	AAA AAA AGC ACC GAC TCG GTG C
AES276_miR7_sgRNA_1_Olap	guidePCR	TAATACGACTCACTATA G GAAATGAGAAGTTTGCTTGG GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES277_miR7_PAMmer_1	PAMmer	TmUmGmCTTmGmGmUGGmAmGmGCTmUmCmUTC
AES278_miR7_sgRNA_2_Olap	guidePCR	TAATACGACTCACTATA G GAACAAGTGGTTTTGGCAGCA GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES279_miR7_PAMmer_2	PAMmer	TmGmGmCAGmCmAmGGGmUmUmCACmCmAmGAG
AES280_miR7_sgRNA_3_Olap	guidePCR	TAATACGACTCACTATA G GACATTAGTAGAACAGAAT GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES281_miR7_PAMmer_3	PAMmer	AmAmCmAGAmAmUmUGGmGmAmAAAmCmGmGAA
AES241_miR17_sgRNA_1_Olap	guidePCR	TAATACGACTCACTATA GGAGTGTTATCAGTCAATTGTA GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES242_miR17_PAMmer_1	PAMmer	CmAmAmUTGmUmAmGGGmGmAmACGmUmAmAAT
AES243_miR17_sgRNA_2_Olap	guidePCR	TAATACGACTCACTATA GGAAGATATGAGGTCCATTTAG GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES244_miR17_PAMmer_2	PAMmer	CmCmAmUTTmAmGmAGGmCmGmUATmUmUmATG
AES245_miR17_sgRNA_3_Olap_F	guidePCR	TAATACGACTCACTATA GGACAGTATGTTTAGTCATATC GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES246_miR17_PAMmer_3	PAMmer	GmUmCmATAmUmCmGGGmAmUmACTmUmAmACA
AES247_miR17_sgRNA_4_Olap	guidePCR	TAATACGACTCACTATA GGAGTTACTCATGTATTGGTAA GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES248_miR17_PAMmer_4	PAMmer	AmUmUmGGTmAmAmAGGmGmGmUACmUmUmGCT
AES247b_miR17_sgRNA_5_Olap	guidePCR	TAATACGACTCACTATA GGAAAGCAATTCCTATTAAACC GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES248b_miR17_PAMmer_5	PAMmer	AmUmUmAAAmCmCmCGGmUmAmGAAmCmCmACA
AES282_miR17_sgRNA_6_Olap	guidePCR	TAATACGACTCACTATA GGCACATCAGATAGACCAGGCA GTT TAA GAG CTA TGC TGG AAA CAG CAT AGC AAG TTT AAA TAA GG
AES283_miR17_PAMmer_6	PAMmer	AmCmCmAGGmCmAmGGGmUmCmUACmAmUmCGA
AES119_GAPDH_7	qPCR	GAGTCAACGGATTTGGTCGT
AES120_GAPDH_8	qPCR	GACAAGCTTCCCCTTCTCAG
AES125_Actin_3	qPCR	AGAAAATCTGGCACACACC
AES126_Actin_4	qPCR	AGGGCATACCCCTCGTAGAT
AES288_pri-miR7_F	qPCR	ATGCAGATTTACACCTTACT
AES289_pri-miR7_R	qPCR	AGAAGAGTTCAAAAACCATTAGCA
AES290_pri-miR17_exon2	qPCR	TTGCCACGTGGATGTGAAGAT
AES291_pri-miR17_exon2	qPCR	GGTGGCTCTTCCAATGGCT
AES268_miR-092-1	qPCR	TCTACACAGTTGGGATCGG
AES269_miR-092-1	qPCR	CGGGACAAGTGAATACCATA
AES254_miR17_qPCR_intron3	qPCR	GACCAAGGTCTCAGACTGC

AES255 miR17 qPCR intron4	qPCR	ATCTGAATCTCACCCCACCG
AES260 miR-019a(p)	qPCR	CCAATAATTCAAGCCAAGCA
AES265 miR-019a(p)	qPCR	CAGGCAGATTCTACATCGACA

### Table 2.3 Table of oligos used for pri-miRNA studies

All sequences are listed 5'-3', 2'OMe modifications are noted as lowercase m preceding the modified base.

#### 2.3.8 qPCR validation of pri-miRNA pulldown

Pulldown elutions were mixed well with Trizol to isolate RNA from samples. After 5 min incubation, chloroform was added and samples spun for phase separation. Aqueous layer was removed and RNA containing solution was ethanol precipitated overnight using GlycoBlue Coprecipitant (ThermoFisher) to increase yields. Isolated RNA was resuspended in 30  $\mu$ l DEPC and 1  $\mu$ l was added to a 10  $\mu$ l cDNA synthesis reaction using random hexamer containing Vilo Master Mix (Invitrogen) using standard conditions. Resulting cDNA was further diluted 1:4, and then used as a template for qPCR analysis. Standard SybrGreen qPCR assay conditions were generally used with pre-purchased master mixes. 10% input samples were Trizol extracted in parallel to all experimental samples for reference. Relative enrichment values were calculated as  $2^{(\Delta\Delta Ct)}$  relative to the input sample beta-actin mRNA levels. Table 2.3 contains all qPCR primers used in this study.

#### 2.3.9 pri-miRNA isolation and mass spectrometry of binding proteins

For mass spectrometry data presented in this work, six 20 cm plates of 293T cells were harvested at 90% confluency by scraping into PBS. Cells were washed twice with PBS to remove any residual serum, and then incubated in 70 mL of PBS containing a final concentration of 1% formaldehyde. Cells were incubated in formaldehyde for 10 mins at room temperature with rocking to non-specifically crosslink protein:RNA complexes prior to lysis. The crosslinking reaction was quenched by the addition of 1 M Tris-HCl pH 7.4 in DEPC-H<sub>2</sub>O to a final concentration of 250 mM, and cell suspension rocked for 10 mins at room temperature. Cells were then gently washed twice with PBS to remove any residual formaldehyde. After final spin, cells were resuspended in 2 mL of hypotonic lysis buffer (Lee et al., 2013) and incubated on ice for 10 mins. Cells were then sonicated for 2 mins, 10 sec On, 40 sec Off, at power level 2.0. Lysates were spun down at 15,000 in benchtop centrifuge, and the supernatant was passed through a 0.22 syringe filter. This lysate could then be used immediately or stored at -80C. Generally, it is advised to re-filter lysates after freezing due to aggregation of the samples during freeze-thaw cycles.

Crosslinked pulldowns were performed using desthiobiotinylated labeled Cas9 and additional modifications to the PAMmers, which is not covered in this work. (unpublished, Mitchell R. O'Connell). These experiments were carried out similarly to the northern-blot monitored pulldowns described above with the following changes. Desthiobiotin-Cas9-sgRNA complexes were reconstituted before pull-down experiments by incubating a two-fold molar excess of Cas9 with sgRNA for 10 min at 37 °C in reaction buffer (generally 2  $\mu$ M Cas9: 1  $\mu$ M sgRNA). Clarified, crosslinked lysate (200  $\mu$ l, approximately  $30 \times 10^6$  cells/sample) was added to reaction buffer (75 mM KCl, 20 mM Tris pH 7.5, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 5 % glycerol, 10  $\mu$ g/mL Heparin)



with 40 U RNasin (Promega), PAMmer (2.5  $\mu$ M) and desthiobiotin-dCas9 (100 nM): sgRNA (50 nM) complex in a total volume of 400  $\mu$ L and incubated at 37 °C for 1 h. This mixture was then added to 50  $\mu$ L magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1; Life Technologies) pre-equilibrated in reaction buffer and agitated at 4 °C for 1 h. Three washes were performed using a similar buffer to before, except that the salt concentration was raised to 500 mM KCl. Complexes were eluted off the beads using 150  $\mu$ L of 25 mM biotin-containing reaction buffer by vigorous shaking at 37 °C for 1 h. Eluants were separated from the beads using the magnetic rack.

Elutions were analyzed by silver-stained SDS-PAGE prior to mass spectrometry analysis. Silver staining was done using previously described protocol (Chevallet et al., 2006). For mass spectrometry analysis, lysate crosslinks were reversed by boiling samples at 95 °C for 10 mins in SDS-PAGE loading dye. 40  $\mu$ L of elution was loaded into wide wells of a stacked 15% SDS-PAGE gel and gel was run for 5-10 mins, so that the control lane containing protein ladder had fully left the stacking gel. Using the ladder as a guide, each sample was cut out of the gel and submitted for a tryptic digest and mass spectroscopy analysis performed by the QB3 Mass Spectrometry Facility at UC Berkeley.

### **2.3.10 Mass spectrometry of elutions.**

Mass spectrometry was performed by the Vincent J.Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley. A nano-LC column was packed in a 100  $\mu$ m inner diameter glass capillary with an emitter tip. The column consisted of 10 cm of Polaris C18 5  $\mu$ m packing material (Varian), followed by 4 cm of Partisphere 5 SCX (Whatman). The column was loaded using a pressure bomb and washed extensively with buffer A (see below). The column was then directly coupled to an electrospray ionization source mounted on a Thermo-Fisher LTQ XL linear ion trap mass spectrometer. An Agilent 1200 HPLC equipped with a split line to deliver a flow rate of 300 nl/min was used for chromatography. Peptides were eluted using a 4-step MudPIT procedure (Washburn et al., 2001). Buffer A was 5% acetonitrile/ 0.02% heptafluorobutyric acid (HBFA); buffer B was 80% acetonitrile/ 0.02% HBFA. Buffer C was 250 mM ammonium acetate/ 5% acetonitrile/ 0.02% HBFA; buffer D was same as buffer C, but with 500 mM ammonium acetate.

Protein identification and quantification were done with Integrated Proteomics Pipeline (IP2, Integrated Proteomics Applications, Inc. San Diego, CA) using ProLuCID/Sequest, DTASelect2 and Census (Cociorva et al., 2007; Park et al., 2008; Tabb et al., 2002; Xu et al., 2015). Tandem mass spectra were extracted into ms1 and ms2 files from raw files using RawExtractor (McDonald et al., 2004) and were searched against the human protein database supplemented with Cas9 sequences and sequences of common contaminants, concatenated to a decoy database in which the sequence for each entry in the original database was reversed (Peng et al., 2003). The human database was downloaded from NCBI; the version date of the database is February 4, 2014. LTQ data was searched with 3000.0 milli-amu precursor tolerance and the fragment ions were restricted to a 600.0 ppm tolerance. All searches were parallelized and searched on the VJC proteomics cluster. Search space included all fully tryptic peptide candidates with no missed cleavage restrictions. Carbamidomethylation (+57.02146) of cysteine was considered a static modification.

We required 1 peptide per protein and both tryptic termini for each peptide identification. The ProLuCID search results were assembled and filtered using the DTASelect program (Cociorva et al., 2007; Tabb et al., 2002; Xu et al., 2015) with a peptide false discovery rate (FDR) of 0.001 for single peptides and a peptide FDR of 0.005 for additional peptides for the same protein. Under such filtering conditions, the estimated false discovery rate was zero for the dataset used.

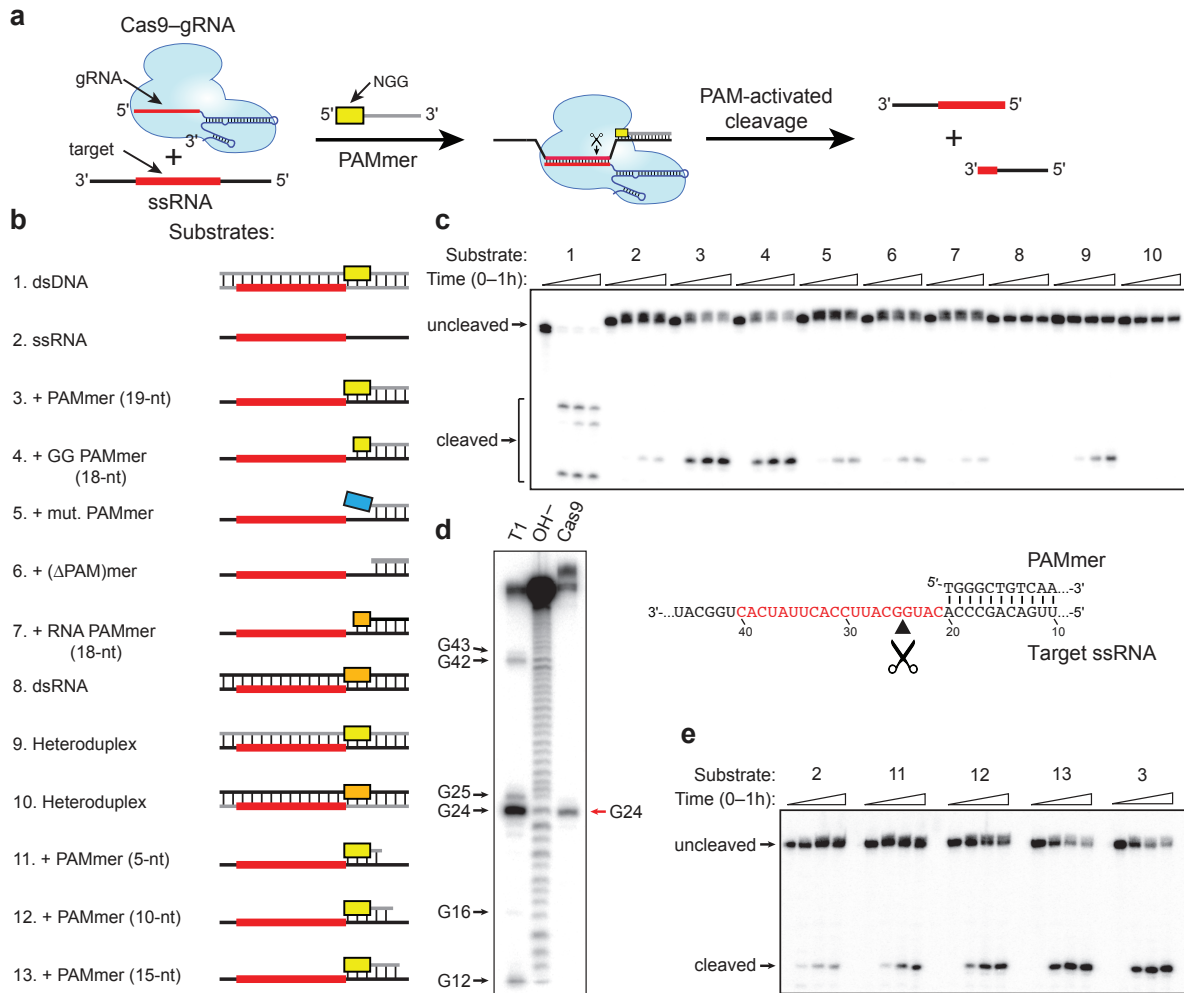
### **2.3.11 Processing of mass spectrometry data.**

For the data presented in this dissertation, observed peptide counts of a given protein were summed to assign protein level abundance values. As this data is considered preliminary, no trimming of the dataset was done to eliminate low abundance peptides. In order to compare across samples, protein level peptide abundances were normalized to the number of Cas9 peptides present in each sample. Relative enrichment scores were then calculated by dividing each protein abundance in targeted samples, against the abundance in the non-targeting  $\lambda$ 2 control. For simplicity, all non-observed proteins were assigned a value of 0.5 for this calculation. Specific classes of proteins were also omitted for all presented analysis including ribosomal proteins, chaperones, cytoskeleton proteins and histones.

## **2.4 Results**

### **2.4.1 RNA-guided Cas9 cleaves ssRNA targets in the presence of a short PAM-presenting DNA oligonucleotide (PAMmer)**

Using *S. pyogenes* Cas9 and dual-guide RNAs, we performed *in vitro* cleavage experiments using a panel of RNA and DNA targets (Fig. 2.2 and Table 2.1). Deoxyribonucleotide-comprised PAMmers specifically activated Cas9 to cleave ssRNA (Fig. 2.2), an effect that required a 5'-NGG-3' or 5'-GG-3' PAM. RNA cleavage was not observed using ribonucleotide-based PAMmers, suggesting that Cas9 may recognize the local helical geometry and/or deoxyribose moieties within the PAM. Consistent with this idea, dsRNA targets were not cleavable, and RNA–DNA heteroduplexes could only be cleaved when the non-target strand was composed of deoxyribonucleotides. Interestingly, we found that Cas9 cleaved the ssRNA target strand between positions 4 and 5 of the base-paired guide RNA-target RNA hybrid (Fig. 2.2d), in contrast to the cleavage between positions 3 and 4 observed for dsDNA (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012) likely due to subtle differences in substrate positioning.

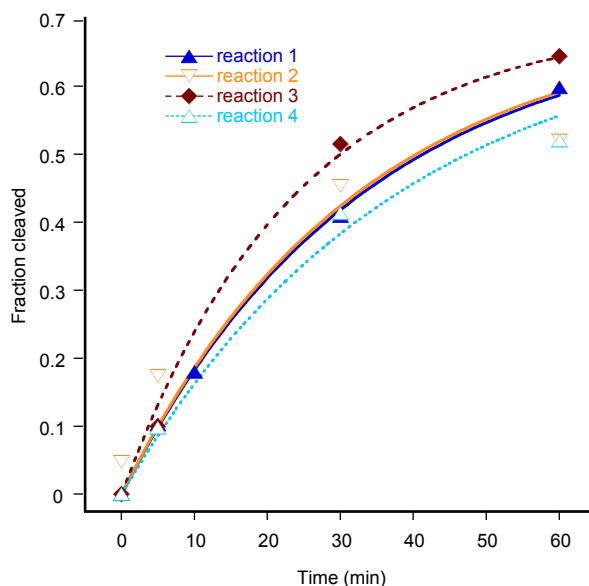


**Figure 2.2 RNA-guided Cas9 cleaves ssRNA targets in the presence of a short PAM-presenting DNA oligonucleotide (PAMmer).**

**a**, Schematic depicting the approach used to target ssRNA for programmable, sequence-specific cleavage. **b**, The panel of nucleic acid substrates examined in this study. Substrate elements are colored as follows: DNA (grey), RNA (black), guide RNA target sequence (red), DNA PAM (yellow), mutated DNA PAM (blue), RNA PAM (orange). **c**, Representative cleavage assay for 5'-radiolabeled nucleic acid substrates using Cas9-gRNA, numbered as in (b). **d**, Cas9-gRNA cleavage site mapping assay for substrate 3. T1 and OH<sup>-</sup> denote RNase T1 and hydrolysis ladders, respectively; the sequence of the target ssRNA is shown at right. **e**, Representative ssRNA cleavage assay in the presence of PAMmers of increasing length, numbered as in (b).

We hypothesized that PAMmer nuclease activation would depend on the stability of the hybridized PAMmer-ssRNA duplex and tested this by varying PAMmer length. As expected, ssRNA cleavage was lost when the predicted melting temperature for the duplex decreased below the temperature used in our experiments (Fig. 2.2). In addition, we did observe a significant reduction in the pseudo-first order cleavage rate constant of PAMmer-activated ssRNA as compared to ssDNA (Sternberg et al., 2014) (Fig. 2.3).

Collectively, these data demonstrate that hybrid substrate structures composed of ssRNA and deoxyribonucleotide-based PAMmers that anneal upstream of the RNA target sequence can be cleaved efficiently by RNA-guided Cas9.

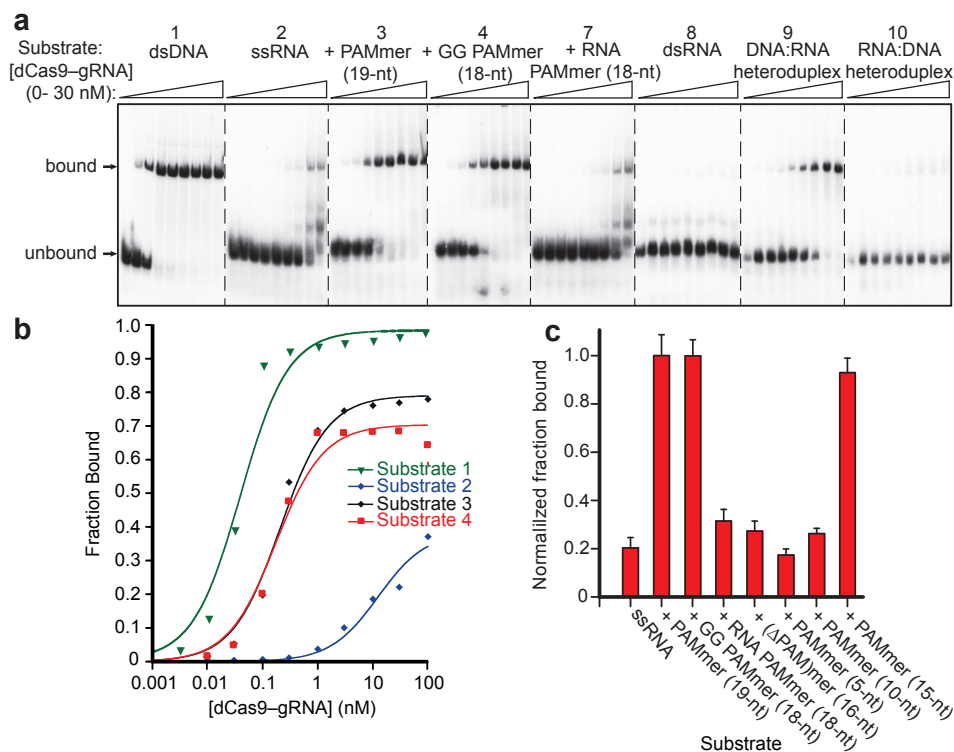


**Figure 2.3 Quantified data for cleavage of ssRNA by Cas9–gRNA in the presence of a 19-nt PAMmer.**

Cleavage assays were conducted as described in the Methods, and the quantified data were fit with single-exponential decays. Results from four independent experiments yielded an average apparent pseudo-first order cleavage rate constant of  $0.032 \pm 0.007 \text{ min}^{-1}$ . This is slower than the rate constant determined previously for ssDNA in the presence of the same 19-nt PAMmer ( $7.3 \pm 3.2 \text{ min}^{-1}$ ) (Sternberg et al., 2014).

#### **2.4.2 dCas9–gRNA binds ssRNA targets with high affinity in the presence of PAMmers**

We investigated the binding affinity of catalytically inactive (dCas9; D10A/H840A) dCas9–gRNA for ssRNA targets with and without PAMmers using a gel mobility shift assay. Intriguingly, while our previous results showed that ssDNA and PAMmer-activated ssDNA targets are bound with indistinguishable affinity (Sternberg et al., 2014), PAMmer-activated ssRNA targets were bound >500-fold tighter than ssRNA alone (Fig. 2.4). A recent crystal structure of Cas9 bound to a ssDNA target revealed deoxyribose-specific van der Waals interactions between the protein and the DNA backbone (Nishimasu et al., 2014), suggesting that energetic penalties associated with ssRNA binding must be attenuated by favorable compensatory binding interactions with the provided PAM. The equilibrium dissociation constant measured for a PAMmer–ssRNA substrate was within 5-fold of that for dsDNA (Fig. 2.4b), and this high-affinity interaction again required a cognate deoxyribonucleotide-comprised 5'-GG-3' PAM (Fig. 2.4a). Tight binding also scaled with PAMmer length (Fig. 2.4c), consistent with the cleavage data presented above.



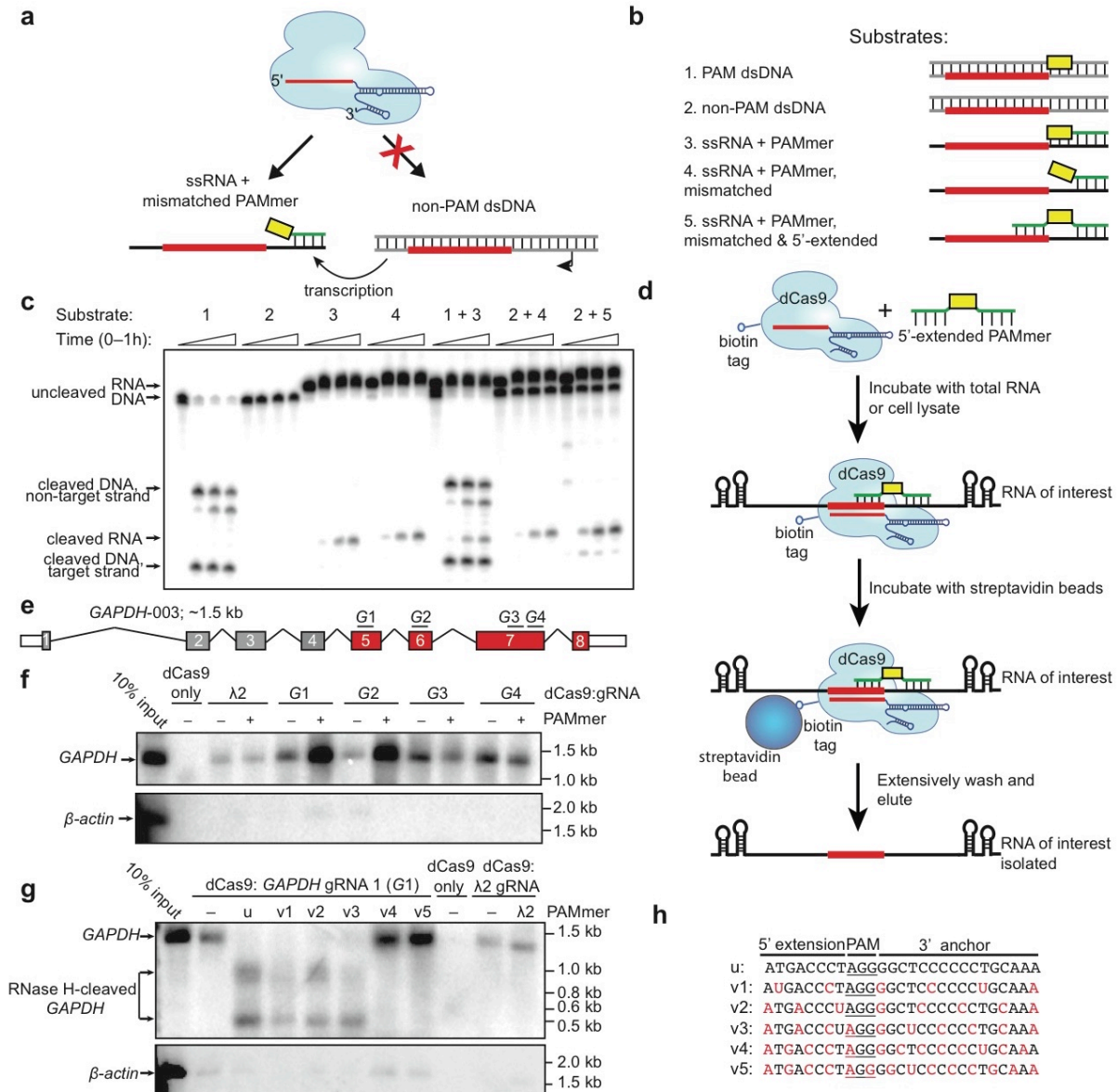
**Figure 2.4 dCas9-gRNA binds ssRNA targets with high affinity in the presence of PAMmers.**

**a**, Representative electrophoretic mobility shift assay for binding reactions with dCas9-gRNA and a panel of 5'-radiolabeled nucleic acid substrates, numbered as in Fig. 2.2, Quantified binding data for substrates 1-4 from (a) fit with standard binding isotherms. Measured dissociation constants from three independent experiments (mean  $\pm$  s.d.) were  $0.036 \pm 0.003$  nM (1),  $>100$  nM (2),  $0.20 \pm 0.09$  nM (3), and  $0.18 \pm 0.07$  nM (4). **c**, Relative binding data for 1 nM dCas9-gRNA and 5'-radiolabeled ssRNA with a panel of different PAMmers. The data are normalized to the amount of binding observed at 1 nM dCas9-gRNA with a 19-nt PAMmer; error bars represent the standard deviation from three independent experiments.

### 2.4.3 PAMmer guided dCas9-gRNA can isolate GAPDH mRNA from total RNA and cell lysates

Next, we investigated whether nuclease activation by PAMmers requires base-pairing between the 5'-NGG-3' and corresponding nucleotides on the ssRNA. Prior studies showed that DNA substrates containing a cognate PAM that is mismatched with the corresponding nucleotides on the target strand are cleaved as efficiently as a fully base-paired PAM (Jinek et al., 2012). Importantly, this could enable targeting of RNA while precluding binding or cleavage of corresponding genomic DNA sites lacking PAMs (Fig. 2.5). To test this possibility, we first demonstrated that Cas9-gRNA cleaves PAMmer-ssRNA substrates regardless of whether or not the PAM is base-paired (Fig. 2.5). When Cas9-RNA was incubated with both a PAMmer-ssRNA substrate and the corresponding dsDNA template containing a cognate PAM, both targets were cleaved.

In contrast, when a dsDNA target lacking a PAM was incubated together with a PAMmer-ssRNA substrate bearing a mismatched 5'-NGG-3' PAM, Cas9-gRNA selectively targeted the ssRNA for cleavage (Fig. 2.5). The same result was obtained using a mismatched PAMmer with a 5' extension (Fig. 2.5), demonstrating that this general strategy enables the specific targeting of RNA transcripts while effectively eliminating any targeting of their corresponding dsDNA template loci.



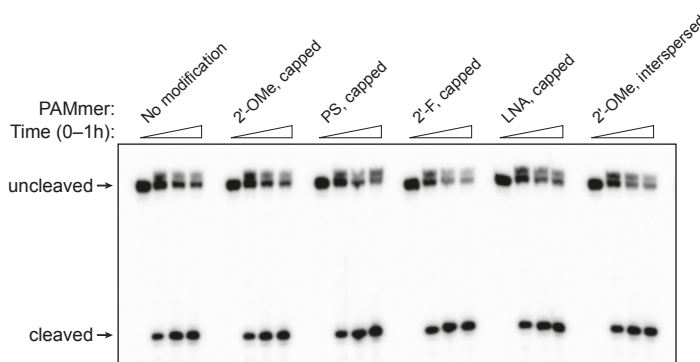
**Figure 2.5 RNA-guided Cas9 can target non-PAM sites on ssRNA and isolate GAPDH mRNA from HeLa cells in a tagless manner.**

**a**, Schematic of the approach designed to avoid cleavage of template DNA by targeting non-PAM sites in the ssRNA target. **b**, The panel of nucleic acid substrates tested in (c). **c**, Cas9-gRNA cleaves ssRNA targets with equal efficiency when the 5'-NGG-3' of the



PAMmer is mismatched with the ssRNA. This strategy enables selective cleavage of ssRNA in the presence of non-PAM target dsDNA. **d**, Schematic of the dCas9 RNA pull-down experiment. **e**, *GAPDH* mRNA transcript isoform 3 shown schematically, with exons common to all *GAPDH* protein-coding transcripts in red and gRNA/PAMmer targets G1-G4 indicated. **f**, Northern blot showing that gRNAs and 5'-extended PAMmers enable tagless isolation of *GAPDH* mRNA from HeLa total RNA;  $\beta$ -actin mRNA is shown as a control. **g**, Northern blot showing tagless isolation of *GAPDH* mRNA from HeLa cell lysate with varying 2'-OMe-modified PAMmers. RNase H cleavage is abrogated with v4 and v5 PAMmers;  $\beta$ -actin mRNA is shown as a control. **h**, Sequences of unmodified and modified *GAPDH* PAMmers used in (**g**); 2'-OMe-modified nucleotides are shown in red.

We next explored whether Cas9-mediated RNA targeting could be applied for tagless transcript isolation from HeLa cells (Fig. 2.5). The immobilization of Cas9 on a solid-phase resin is described in Methods. As a proof of concept, we first isolated *GAPDH* mRNA from HeLa total RNA using biotinylated dCas9, gRNAs and PAMmers (Table 2.2) that target four non-PAM-adjacent sequences within exons 5–7 (Fig. 2.5). We observed a substantial enrichment of *GAPDH* mRNA relative to a control  $\beta$ -actin mRNA by Northern blot analysis, but saw no enrichment using a non-targeting gRNA or dCas9 alone (Fig 2.5).



**Figure 2.6 RNA-guided Cas9 can utilize chemically modified PAMmers.**

19-nt PAMmer derivatives containing various chemical modifications on the 5' and 3' ends (capped) or interspersed still activate Cas9 for cleavage of ssRNA targets. These types of modification are often used to increase the *in vivo* half-life of short oligonucleotides by preventing exo- and endonuclease-mediated degradation. Cleavage assays were conducted as described in the Methods. PS, phosphorothioate bonds; LNA, locked nucleic acid.

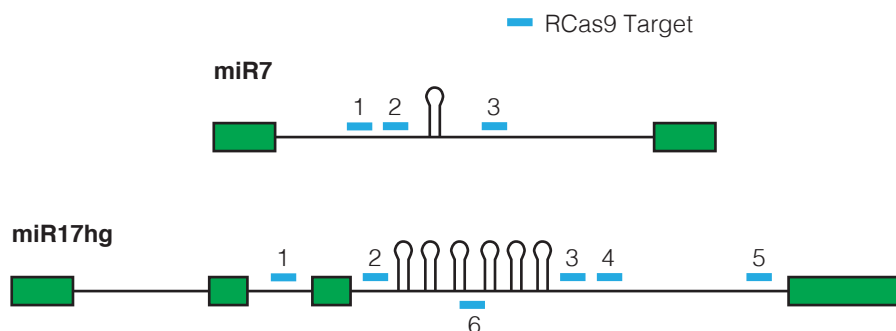
We then used this approach to isolate endogenous *GAPDH* transcripts from HeLa cell lysate under physiological conditions. In initial experiments, we found that Cas9–gRNA captured two *GAPDH*-specific RNA fragments rather than the full-length mRNA (Fig 2.5). Based on the sizes of these bands, we hypothesized that RNA:DNA heteroduplexes formed between the mRNA and PAMmer were cleaved by cellular RNase H. Previous studies have shown that modified DNA oligonucleotides can abrogate RNase H activity (Wu et al., 1999), and therefore we investigated whether

Cas9 would tolerate chemical modifications to the PAMmer. We found that a wide range of modifications (locked nucleic acids, 2'-OMe and 2'-F ribose moieties) still enabled PAMmer-mediated nuclease activation (Fig 2.6). Importantly, by varying the pattern of 2'-OMe modifications in the PAMmer, we could completely eliminate RNase H-mediated cleavage during the pull-down and successfully isolate intact *GAPDH* mRNA (Fig. 2.5). Interestingly, we consistently observed specific isolation of *GAPDH* mRNA in the absence of any PAMmer, albeit with lower efficiency, suggesting that Cas9-gRNA can bind to *GAPDH* mRNA through direct RNA:RNA hybridization (Fig. 2.5) These experiments demonstrate that RNA-guided Cas9 can be used to purify endogenous untagged RNA transcripts. We hereby refer to this approach as RCas9. In contrast to current oligonucleotide-mediated RNA-capture methods, this approach works well under physiological salt conditions and does not require crosslinking or large sets of biotinylated probes (Chu et al., 2011; Engreitz et al., 2013; Simon et al., 2011).

## 2.5 Preliminary Results

### 2.5.1 pri-miRNA target design and validation

With these results in hand, we set out to develop a tagless method for RNA-protein complex isolation using RCas9. Our goal was to develop a facile method for mass spectrometry analysis of these RNA:protein complexes. To facilitate initial method development, we chose a model RNA:protein system, namely primary microRNAs (pri-miRNAs), which have well-established binding proteins including DCGR8 and Drosha, combined with recently identified cis-regulatory motifs required for efficient processing, with unknown binding partners (Auyeung et al., 2013; Ha and Kim, 2014). Using this system, we hoped to identify novel pri-miRNA processing modulators in addition to using known partners as positive controls for our studies.



**Figure 2.7 pri-miRNA target design**

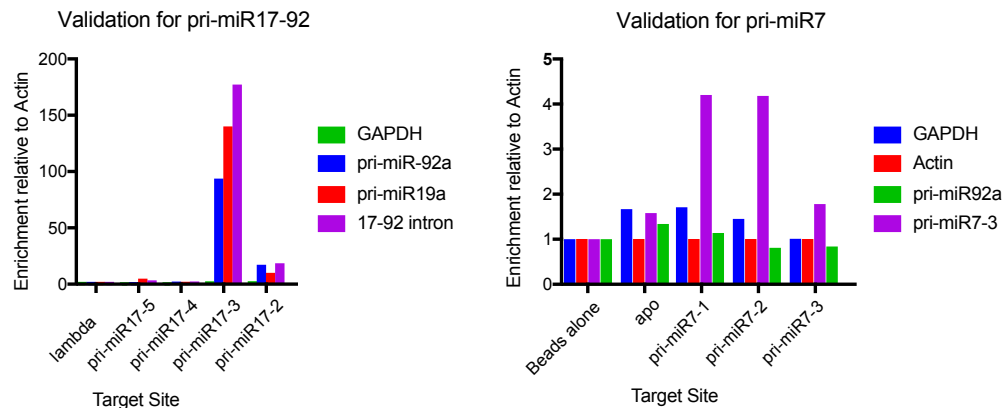
Schematic of targeting sites of pri-miR7 and pri-miR17-92. Green boxes represent exons while the thin black line represents intronic sequences. A variety of targeting sites were chosen to flank the pri-miRNA hairpins of both loci.

Two pri-miRNA species were chosen for initial studies, miR7 and miR17-92 (Fig 2.7). miR17-92 is a well characterized cluster of six pri-miRNA hairpins that drives cancer progression (Olive et al., 2010). The six hairpins of the cluster are pleiotropic with their downstream targets impacting cell proliferation, apoptosis, and angiogenesis. The high expression levels of miR17-92 off a long non-coding transcript, differential processing patterns, and known binding partners, made this pri-miRNA an attractive



target for method development (Chakraborty et al., 2012; Guil and Cáceres, 2007; Thomas et al., 2013). miR7 was chosen due to distinct features from miR17-92. For example, miR7 is not part of a miRNA cluster, and resides in the last intron of a highly expressed gene hnRNP K (Choudhury et al., 2013). Despite ubiquitously high expression of the hnRNP K locus, miR7 is lowly expressed in most tissues except for the brain and pancreas, suggesting a strong post-transcriptional repression of the pri-miRNA or pre-miRNA. Studied together, we hoped these two pri-miRNAs would provide contrasting modulators of miRNA processing.

RCas9 target sites on both pri-miRNA targets were chosen as described in the methods. To validate the design and chose the best guides for further development, we perform qPCR on pulldown elutions from total RNA and cell lysates (Fig 2.8). Similar to the northern blotting analysis of GAPDH pulldowns, not all target sites were equally efficient at isolating the particular RNA of interest. The basis of this variability is unknown. For miR17-92, multiple primer pairs were used to validate intact cluster isolation including the intronic region, the second (pri-miR19a) and the sixth (pri-miR92a) hairpins. Target site 3, and to a lesser extent site 2 showed the most enrichment by qPCR in total RNA (Fig 2.8). For pri-miR7 initial testing was performed in cell lysate, which is generally results in lower enrichment values. Target sites 1 and 2 which both reside on the 5' side of the miR7 hairpins both performed the most robustly for pri-miR7 isolation (Fig 2.8).

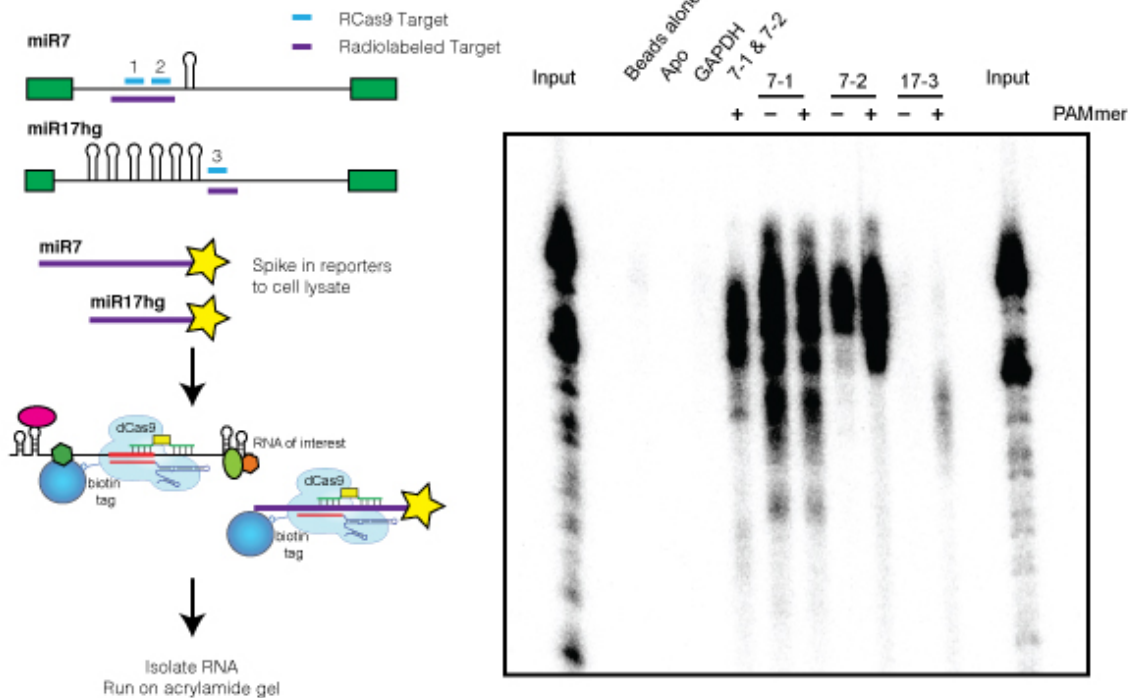


**Figure 2.8 qPCR validation of pri-miRNA pulldown**

qPCR validation of RCas9 pulldown of pri-miR17-92 and pri-miR7 in total RNA for miR17-92 and cell lysate for pri-miR7. Pulldown was performed with 100 nM biotin-dCas9: 50 nM guide. Pulldown wash condition used here were unoptimized at this point, using 150 mM KCl containing wash buffer for both sets of reactions.

Additional *in vitro* validation of these sgRNA's was performed using radiolabeled reporters. For these experiments, radiolabeled reporters for target sites 1 & 2 of pri-miR7 and target site 3 of pri-miR17-92 were transcribed *in vitro*, radiolabeled and then used to monitor pulldown efficiencies. Protection and isolation of these fragments was measured by denaturing PAGE analysis. By this analysis pulldown of pri-miR7 by target site 2 was more PAMmer dependent and specific for the miR7 fragment than

target 1 site. This biochemical validation of the pulldown method again highlights the variability observed in target site efficiencies.



**Figure 2.9 Radiolabeled reporter validation of pri-miR17-92 and pri-miR7 targets.**

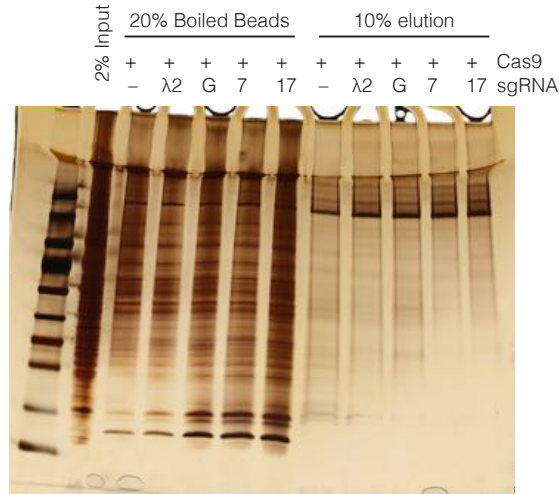
**a.** Schematic of radiolabeled pull down strategy. Two different size ssRNA reporters were transcribed, a 70-nt length reporter for pri-miR17 and 90-nt length reporter for pri-miR7. **b.** These two reporters were radiolabeled, and then co-incubated with biotin-RCas9 complexed to pri-miRNA guides with and without PAMmers. Elutions were phenol extracted and then analyzed using a 10% denaturing PAGE gel.

### 2.5.2 Mass spectrometry analysis of pri-miRNA binding proteins

Next, we wanted to identify the proteins bound to each of these pri-miRNAs by mass spectrometry. One significant problem to overcome during this phase of the method development was reducing the non-specific association of proteins to the magnetic beads. Initial mass spectrometry results were highly polluted with ribosomal proteins and other highly abundant housekeeping proteins, yielding little difference between the control and pri-miRNA pulldown samples. To overcome this problem, we transitioned to using an elutable desoethiobiotin tag on Cas9 and increasing the salt concentration to 500 mM during the washing steps, which both helped to decrease these non-specific contaminants. Another key change within the protocol was the addition of crosslinking to the lysate preparation. 1% formaldehyde treatment of the total cell lysates yield higher concentrations of eluted protein, and reduced concerns of the higher salt wash removing bound proteins. Crosslinks were reversed by boiling sample in SDS-PAGE loading dye prior to sample submission to the Mass Spectroscopy core facility.

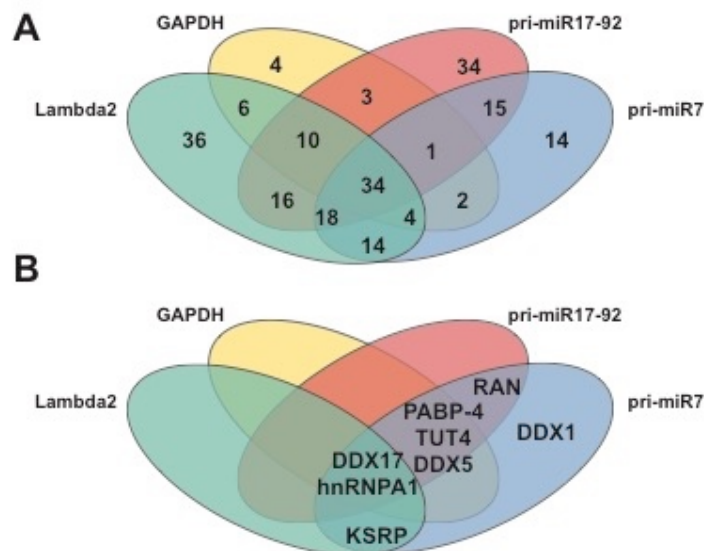
With these protocol changes we performed proof-of-concept pulldowns on pri-miR7 target site 2 and pri-miR17-92 target site 6 (validation data not shown, but we

generally see fold-enrichment values between 6-10 in cell lysates). Each Pulldown was performed as described in the methods and prior to any mass spectrometry analysis, successful RNA:protein complex pulldown was confirmed by silver-stained SDS-PAGE (Fig 2.10).



**Figure 2.10 Silver-stained SDS-PAGE of Pulldown elutions**

a. Pulldown elutions and beads were boiled in SDS-PAGE loading dye prior to loading on 4-20% pre-cast Tris-PAGE gel. Cas9 band is visible in all eluted samples, and appears to be at a constant concentration. Slight banding differences are observed between pulldown samples, suggesting that the pulldown strategy is successful. A significant fraction of the proteins isolated in the pulldown remains bound to the beads after specific elution of the desoethiobiotin labeled Cas9.



**Figure 2.11 Summary of isolated proteins by RCas9 pulldown of pri-miR7, pri-miR17-92 and GAPDH RNAs**

**a.** Venn diagrams of isolated proteins in each sample. Numbers in each section correspond to the number of proteins identified in within the indicated samples. **b.** Short protein names of known pri-miRNA binding proteins are highlighted in their locations on the Venn diagram.

Using mass spectroscopy data obtained from these pulldown samples, we globally categorized the proteins identified in all samples. Although the high salt wash had reduced background binding, a significant fraction of our peptides originated from ribosomal, cytoskeleton or histone proteins. These peptides were omitted from the processed dataset, as they are expected to be background non-specific RNA-binders. In the future, it is suggested that a low-salt (50 mM), no  $Mg^{2+}$  wash should be incorporated into the protocol, especially on cross linked samples to reduce the quantity of ribosomes recovered. Overall, the global analysis of this data indicates that while background binding is still high (large number of proteins bound to non-targeting guide), the pri-miRNA samples have many unique and overlapping hits between them (34 unique binders for pri-miR17-92, 14 for pri-miR7, and 15 shared binders) (Fig 2.11).

Many positive controls proteins were isolated as highlighted in the Venn diagram in Fig. 2.11b. For the microRNA pathway genes, we were able to recover general pathway proteins including DDX5 (p68), DDX17 (p72), KSRP, TUT4, and RAN (Ha and Kim, 2014). Additionally previously established specific pri-miR7 and pri-miR17-92 binding partners like hnRNPA1 (both) and DDX1 (only pri-miR7) were isolated in the expected samples (Choudhury et al., 2013; Guil and Cáceres, 2007; Han et al., 2014). Alarmingly, a significant number of these proteins were present in the  $\lambda 2$  and GAPDH samples which would suggest they were not specifically bound to the pri-miRNA samples. To control for the relative abundance in each sample, enrichment scores relative to the  $\lambda 2$  sample were calculated for all targeting pulldowns (GAPDH, pri-miR7, pri-miR17-92) (Table 2.4). All of the positive controls except hnRNPA1 demonstrated more than 2-fold enrichment in the pri-miRNA samples, suggesting specific binding over the background. A few proteins are conspicuously absent from this list, most surprisingly the components of microprocessor, DCGR8 and Drosha. This absence could be accounted for in the kinetics of pri-miRNA processing, with the assumption that binding of microprocessor leads to the immediate release of pre-miRNA species from the pri-miRNA. Additionally, since this was considered a pilot run, these proteins might be missing due to the limited depth of MS analysis

Positive Controls	Category	pri-miR17-92	pri-miR7	GAPDH
HNRNPA1	miRNA pathway	1.88	0.93	2.76
DDX5	miRNA pathway	32.35	31.75	1.09
DDX1	miRNA pathway	9.24	ND	ND
KSRP	miRNA pathway	ND	21.17	5.80
TUT4	miRNA pathway	46.22	19.05	0.90
RAN	miRNA pathway	13.87	16.93	ND
DDX17	miRNA pathway	2.77	3.39	0.70
PABPC4	Splicing	27.73	23.28	0.77
SF3B3	Splicing	27.73	25.40	ND
SFPQ	Splicing	9.24	19.05	ND
SF3B2	Splicing	ND	8.47	ND
U2AF2	Splicing	ND	8.47	ND

**Table 2.4 Enrichment Scores for positive controls in all samples**

Observed positive control proteins in mass spectrometry. ND - not detected in sample.

Another class of positive control genes relevant for this analysis is mRNA splicing factors since the GAPDH target site is exonic, while both pri-miRNA samples are intronic. As expected, many more splicing proteins were present in the pri-miRNAs samples compared to GAPDH. Overall the top enriched proteins in each pulldown are noted in Tables 2.5-2.7. Another encouraging feature of this dataset is the prevalence of RNA-biology related proteins in the top enriched genes as annotated.

	Protein Name	Enrichment Score	Annotation
1	PKM	78.32	
2	KCTD1	67.73	transcription factor
3	SHC1	52.92	
4	ENO1	46.57	
5	GAPDH	44.45	
6	HNRNPM	38.10	RNA binding
7	NONO	35.98	RNA binding
8	ACLY	33.87	
8	STIP1	33.87	
10	DHX9	31.75	RNA helicase
10	SYNCRIP	31.75	RNA binding
10	DDX5	31.75	RNA helicase
13	NCL	27.52	
14	MTHFD1	25.40	
15	ILF2	25.40	transcription factor
15	RUVBL2	25.40	DNA helicase
15	SF3B3	25.40	RNA splicing
18	HNRNPA1L2	23.28	RNA binding
18	LDHB	23.28	
18	PABPC4	23.28	RNA binding

**Table 2.5 Top protein binding candidates for pri-miR7**

Proteins with the top 20 enrichment scores are lists above. Housekeeping proteins (ribosomal, cytoskeletal, and histones) were removed for the list for clarity.

	Protein Name	Enrichment Score	Annotation
1	PKM	87.82	
1	HNRNPU	87.82	RNA binding
3	SYNCRIP	60.09	RNA binding
4	SND1	50.84	Transcription factor
5	UBB	46.22	
5	TUTase 4	46.22	RNA uridylation
7	ENO1	41.60	Transcription factor
8	TENM3	32.35	
8	QARS	32.35	
8	DDX3X	32.35	RNA helicase
8	NCL	32.35	
8	DDX3Y	32.35	RNA helicase
8	DDX5	32.35	RNA helicase
14	SUPT16H	27.73	Chromatin factor
14	CSDE1	27.73	RNA binding
14	TRIM28	27.73	Transcription factor
14	SF3B3	27.73	RNA splicing
14	ILF2	27.73	Transcription factor
14	PABPC4	27.73	RNA binding
20	RUVBL2	23.11	DNA helicase
20	MCM3	23.11	
20	DHX9	23.11	RNA helicase
20	DARS	23.11	

**Table 2.6 Top protein binding candidates for pri-miR17-92**

Proteins with the top 20 enrichment scores are lists above. Housekeeping genes (ribosomal, cytoskeletal, and histones) were removed for the list for clarity.

	Protein Name	Enrichment Score	Annotation
1	PRMT5	5.80	
1	DHX36	5.80	RNA helicase
1	PUF60	5.80	RNA binding
1	PARP1	5.80	Polymerase
1	PRDX2	5.80	
1	NAP1L1	5.80	
1	IPO5	5.80	
1	KHSRP	5.80	RNA binding
1	C1QBP	5.80	RNA binding
1	KCTD1	5.80	Transcription factor
1	AIFM1	5.80	
1	PFKM	5.80	
13	HNRNPA1	2.76	RNA binding
13	HNRNPD	2.26	RNA binding
15	PKM	2.15	

**Table 2.7 Top protein binding candidates for GAPDH**

Proteins with the top 20 enrichment scores are lists above. Housekeeping genes (ribosomal, cytoskeletal, and histones) were removed for the list for clarity.

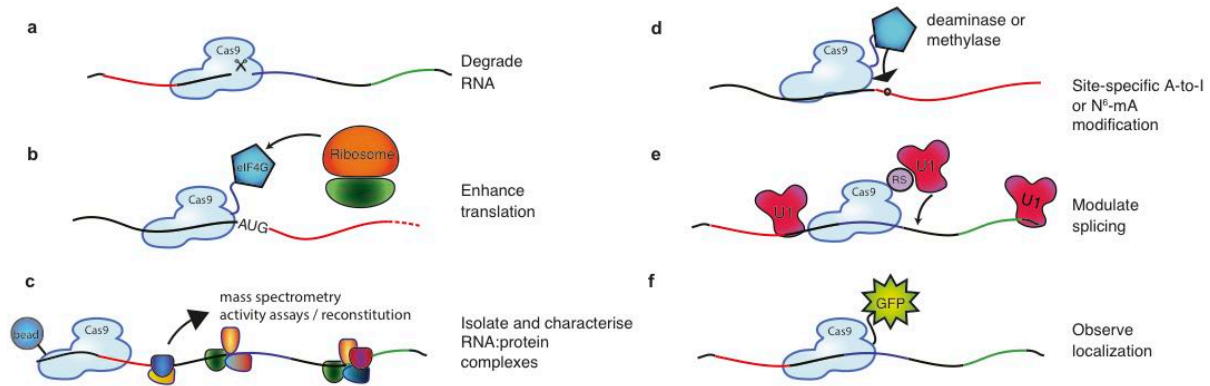
Overall, we have high confidence that desothiobiotin-RCas9 successfully isolated protein complexes bound to the specific RNAs of interest. For this particular dataset, further validation both *in vitro* and *in vivo* of binding partners will be required to confirm these interactions. In the future, it is recommended that additional low salt, no magnesium wash steps are added to the protocol to further reduce ribosomal binding to samples. While more complicated label-free mass spectrometry analysis could be performed on similar samples, this dataset provides proof-of-concept for RNA-binding partner identification by RCas9.

## 2.6 Discussion

Here we have demonstrated the ability to re-direct the dsDNA targeting capability of CRISPR/Cas9 for RNA-guided ssRNA binding and/or cleavage (RCas9). Programmable RNA recognition and cleavage has the potential to transform the study of RNA function much as site-specific DNA targeting is changing the landscape of genetic and genomic research (Mali et al., 2013) (Fig. 2.12). Although certain engineered proteins such as PPR proteins and Pumilio/FBF (PUF) repeats show promise as platforms for sequence-specific RNA targeting (Filipovska and Rackham, 2011; Mackay et al., 2011; Yagi et al., 2014; Yin et al., 2013), these strategies require re-designing the protein for every new RNA sequence of interest. While RNA interference has proven useful for manipulating gene regulation in certain organisms (Kim and Rossi, 2008), there has been a strong motivation to develop orthogonal nucleic acid-based RNA recognition systems, such as the CRISPR/Cas Type III-B Cmr complex (Hale et al., 2009; 2012; Spilman et al., 2013; Staals et al., 2013; Terns and Terns, 2014) and the atypical Cas9 from *Francisella novicida* (Sampson and Weiss, 2014; Sampson et al., 2013). In contrast to these systems, the molecular basis for RNA



recognition by RCas9 is now clear and requires only the design and synthesis of a matching gRNA and complementary PAMmer. The ability to recognize endogenous RNAs within complex mixtures with high affinity and in a programmable manner paves the way for direct transcript detection, analysis and manipulation without the need for genetically encoded affinity tags.



**Figure 2.12 Potential applications of RCas9 for untagged transcript analysis, detection, and manipulation.**

**a**, Catalytically-active RCas9 could be used to target and cleave RNA, particularly those for which RNAi-mediated repression/degradation is not possible. **b**, Tethering the eukaryotic initiation factor eIF4G to a catalytically inactive dRCas9 targeted to the 5' untranslated region of an mRNA could drive translation. **c**, dRCas9 tethered to beads could be used to specifically isolate RNA or native RNA:protein complexes of interest from cells for downstream analysis or assays including identification of bound protein complexes, probing of RNA structure under native protein-bound conditions, and enrichment of rare transcripts for sequencing analysis. **d**, dRCas9 tethered to RNA deaminase or N<sup>6</sup>-mA methylase domains could direct site-specific A-to-I editing or methylation of RNA, respectively. **e**, dRCas9 fused to a U1 recruitment domain (arginine- and serine-rich (RS) domain) could be programmed to recognize a splicing enhancer site and thereby promote the inclusion of a targeted exon. **f**, dRCas9 tethered to a fluorescent protein such as GFP could be used to observe RNA localization and transport in living cells.

# Chapter 3

## Discovery of RNA targeting single effector CRISPR systems

A portion of the content presented in this chapter has been previously published as part of the following research article: Alexandra East-Seletsky, Mitchell R. O’Connell, Spencer C. Knight, David Burstein, Jamie H. D. Cate, Robert Tjian, and Jennifer A. Doudna, (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538, 270–273.

Alexandra Seletsky and Mitchell O’Connell were designated co-first authors. Alexandra Seletsky, Mitchell O’Connell, Spencer Knight conceived the study and designed experiments with input from Jamie Cate, Robert Tjian, and Jennifer Doudna. David Burstein performed bioinformatic analyses. Alexandra Seletsky, and Mitchell O’Connell, executed all experimental work with assistance from Spencer Knight. All authors discussed the data and wrote the manuscript.



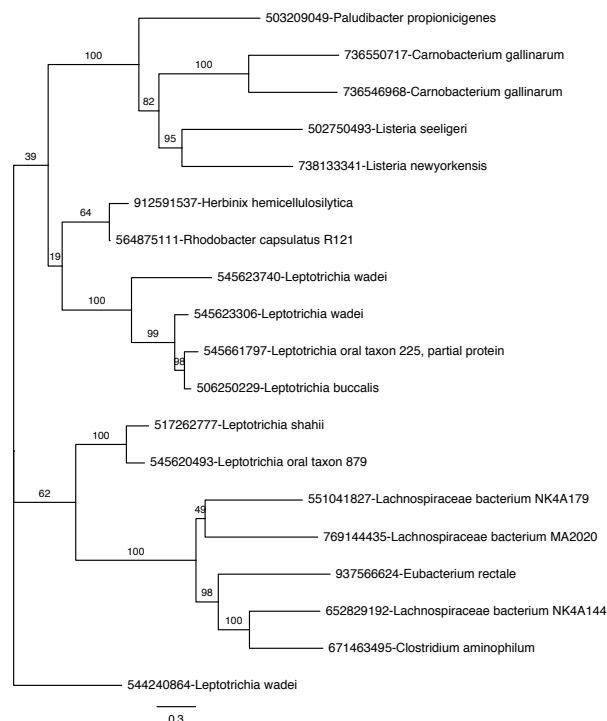
### 3.1 Summary and Introduction

Bacterial adaptive immune systems employ CRISPRs (clustered regularly interspaced short palindromic repeats) and CRISPR-associated (Cas) proteins for RNA-guided nucleic acid cleavage (van der Oost et al., 2014; Wright et al., 2016). Although generally targeted to DNA substrates (Brouns et al., 2008; Garneau et al., 2010; Marraffini and Sontheimer, 2008), the Type III and Type VI CRISPR systems direct interference complexes against single-stranded RNA (ssRNA) substrates (Abudayyeh et al., 2016; Hale et al., 2009; Samai et al., 2015; Staals et al., 2013). In Type VI systems, the single-subunit C2c2 protein functions as an RNA-guided RNA endonuclease (Abudayyeh et al., 2016; Shmakov et al., 2015). How this enzyme acquires mature CRISPR RNAs (crRNAs) essential for immune surveillance and its mechanism of crRNA-mediated RNA cleavage remain unclear. Here we show that C2c2 possesses a unique ribonuclease activity responsible for CRISPR RNA maturation that is distinct from its RNA-activated ssRNA-degradation activity. These dual ribonuclease functions are chemically and mechanistically different from each other and from the crRNA-processing behavior of the evolutionarily unrelated CRISPR enzyme Cpf1 (Fonfara et al., 2016). We show that the two ribonuclease activities of C2c2 enable multiplexed processing and loading of guide RNAs that in turn allow for sensitive cellular transcript detection.

### 3.2 Methods

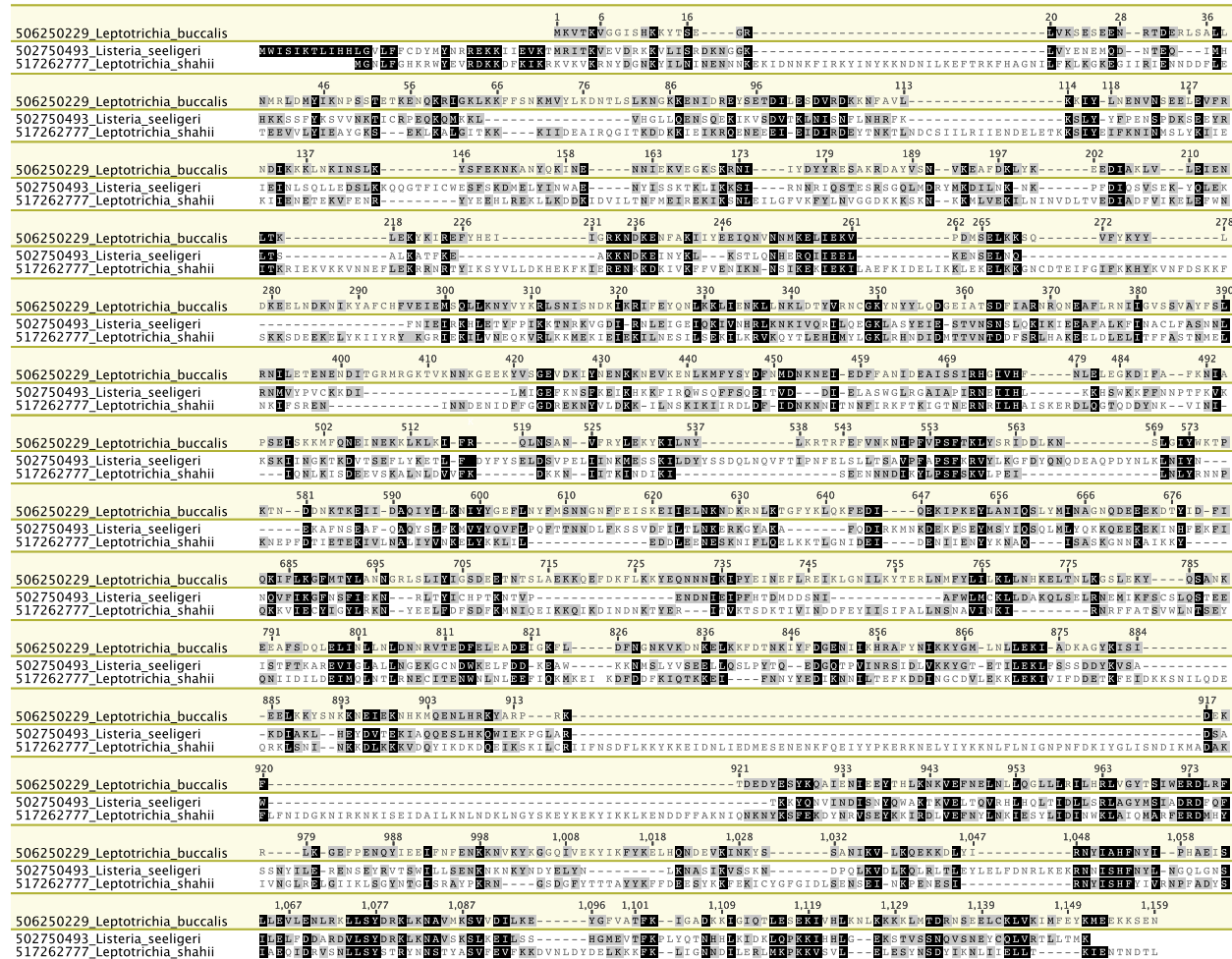
#### 3.2.1 C2c2 phylogenic and candidate selection

C2c2 maximum-likelihood phylogenies were computed using RAxML (Stamatakis, 2014b) with the PROTGAMMALG evolutionary model and 100 bootstrap samplings. Sequences were aligned by MAFFT with the 'einsi' method (Kato and Standley, 2013b).



### Figure 3.1 Phylogenetic tree of C2c2 family

Maximum-likelihood phylogenetic reconstruction of C2c2 proteins. Leaves include GI protein numbers and organism of origin; bootstrap support values, out of 100 resamplings, are presented for inner split. Scale is in substitutions per site.



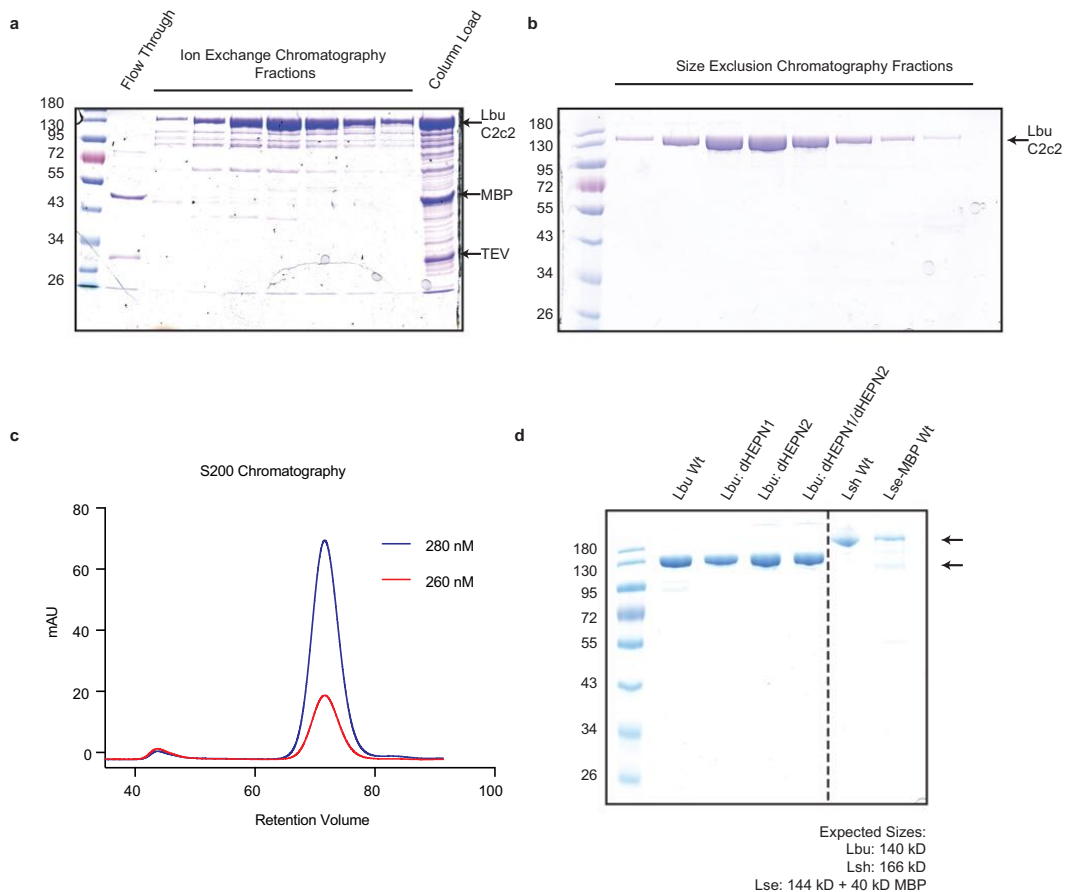
### Figure 3.2 Alignment of protein sequences from three C2c2 homologs

Multiple sequence alignment of the three analyzed homologs of C2c2; coordinates are based on LbuC2c2.

#### 3.2.2 C2c2 protein production and purification

Expression vectors for protein purification were assembled using synthetic gBlocks ordered from Integrated DNA Technologies. The codon-optimized C2c2 genomic sequence was N-terminally tagged with a His<sub>6</sub>-MBP-TEV cleavage site, with expression driven by a T7 promoter. Mutant proteins were cloned via site-directed mutagenesis of wild-type C2c2 constructs. Expression vectors were transformed into Rosetta2 *E. coli* cells grown in 2xYT broth at 37 °C. *E. coli* cells were induced during log phase with 0.5 M IPTG, and the temperature was reduced to 16 °C for overnight expression of His-MBP-C2c2. Cells were subsequently harvested, resuspended in lysis buffer (50 mM Tris-HCl pH 7.0, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 0.5mM PMSF, and EDTA-

free protease inhibitor (Roche)) and lysed by sonication, and the lysates were clarified by centrifugation. Soluble His-MBP-C2c2 was isolated over metal ion affinity chromatography, and protein-containing eluate was incubated with TEV protease at 4 °C overnight while dialyzing into ion exchange buffer (50 mM Tris-HCl pH 7.0, 250 mM KCl, 5% glycerol, 1 mM TCEP) in order to cleave off the His<sub>6</sub>-MBP tag. Cleaved protein was loaded onto a HiTrap SP column and eluted over a linear KCl (0.25-1.5M) gradient. Cation exchange chromatography fractions were pooled and concentrated with 30 kD cutoff concentrators (Thermo Fisher). The C2c2 protein was further purified via size-exclusion chromatography on an S200 column and stored in gel filtration buffer (20 mM Tris-HCl pH 7.0, 200 mM KCl, 5% glycerol, 1 mM TCEP) for subsequent enzymatic assays. Expression plasmids are deposited with Addgene. Representative PAGE analysis of purification fractions and size-exclusion chromatography trace is shown in Fig 3.3.



### Figure 3.3 Purification and Production of C2c2

All C2c2 homologs were expressed in *E. coli* as His-MBP fusions and purified by a combination of affinity, ion exchange and size exclusion chromatography. The Ni<sup>+</sup> affinity tag was removed by incubation with TEV protease. Representative SDS-PAGE gels of chromatography fractions are shown in (a, b). c, The chromatogram from Superdex 200 (16/60) column demonstrating that C2c2 elutes as a single peak, devoid of nucleic acid. d, SDS PAGE analysis of purified proteins used in this manuscript.

### 3.2.3 Generation of RNA

All RNAs used in this study were transcribed *in vitro* except for crRNA AES461 which was ordered synthetically (Integrated DNA Technologies) [see Table 3.1]. *In vitro* transcription reactions were performed as previously described with the following modifications: the T7 polymerase concentration was reduced to 10 µg/mL, and the UTP concentration was reduced to 2.5 mM (Sternberg et al., 2012). Transcriptions were incubated at 37°C for 1-2 hr to reduce non-template addition of nucleotides and quenched via treatment with DNase I at 37°C for 0.5-1 hr. Transcription reactions were purified by 15% denaturing polyacrylamide gel electrophoresis (PAGE), and all RNAs were resuspended in cleavage buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, and 5% glycerol). For radioactive experiments, 5' triphosphates were removed by calf intestinal phosphate (New England Biolabs) prior to radiolabeling and ssRNA substrates were then 5'-end labeled using T4 polynucleotide kinase (New England Biolabs) and [<sup>32</sup>P]-ATP (Perkin Elmer) as described previously (Sternberg et al., 2012).

Oligo Name	Sequence
Lbu_pre-crRNA_A_SCK314	GGAUUUAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lse_pre-crRNA_B_AES484	GGUAAGAGACUACCUCUAUUAUGAAAGAGGACUAAAACCAAACAUGAUCUGGGUCAUC
Lsh_pre-crRNA_A_SCK339	GGAUUUAGACCACCCCAAAUUAUCGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-crRNA_invert_SCK321	GGAUUUAGACCAGGGGAAGUAAAACCCACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-crRNA_5stem_SCK331	GGAUUUAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-crRNA_7bubble_SCK334	GGAUUUAGACCACCCCAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-crRNA_5bubble_SCK335	GGAUUUAGACCACCCCAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-crRNA_3stem_SCK342	GGAUUUAGACCACCCCAAAAAUGAAGGGGCACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut1_AES497	GGCGUUAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut2_AES496	GGAGCUAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut3_AES495	GGAUCCAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut4_AES477	GGAUUCGAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut5_AES482	GGAUUUACCCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut6_AES478	GGAUUUAAUCCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre-cr_5'_mut7_AES480	GGAUUUAGAAAACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_5'_mut8_AES498	GGAUUUAGACCGCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_stem_mut1_AES502	GGAUUUAGACCAGCCCAAAAAUGAAGGGGCACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_stem_mut2_AES501	GGAUUUAGACCACCGCAAAAAUGAAGCGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_stem_mut3_AES500	GGAUUUAGACCACCAAAAAUGAAGGUGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_stem_mut4_AES499	GGAUUUAGACCACCAAAAAUGAAGGUGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_stem_mut5_AES504	GGAUUUAGACCACUCCAAAAAUGAAGGAGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_3'_mut1_AES505	GGAUUUAGACCACCCCAAAAAUGAAGGGGCAUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_3'_mut2_AES506	GGAUUUAGACCACCCCAAAAAUGAAGGGGACGCAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_3'_mut3_AES507	GGAUUUAGACCACCCCAAAAAUGAAGGGGACUAGCACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_3'_mut4_AES508	GGAUUUAGACCACCCCAAAAAUGAAGGGGACUAAAGUAGGGGCAGAGAUGAUGACCCU
crLbu_A_GG_AES432	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
crLbu_B_AES451	GGCCACCCCAAAAAUGAAGGGGACUAAAACCAAACAUGAUCUGGGUCAUC
A_0_target_AES450	GGCACCCCGCAGGGGGGAGCCAAAAGGGUCAUCAUCUCUGCCCCACAGCAGAAGCC
B_target_AES452	GGGAACCCCAAGGCCAACCGCGAGAAGAUGACCCAGAUCUUGUUGAGACCUUCAACAC
crLbu_Lambda2_AES453	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGUGAUAGUGGAAUGCCAUG
crLbu_Lambda3_MOC410	GGCCACCCCAAAAAUGAAGGGGACUAAAACACUGGUGAACUUCGGAUAGUG
crLbu_Lambda4_MOC411	GGCCACCCCAAAAAUGAAGGGGACUAAAACACAGAUUAGCCUGGUGGUUC
Lambda2_target_MOC28	GGCUCAUUUUGACAGCGGUCUAGGCAUUGCCACUUAUCACUGGCAUCCUCCACUC
Lambda3_target_MOC36	GGAAUCAUUCACACCCGCACUUAUCGGAAGUUCACCAGCCAGCCGCAGCACGUU
Lambda4_target_MOC37	GGCAAUAAAAUUGCGCCGCCUGAACCACCAGGCUAUUAUCUGCCACUCAUUGUUGUGA
crLbu_betaActin_1_AES451	GGCCACCCCAAAAAUGAAGGGGACUAAAACCAAACAUGAUCUGGGUCAUC
pre-crLbu_dimer_SCK324	GGAUUUAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGGGGCAGAGAUGAUGACCCUA
	UUUAGACCACCCCAAAAAUGAAGGGGGACUAAAACAGUGAUUAGUGGAAUGCCAUG

crLbu_lambda2_SCK315	GGAUUUAGACCACCCCAAAAAUGAAGGGGACUAAAACAGUGAUUAGUGGAAUGCCAUG
Lbu_pre_cr_5' 4mer1_AES481	GGAUUUAAUAAACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_5' 4mer2_AES479	GGAUUCGAUCCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_5' 4mer3_SCK343	GGAUUUAGGAAGCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
Lbu_pre_cr_5' 4mer4_AES503	GGAUUUAGACCAGGCCAAAAAUGAAGGGCCACUAAAACAGGGGCAGAGAUGAUGACCCU
crLbu_GuideWalk1_SCK302	GGCCACCCCAAAAAUGAAGGGGACUAAAACAUUUUGGCUCSCCCUGCAAUAG
crLbu_GuideWalk2_SCK303	GGCCACCCCAAAAAUGAAGGGGACUAAAACACCCUUUUUGGCUCSCCCUGCAA
crLbu_GuideWalk3_SCK304	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGAUAGCCUUUUUGGCUCSCCCUG
crLbu_GuideWalk4_SCK305	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGAUAGCCUUUUUGGCUCSCCC
crLbu_GuideWalk5_SCK306	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGCAGAGAUGAUGACCCUUUUUGGC
crLbu_GuideWalk6_SCK307	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGGGGGCAGAGAUGAUGACCCUUUU
crLbu_GuideWalk7_SCK308	GGCCACCCCAAAAAUGAAGGGGACUAAAACACAGAGGGGGCAGAGAUGAUGACCC
crLbu_GuideWalk8_SCK309	GGCCACCCCAAAAAUGAAGGGGACUAAAACAUCAAGCAGAGGGGGCAGAGAUGAUG
A.1_target_U_MOC279	GGCUCAUUUGCAGGGGGGAGCCAAAAGGGUCAUCAUCUCUGSCCCUCUGCUGAUGCC
A.2_target_70nt_AES447	GGCCUGACUGCUCUCAUUUGCAGUUGGGAGCCAAAAGGGUCAUCAUCUCUGSCCCUC
A.3_target_80nt_AES448	GGACCUUGAUUCCUGACUGCUCUCAUUUGCAGUUGGGAGCCAAAAGGGUCAUCAUCU
A.4_5'_ts_shift_AES449	GGCACCCCGCAGGGUUUAGCCAAAAGGGUCAUCAUCUCUGSCCCUCUGCUGAUGCC
crLbu_A_16nt_trunc_SCK282	GGCCACCCCAAAAAUGAAGGGGACUAAAACACAGAGAUGAUGACCCU
crLbu_A_24nt_ext_SCK283	GGCCACCCCAAAAAUGAAGGGGACUAAAACAGAGGGGGCAGAGAUGAUGACCCU
crLbu_A_mature_GA_SCK340	GACCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
crLbu_A_mature_GGGA_SCK341	GGGACCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
crLbu_A_mature_CCA_AES461	CCACCCCAAAAAUGAAGGGGACUAAAACAGGGGCAGAGAUGAUGACCCU
T7 Forward (DNA)	TAATACGACTCACTATAGG

**Table 3.1 Oligonucleotides used in study**

### 3.2.4 Pre-crRNA processing assays

Pre-crRNA cleavage assays were performed at 37 °C in RNA processing buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10 µg/mL BSA, 100 µg/mL tRNA, 0.01% Igepal CA-630 and 5% glycerol) with a 100-fold molar excess of C2c2 relative to 5'-labeled pre-crRNA (final concentrations of 100 nM and <1 nM, respectively). Unless otherwise indicated, reaction was quenched after 60 min with 1.5X RNA loading dye (100% formamide, 0.025 w/v% bromophenol blue, and 200 µg mL<sup>-1</sup> heparin). After quenching, reactions were denatured at 95 °C for 5 min prior to resolving by 12% or 15% denaturing PAGE (0.5X TBE buffer). Metal dependence of the reaction was tested by addition of EDTA or EGTA to reaction buffer at concentrations varying from 10-100 mM. Bands were visualized by phosphorimaging and quantified with ImageQuant (GE Healthcare). The percent cleavage was determined as the ratio of the product band intensity to the total intensity of both the product and uncleaved pre-crRNA bands and normalized for background within each measured substrate using ImageQuant TL Software (GE Healthcare) and fit to a one phase exponential association using Prism (GraphPad).

### 3.2.5 Product Size Mapping and 3' end moiety identification

Cleavage product length was determined biochemically by comparing gel migration of product bands to alkaline hydrolysis and RNase T1 digestion ladders using the RNase T1 Kit from Ambion. For hydrolysis ladder, 15 nM full-length RNA substrates were incubated at 95°C in 1X alkaline hydrolysis buffer (Ambion) for 5 min. Reactions were quenched with 1.5X RNA loading buffer, and cooled to -20°C to immediately stop hydrolysis. For RNase T1 ladder, 15 nM full length RNA substrates were unfolded in 1X

RNA sequencing buffer (Ambion) at 65°C. Reactions were cooled to ambient temperature, and then 1 U of RNase T1 (Ambion) was added to reaction. After 15 min, reactions were stopped by phenol-chloroform extraction and 1.5X RNA loading buffer was added for storage. Hydrolysis bands were resolved in parallel to cleavage samples on 15% denaturing PAGE and visualized by phosphorimaging. For 3' end moiety identification, products from the processing reaction were incubated with 10 U of T4 polynucleotide kinase (New England Biolabs) for 1 hr at 37°C in processing buffer. Reactions were quenched with 1.5X RNA loading buffer, resolved on 20% denaturing PAGE and visualized by phosphorimaging.

### 3.2.6 Small RNA sequencing analysis

RNA reads from Shmakov *et al.* (2015) were downloaded from SRA runs SRR3713697, SRR3713948, and SRR3713950. The paired-end reads were locally mapped to the reference sequences using Bowtie2 (Langmead and Salzberg, 2012) with the following options: "--reorder --very-fast-local --local". The mapping was then filtered to retain only alignments that contained no mismatch using mapped.py (<https://github.com/christophertbrown/mapped>) with the "-m 0 -p both" options. BAM file of the resulting mapping are in the supplementary files for this manuscript. Read coverage was visualized using Geneious and plotted using Prism (GraphPad).

### 3.2.7 Target cleavage assays

Target cleavages assays were performed at 25 °C or 37 °C in cleavage buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, and 5% glycerol). crRNA guides were pre-folded by heating to 65 °C for 5 min and then slowly cooling to ambient temperature in cleavage buffer. C2c2:crRNA complex formation was performed in cleavage buffer, generally at a molar ratio of 2:1 protein to crRNA at 37 °C for 10 min, prior to adding 5'-end labeled target and/or other non-radiolabeled RNA target substrates. Unless otherwise indicated, final concentrations of protein, guide, and targets were 100 nM, 50 nM, and <1 nM respectively for all reactions. Reactions were quenched with 1.5X RNA loading dye and resolved by 15% denaturing PAGE (0.5X TBE buffer). Bands were visualized by phosphorimaging and quantified with ImageQuant (GE Healthcare). The percent cleavage was determined as the ratio of total banding intensity for all shorter products relative to the uncleaved band and normalized for background within each measured substrate using ImageQuant TL Software (GE Healthcare) and fit to a one phase exponential association using Prism (GraphPad).

### 3.2.8 crRNA filter-binding assays

Filter binding assays was carried out in RNA processing buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10 µg/mL BSA, 100 µg/mL yeast tRNA, 0.01% Igepal CA-630 and 5% glycerol). LbuC2c2 was incubated with radiolabeled crRNA (<0.1 nM) for 1hr at 37°C. Tufryn, Protran and Hybond-N+ were assembled onto a dot-blot apparatus in the order listed above. The membranes were washed twice with 50µL Equilibration Buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub> and 5% glycerol) before the sample was applied to the membranes. Membranes were again washed with 50 µL Equilibration Buffer, dried and visualized by phosphorimaging. Data were quantified with ImageQuant TL Software (GE Healthcare) and fit to a binding isotherm using Prism



(GraphPad Software). All experiments were carried out in triplicate. Dissociation constants and associated errors are reported in the figure legends.

### 3.2.9 Electrophoretic mobility-shift assays

In order to avoid the dissociation of the LbuC2c2-dHEPN1/dHEPN2: crRNA complex at low concentrations during ssRNA-binding experiments, binding reactions contained a constant excess of LbuC2c2-dHEPN1/dHEPN2 (200 nM), and increasing concentrations of crRNA-A and < 0.1 nM target ssRNA. Assays were carried out in C2c2 EMSA buffer (20 mM HEPES pH 6.8, 50 mM KCl, 10 µg/mL BSA, 100 µg/mL yeast tRNA, 0.01% Igepal CA-630 and 5% glycerol). LbuC2c2-crRNA-A complexes were pre-formed as described above for 10 min at 37°C before the addition of 5'-radiolabelled ssRNA substrate and a further incubation for 45 min at 37°C. Samples were then resolved by 8% native PAGE at 4°C (0.5X TBE buffer). Gels were imaged by phosphorimaging, quantified using ImageQuant TL Software (GE Healthcare) and fit to a binding isotherm using Prism (GraphPad Software). All experiments were carried out in triplicate. Dissociation constants and associated errors are reported in the figure legends.

### 3.2.10 Fluorescent RNA detection assay

LbuC2c2:crRNA complexes were preassembled by incubating 1µM of Lbu-C2c2:C2c2 with 500 nM of crRNA for 10 min at 37°C. These complexes were then diluted to 100nM LbuC2c2: 50 nM crRNA-λ2 in RNA processing buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10 µg/mL BSA, 10 µg/mL yeast tRNA, 0.01% Igepal CA-630 and 5% glycerol) in the presence of 185 nM of RNAase-Alert substrate (Thermo-Fisher), 100 ng of HeLa total RNA and increasing amounts of target 60 nt ssRNA (0-1 nM). These reactions were incubated in a fluorescence plate reader for up to 120 min at 37°C with fluorescence measurements taken every 5 min ( $\lambda_{ex}$ : 485 nm;  $\lambda_{em}$ : 535 nm). Background-corrected fluorescence values were obtained by subtracting fluorescence values obtained from reactions carried out in the absence of target ssRNA. Maximal fluorescence was measured by incubating 50 nM RNaseA with 185 nM of RNAase-Alert substrate. For measurement of crRNA-ACTB mediated LbuC2c2 activation by *beta-actin* mRNA in human total RNA, LbuCas9:crRNA complexes were preassembled by incubating 1µM of LbuC2c2 with 500 nM of crRNA-ACTB for 10 min at 37°C and reactions were carried out in the conditions above in the presence of increasing amounts (0-1 µg) of either HeLa cell total RNA or E. Coli total RNA (as a negative control). These reactions were incubated in a fluorescence plate reader for up to 180 min at 37°C with fluorescence measurements taken every 5 min ( $\lambda_{ex}$ : 485 nm;  $\lambda_{em}$ : 535 nm). Background-corrected fluorescence values were obtained by subtracting fluorescence values obtained from reactions carried out in the absence of target ssRNA. For coupled pre-crRNA processing and RNA detection assays, LbuCas9-crRNA complexes were preassembled by incubating 1µM of LbuC2c2 with 500 nM of pre-crRNA-A-λ2 for 20 min at 37°C and reactions carried out as described above in the presence of increasing amounts of ssRNA A and ssRNA λ2 (0-1 nM each). In each case, error bars represent the standard deviation from three independent experiments.

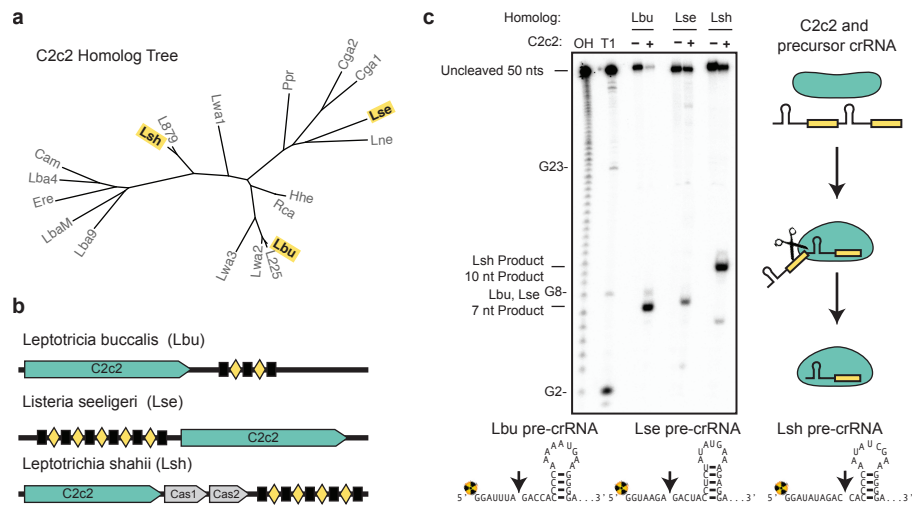
### 3.2.11 Background cleavage in total RNA

LbuC2c2:crRNA $\lambda$ 4 complexes were assembled as previously described for fluorescence RNA detection assay. Complexes were incubated in RNA processing buffer in the presence of 3  $\mu$ g total RNA with and without 10 nM  $\lambda$ 4 ssRNA target. After 2 hr, RNA was isolated by trizol extraction and ethanol precipitation. The RNA fragment size distribution of resuspended samples was resolved using Small RNA Analysis Kit (Agilent) on a Bioanalyzer 2100 (Agilent) using the manufacturer's protocol. Fluorescent intensity curves were normalized in Prism for curve overlay (GraphPad Software).

## 3.3 Results

### 3.3.1 C2c2 catalyzes crRNA maturation for Type VI-A systems

The essential first step of CRISPR immune surveillance requires processing of precursor crRNA transcripts (pre-crRNAs), consisting of repeat sequences flanking viral spacer sequences, into individual mature crRNAs that each contain a single spacer (Charpentier et al., 2015; Hochstrasser and Doudna, 2015; Li, 2015). CRISPR systems employ three known mechanisms to produce mature crRNAs: a dedicated endonuclease (e.g., Cas6 or Cas5d in Type I and III systems) (Carte et al., 2008; Haurwitz et al., 2010; Nam et al., 2012b), coupling of a host endonuclease (e.g., RNase III with a trans-activating crRNA (tracrRNA) in Type II systems) (Deltcheva et al., 2011), or a ribonuclease activity intrinsic to the effector enzyme itself (e.g., Cpf1, Type V systems) (Fonfara et al., 2016).



**Figure 3.4 C2c2 proteins process precursor crRNA transcripts to generate mature crRNAs**

**a**, Maximum-likelihood phylogenetic tree of C2c2 proteins. Homologs used in this study are highlighted in yellow. **b**, Diagram of the three Type VI CRISPR loci used in this study. Black rectangles denote repeat elements, yellow diamonds denote spacer sequences. Cas1 and Cas2 are only found in the genomic vicinity of LshC2c2. **c**, C2c2-mediated cleavage of pre-crRNA derived from the LbuC2c2, LseC2c2 and LshC2c2

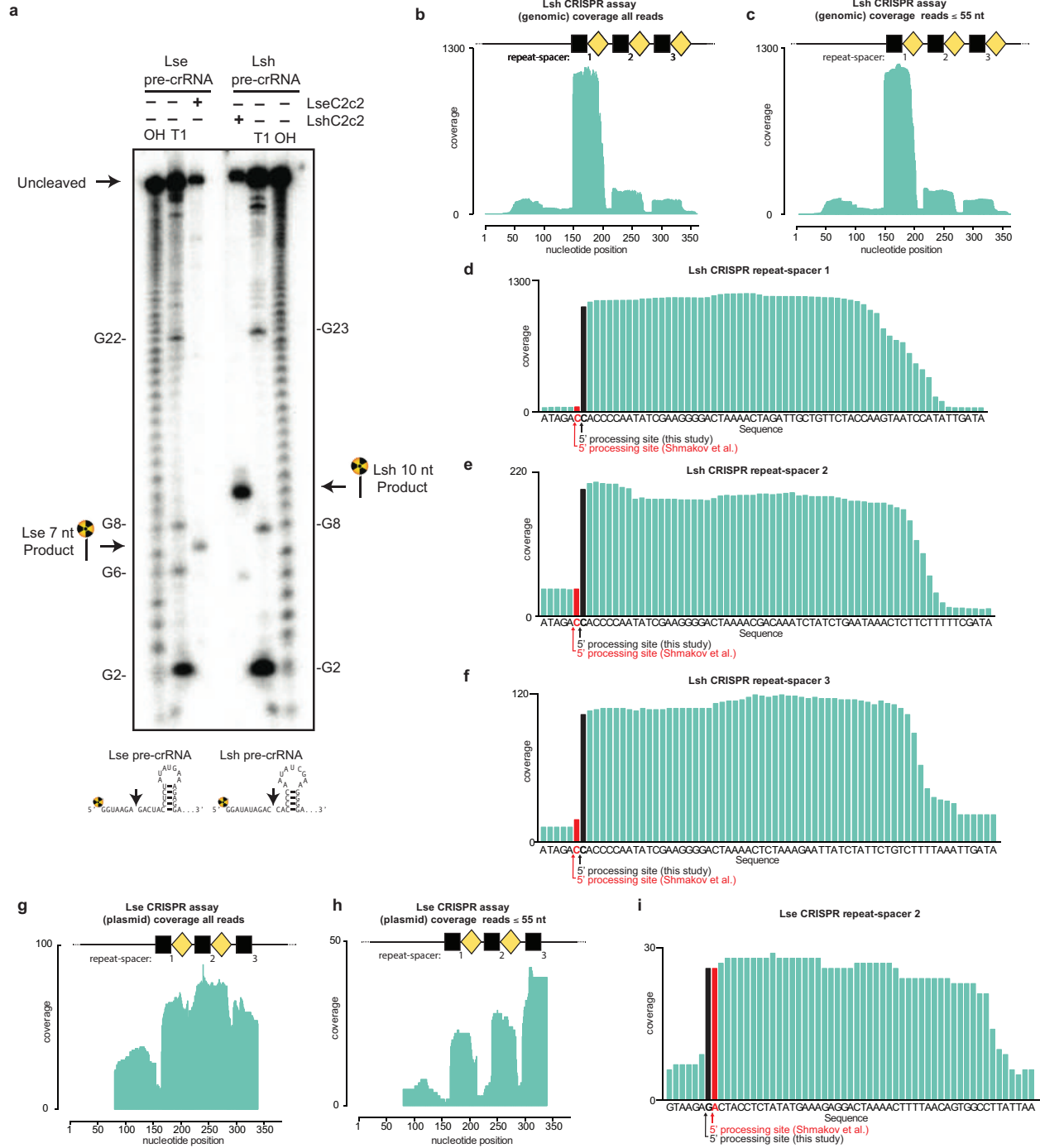


CRISPR repeat loci. OH: alkaline hydrolysis ladder; T1: RNase T1 hydrolysis ladder; processing cleavage reactions were performed with 100 nM C2c2 and <1 nM pre-crRNA. Schematic of cleavage is depicted on right, and predicted pre-crRNA secondary structures are diagrammed below, with arrows indicating the mapped C2c2 cleavage sites

Since Type VI CRISPR loci lack an obvious Cas6 or Cas5d-like endonuclease or tracrRNA (Shmakov et al., 2015), we wondered whether C2c2 itself might possess pre-crRNA processing activity, and if so, whether the mechanism would be distinct from Cpf1, an unrelated class 2 CRISPR effector recently demonstrated to process pre-crRNAs (Fonfara et al., 2016). Using purified recombinant C2c2 protein homologs from three distinct branches of the C2c2 protein family (Fig. 3.1-3.4), we found that all three C2c2 enzymes cleave 5'-end radiolabeled pre-crRNA substrates consisting of a full-length consensus repeat sequence and a 20 nucleotide (nt) spacer sequence (Fig. 3.4c). We mapped the cleavage site for each pre-crRNA:C2c2 homolog pair, revealing that processing occurs at a position either two or five nucleotides upstream of the predicted repeat-sequence hairpin structure, depending on the C2c2 homolog (Fig. 3.4c and 3.5a). Surprisingly, our biochemically mapped 5'-cleavage sites do not agree with previously reported cleavage sites for *Leptotrichia shahii* (LshC2c2) or *Listeria seeligeri* (LseC2c2) pre-crRNAs (Shmakov et al., 2015). Our own analysis of Shmakov *et al.*'s RNA sequencing data set indicates agreement of the *in vivo* cleavage site with the *in vitro* site reported here (Fig. 3.5b-i). Furthermore, cleavage assays using C2c2 from *Leptotrichia buccalis* (LbuC2c2) and a larger pre-crRNA comprising a tandem hairpin-repeat array resulted in two products resulting from two separate cleavage events (Fig. 3.6a), consistent with a role for C2c2 in processing precursor crRNA transcripts generated from Type VI CRISPR loci.

### 3.3.2 Substrate requirements and mechanism of pre-crRNA processing by C2c2

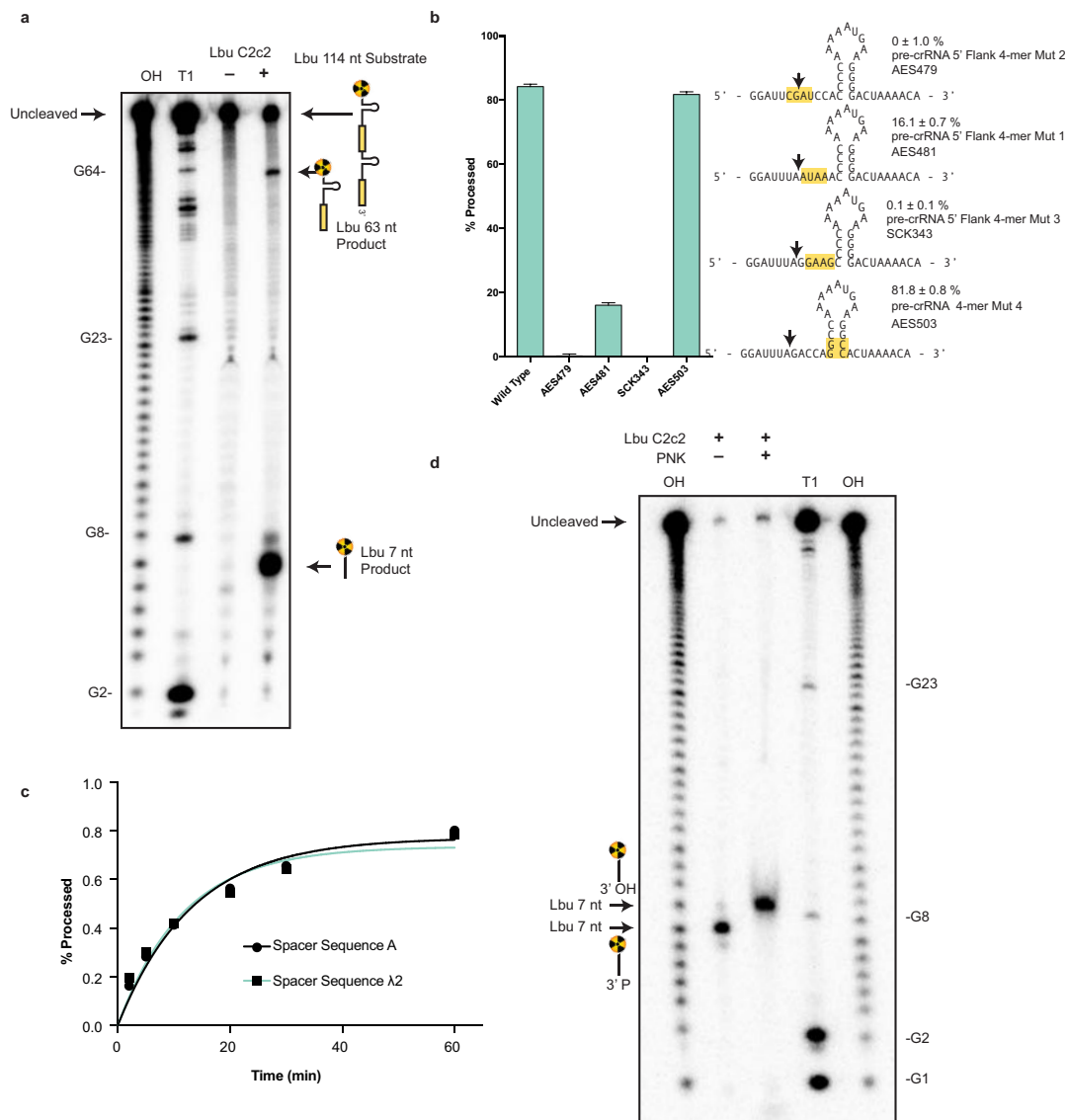
To understand the substrate requirements and mechanism of C2c2 guide RNA processing, we generated pre-crRNAs harboring mutations in either the stem loop or the single-stranded flanking regions of the consensus repeat sequence and tested their ability to be processed by LbuC2c2 (Fig. 3.7). We found that C2c2-catalyzed cleavage was attenuated upon altering the length of the stem in the repeat region (Fig. 3.7a). Inversion of the stem loop or reduction of the loop length also reduced C2c2's processing activity, while contiguous 4-nt mutations including or near the scissile bond completely abolished it (Fig. 3.6b). A more extensive mutational analysis of the full crRNA repeat sequence revealed two distinct regions on either side of the hairpin with marked sensitivity to base changes (Fig. 3.7b). By contrast, there was no dependence on the spacer sequence for kinetics of processing (Fig. 3.6c). This sensitivity to both flanking regions of the hairpin is reminiscent of the sequence and structural motifs required by many Cas6 and Cas5d enzymes (Charpentier et al., 2015; Fonfara et al., 2013; Li, 2015). In contrast, Cpf1 does not have any dependence on the 3' hairpin flanking region, as the variable spacer region abuts the hairpin stem (Fonfara et al., 2016).



**Figure 3.5 Mapping of pre-crRNA processing by C2c2 *in vitro* and *in vivo*.**

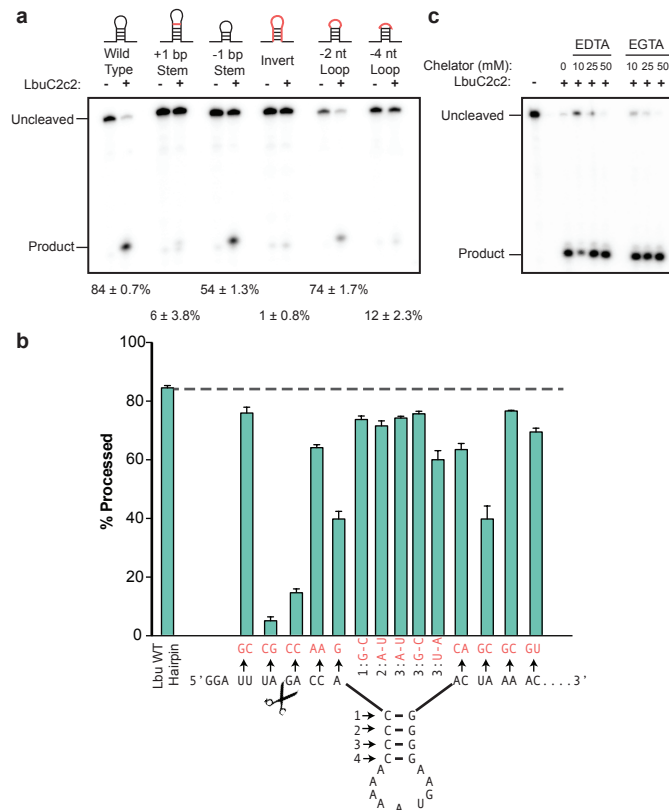
**a**, Cleavage site mapping of LseC2c2 and LshC2c2 cleavage of a single cognate pre-crRNA array. OH: alkaline hydrolysis ladder; T1: T1 RNase hydrolysis ladder. Cleavage reactions were performed with 100 nM C2c2 and <1 nM pre-crRNA. **b-i**, Re-analysis of LshC2c2 (**b-f**) and LseC2c2 (**g-i**) CRISPR array RNA sequencing experiments from Shmakov *et al.* (Fig. S7 and Fig. 5, respectively). All reads (**b,g**) and filtered reads (55 nt or less; as per original Shmakov *et al.* analysis; **c,h**) were stringently aligned to each CRISPR array using Bowtie2 (see Methods). Detailed views of individual CRISPR

repeat-spacers are shown for Lsh (d-f) and Lse (i). Differences in 5' end pre-crRNA processing are indicated by arrows below each sequence. BAM alignment files of our analysis are available in Supplementary Materials. This mapping clearly indicates that the 5' ends of small RNA sequencing reads generated from Lsh pre-crRNAs map to a position 2 nts from the base of the predicted hairpin, in agreement with our *in vitro* processing data (a). This pattern holds for all mature crRNAs detected from both native expression in *L. shahii* and heterologous expression in *E. coli* (data not shown, BAM file available in supplementary methods). Unfortunately, the LseC2c2 crRNA sequencing data (used in g-i) is less informative due to low read depth, and each aligned crRNA exhibits a slightly different 5' end with little obvious uniformity. The mapping for one of the processed repeats (repeat-spacer 2; i) is in agreement with our data but only with low confidence due to the insufficient read depth.



**Figure 3.6** Further investigations into the substrate requirements and mechanism of pre-crRNA processing by C2c2.

**a**, Cleavage site mapping of LbuC2c2 cleavage of a tandem pre-crRNA array. OH: alkaline hydrolysis ladder; T1: T1 RNase hydrolysis ladder. Cleavage reactions were performed with 100 nM LbuC2c2 and <1 nM pre-crRNA. A schematic of cleavage products is depicted on right, with arrows indicating the mapped C2c2 cleavage products. **b**, LbuC2c2 4-mer mutant pre-crRNA processing data demonstrating the importance of the 5' single-stranded flanking region for efficient pre-crRNA processing. Percentage of pre-crRNA processing was measured after 60 min (mean  $\pm$  s.d.,  $n = 3$ ). **c**, Representative LbuC2c2 pre-crRNA cleavage time-course demonstrating that similar rates of pre-crRNA processing occur independent of crRNA spacer sequence pseudo-first-order rate constants ( $k_{obs}$ ) (mean  $\pm$  s.d.) are  $0.07 \pm 0.04 \text{ min}^{-1}$  and  $0.08 \pm 0.04 \text{ min}^{-1}$  for spacer A and spacer  $\lambda_2$ , respectively. **d**, End group analysis of cleaved RNA by T4 polynucleotide kinase (PNK) treatment. Standard processing assay conditions were used to generate cleavage product, which was then incubated with PNK for 1 hr to remove any 2', 3'-cyclic phosphates/3' monophosphates. Retarded migration of band indicates removal of the charged, monophosphate from the 3' end of radiolabeled 5' product.



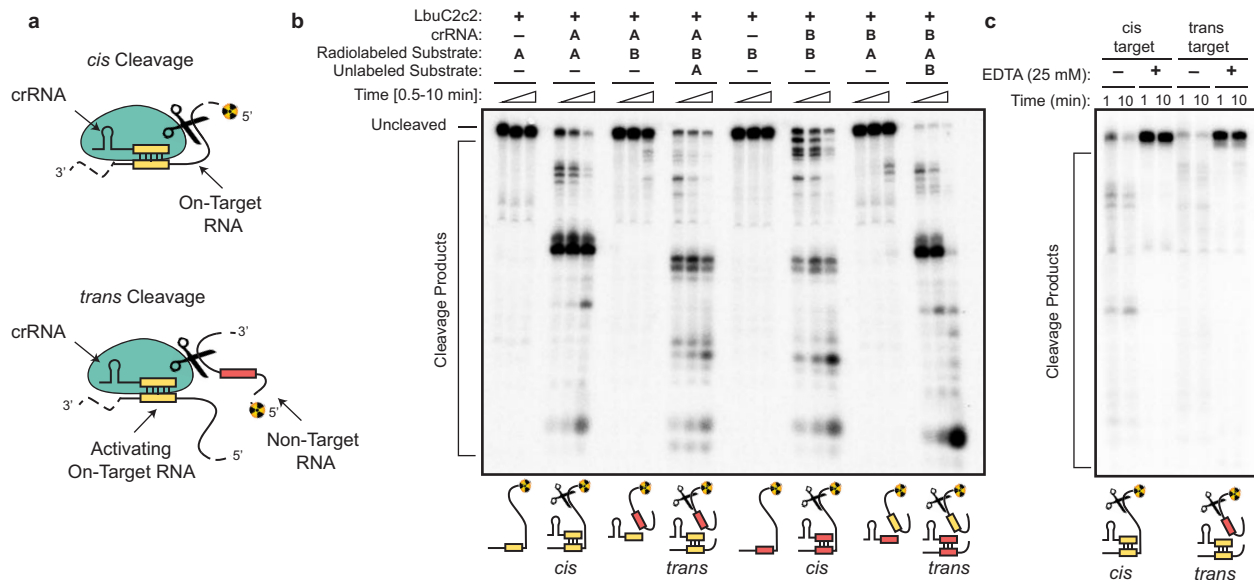
**Figure 3.7 LbuC2c2 mediated crRNA biogenesis depends on both structure and sequence of CRISPR repeats.**

**a**, Representative cleavage assay by LbuC2c2 on pre-crRNAs containing structural mutations within the stem and loop regions of hairpin. Processed percentages listed below are quantified at 60 min (mean  $\pm$  s.d.,  $n = 3$ ). **b**, Bar graph showing the dependence of pre-crRNA processing on the CRISPR repeat sequence. The wild-type

repeat sequence is shown below with individual bars representing tandem nucleotide mutations as noted in red. The cleavage site is indicated by cartoon scissors. Percentage processed was measured after 60 min (mean  $\pm$  s.d.,  $n = 3$ ). **c**, Divalent metal ion dependence of the crRNA processing reaction was tested by addition of 10-50 mM EDTA and EGTA to standard reaction conditions.

Mechanistic studies of LbuC2c2 revealed that processing activity was unaffected by the presence of divalent metal ion chelators EDTA or EGTA (Fig. 3.7c), indicative of a metal ion-independent RNA hydrolytic mechanism. Metal ion-independent RNA hydrolysis is typified by the formation of a 2', 3'-cyclic phosphate and 5'-hydroxide on the 5' and 3' halves of the crRNA cleavage products, respectively (Yang, 2011). To determine the end-group chemical identity of C2c2-processed substrates, we further incubated the 5' flanking products with T4 polynucleotide kinase, which removes 2',3'-cyclic phosphates to leave a 3'-hydroxyl. We observed altered denaturing-gel migration of the 5' flanking product after kinase treatment, consistent with the removal of a 3' phosphate group (Fig. 3.6d). The divalent metal ion independence of C2c2's pre-crRNA processing activity is in stark contrast with the divalent metal ion dependency of Cpf1, the only other single-protein CRISPR effector shown to perform guide processing (Fonfara et al., 2016). Collectively, these data indicate that C2c2-catalyzed pre-crRNA cleavage is a divalent metal ion-independent process that likely uses a general acid-base catalysis mechanism (Yang, 2011).

### 3.3.3 C2c2:crRNA complexes degrade ssRNA upon target complementarity

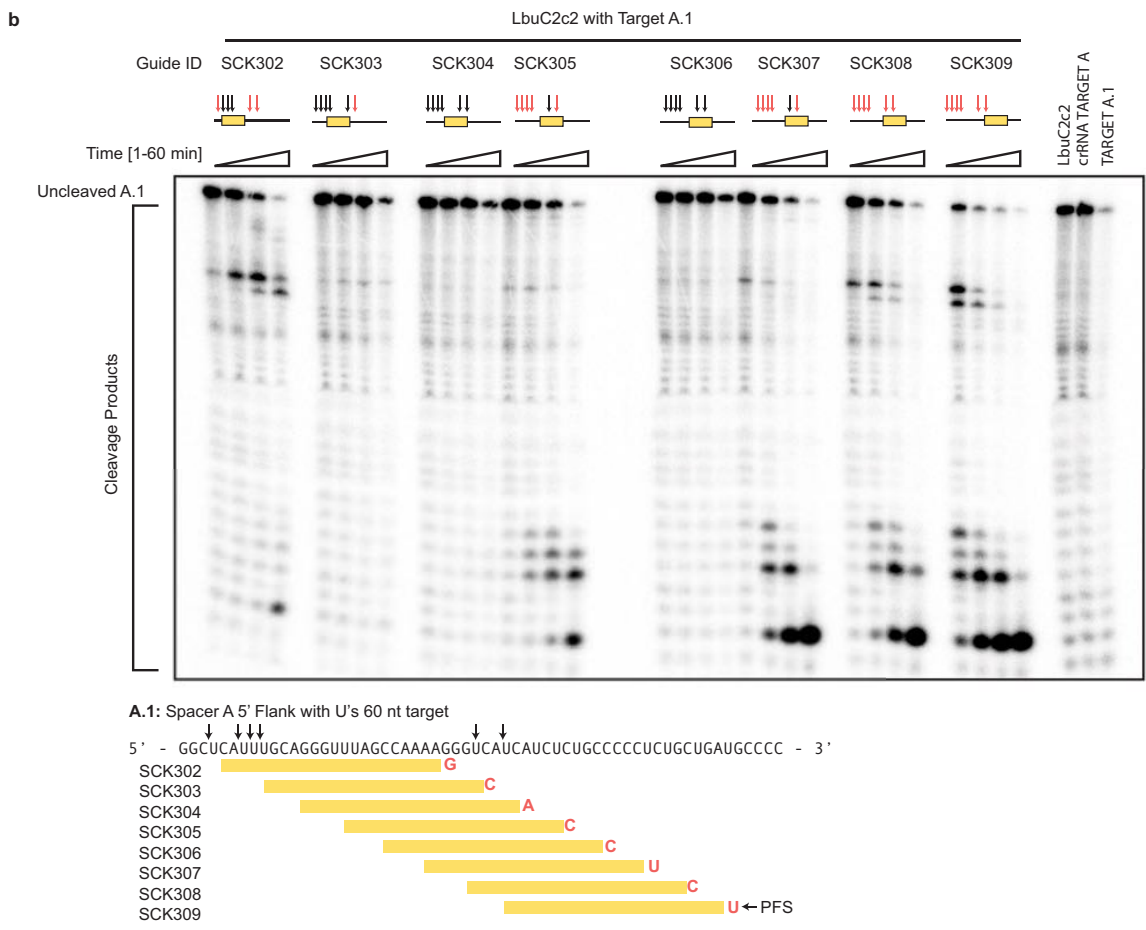
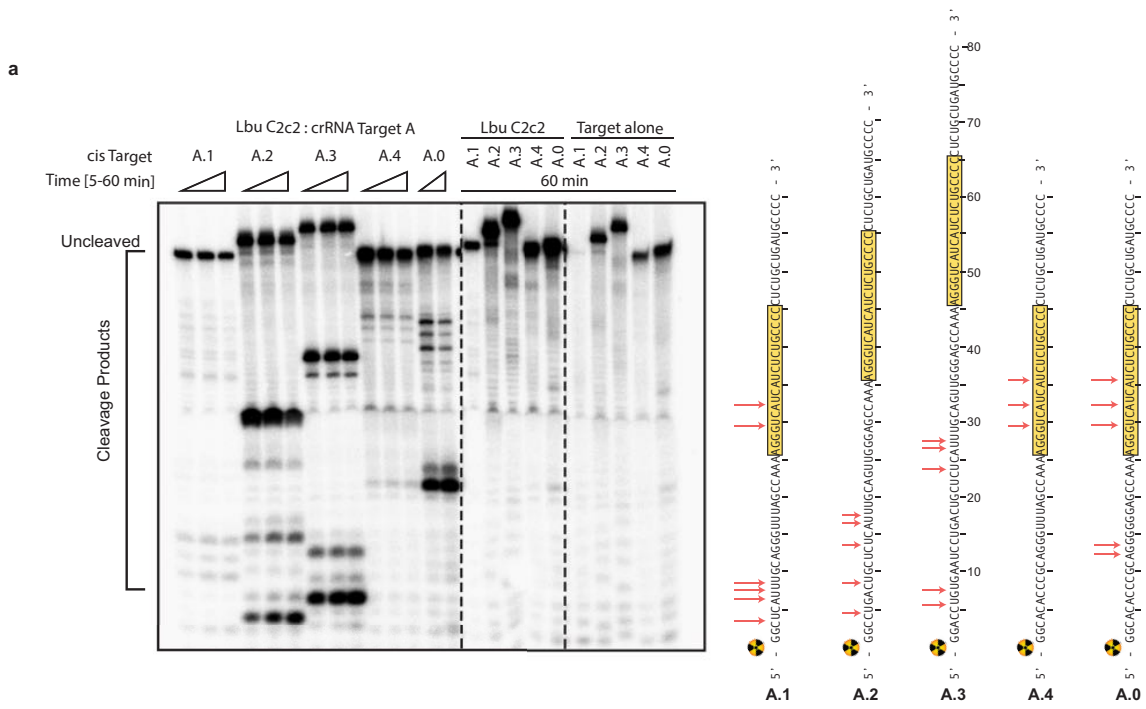


**Figure 3.8 LbuC2c2 catalyzes guide-dependent ssRNA degradation on *cis* and *trans* targets.**

**a**, Schematic of the two modes of C2c2, guide-dependent ssRNA degradation. **b**, Cleavage of two distinct radiolabeled ssRNA substrates, A and B, by LbuC2c2.

Complexes of 100 nM C2c2 and 50 nM crRNA were pre-formed at 37 °C, and reaction was initiated upon addition of <1 nM 5'-labeled target RNA at 25°C. *Trans* cleavage reactions contained equimolar (<1 nM) concentrations of radiolabeled non-guide-complementary substrate, and unlabeled on-target ssRNA. For multiple ssRNA substrates, we observed that LbuC2c2 catalyzed efficient cleavage only when bound to the complementary crRNA, indicating that LbuC2c2:crRNA cleaves ssRNA in an RNA-guided fashion. This activity is hereafter referred to as on-target or *cis*-target cleavage. LbuC2c2-mediated *cis* cleavage resulted in a laddering of multiple products, with cleavage preferentially occurring before uracil residues, analogous to LshC2c2 (Abudayyeh et al., 2016). We repeated non-target cleavage reactions in the presence of unlabeled, on-target (crRNA-complementary) ssRNA. In contrast to non-target cleavage experiments performed in *cis*, we observed rapid degradation of non-target RNA in *trans*. The similar RNA cleavage rates and near identical cleavage products observed for both *cis* on-target cleavage and *trans* non-target cleavage implicate the same nuclease center in both activities. **c**, LbuC2c2 loaded with crRNA targeting spacer A was tested for cleavage activity under both *cis* (target A labeled) and *trans* (target B labeled in the presence of unlabeled target A) cleavage conditions in the presence of 25 mM EDTA.

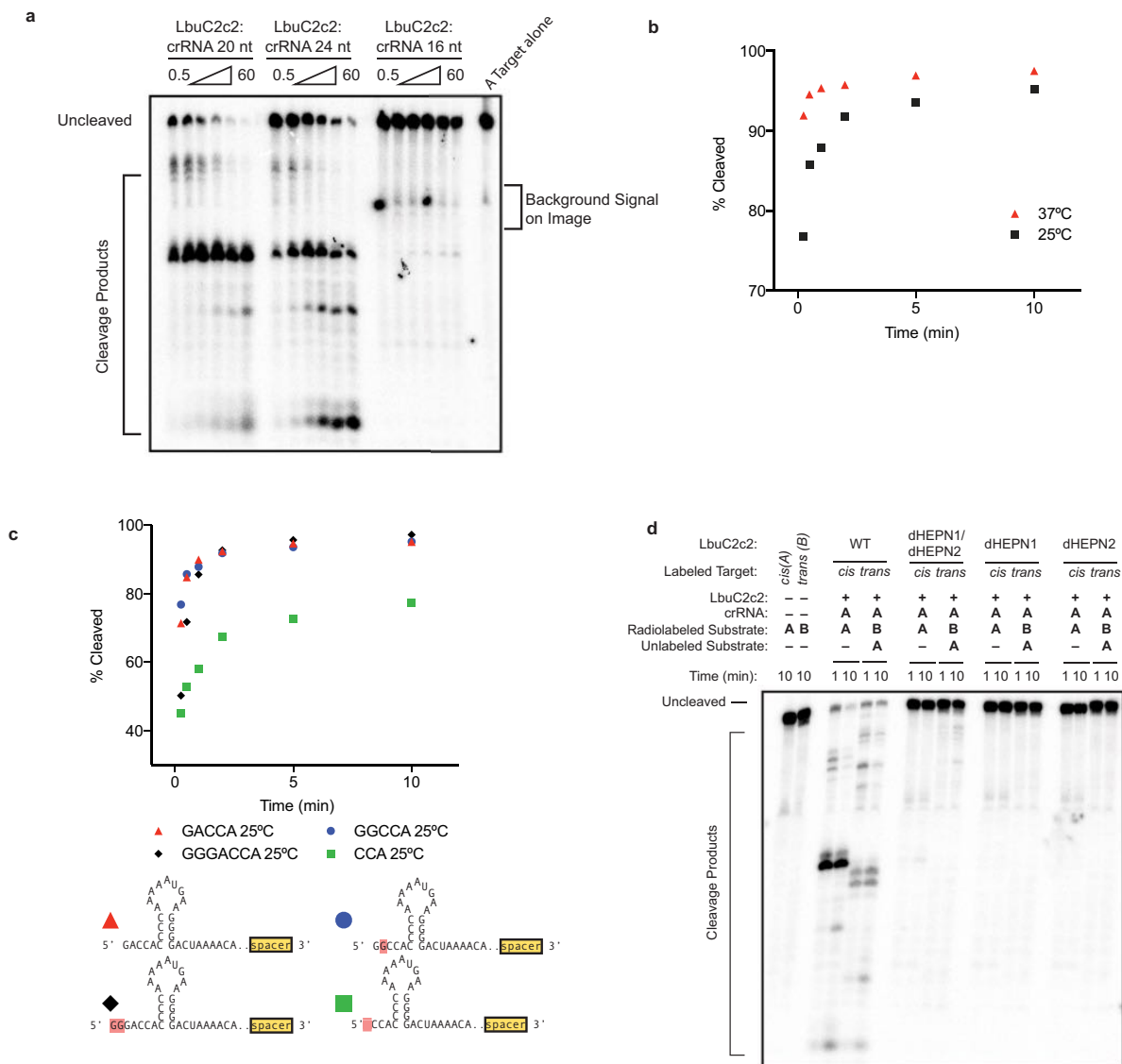
Following maturation, crRNAs typically bind with high affinity to Cas effector protein(s) to create RNA-guided surveillance complexes capable of sequence-specific nucleic acid recognition (Jinek et al., 2012; van der Oost et al., 2014; Wright et al., 2016). In agreement with previous work using LshC2c2 (Abudayyeh et al., 2016), LbuC2c2 catalyzed efficient target RNA cleavage only when such substrates could base pair with a complementary sequence in the crRNA (Figs. 3.8-3.10). Given the promiscuous pattern of cleavage observed for C2c2 (Fig. 3.9), we tested the ability of LbuC2c2 to act as a crRNA-activated non-specific RNA endonuclease in *trans* (Fig. 3.8b). In striking contrast to non-target cleavage experiments performed in *cis* and consistent with observations for LshC2c2 (Abudayyeh et al., 2016), we observed rapid degradation of non-target RNA in *trans* (Fig. 3.8b). This result shows that target recognition activates C2c2 for general non-specific degradation of RNA. Importantly, the similar RNA cleavage rates and near-identical cleavage products observed for both *cis* on-target cleavage and *trans* non-target cleavage of the same RNA substrate implicate the same nuclease center in both activities (Fig. 3.8b).



**Figure 3.9 LbuC2c2 ssRNA target cleavage site mapping**

**a**, ssRNA target cleavage assay conducted per Methods demonstrating LbuC2c2-mediated 'cis'-cleavage of several radiolabeled ssRNA substrates with identical spacer-complementary sequences but distinct 5' flanking sequences of variable length and nucleotide composition. Sequences of ssRNA substrates are shown to the right with spacer-complementary sequences for crRNA-A highlighted in yellow. Arrows indicate detected cleavage sites. Gel was cropped for clarity. It should be noted that the pattern of cleavage products produced on different substrates (e.g. A.1 vs. A.2 vs. A.3) indicates that the cleavage site choice is primarily driven by a uracil preference and exhibits an apparent lack of exclusive cleavage mechanism within the crRNA-complementary target sequence, which is in contrast to what is observed for other Class II CRISPR single effector complexes such as Cas9 and Cpf1 (Fonfara et al., 2016; Jinek et al., 2012). Interestingly, the cleavage pattern observed for substrate A.0 hints at a secondary preference for polyG sequences. **b**, LbuC2c2 ssRNA target cleavage assay as per Methods, using a range of crRNAs that tile the length of the ssRNA target. The sequence of the ssRNA substrates used in this experiment is shown below the gel with spacer-complementary sequences for each crRNA highlighted in yellow. Arrows indicate predicted cleavage sites. Above each set of lanes, a small diagram indicates the location of the spacer sequence along the target (yellow box) and the cleavage products observed (red arrows) or absent (black arrows). Likewise, it should be noted that for every crRNA the cleavage product length distribution is very similar, again indicating an apparent lack of exclusive cleavage within the crRNA-bound sequence. The absence of a several cleavage products in a subset of the reactions might be explained by the presence of bound C2c2:crRNA on the ssRNA target, which could sterically occlude access to uracils by any cis (intramolecular) or trans (intermolecular) LbuC2c2 active sites. While proper analysis for protospacer flanking site (PFS) preference for LbuC2c2 is beyond the scope of this study, minimal impact of the 3' flanking nucleotide was observed. Expected PFS base is noted in diagram next to each guide tested in red.



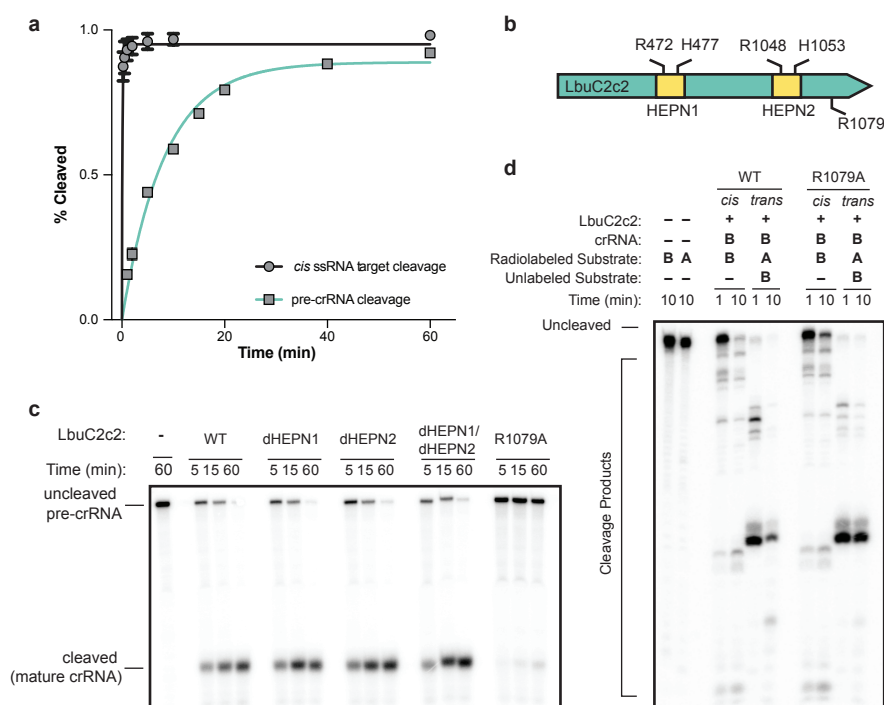


**Figure 3.10 Dependence of RNA targeting on crRNA variants, temperature and point mutations**

**a**, LbuC2c2 ssRNA target cleavage assay carried out, as per Methods with crRNAs possessing 16-nt, 20-nt or 24-nt spacers. **b**, LbuC2c2 ssRNA target cleavage time-course carried out at either 25°C and 37°C as per methods. **c**, LbuC2c2 ssRNA target cleavage timecourse carried out as per Methods with crRNAs possessing different 5'-flanking nucleotide mutations. Mutations are highlighted in red. 1-2 nucleotide 5' extensions negligibly impacted cleavage efficiencies. In contrast, shortening the flanking region to 3 nts slowed cleavage rates. **d** Impact of point mutations on ribonuclease activity of C2c2 in conserved residue mutants within HEPN motifs for ssRNA targeting.

### 3.3.4 pre-crRNA processing and ssRNA targeting activities occupy distinct active sites

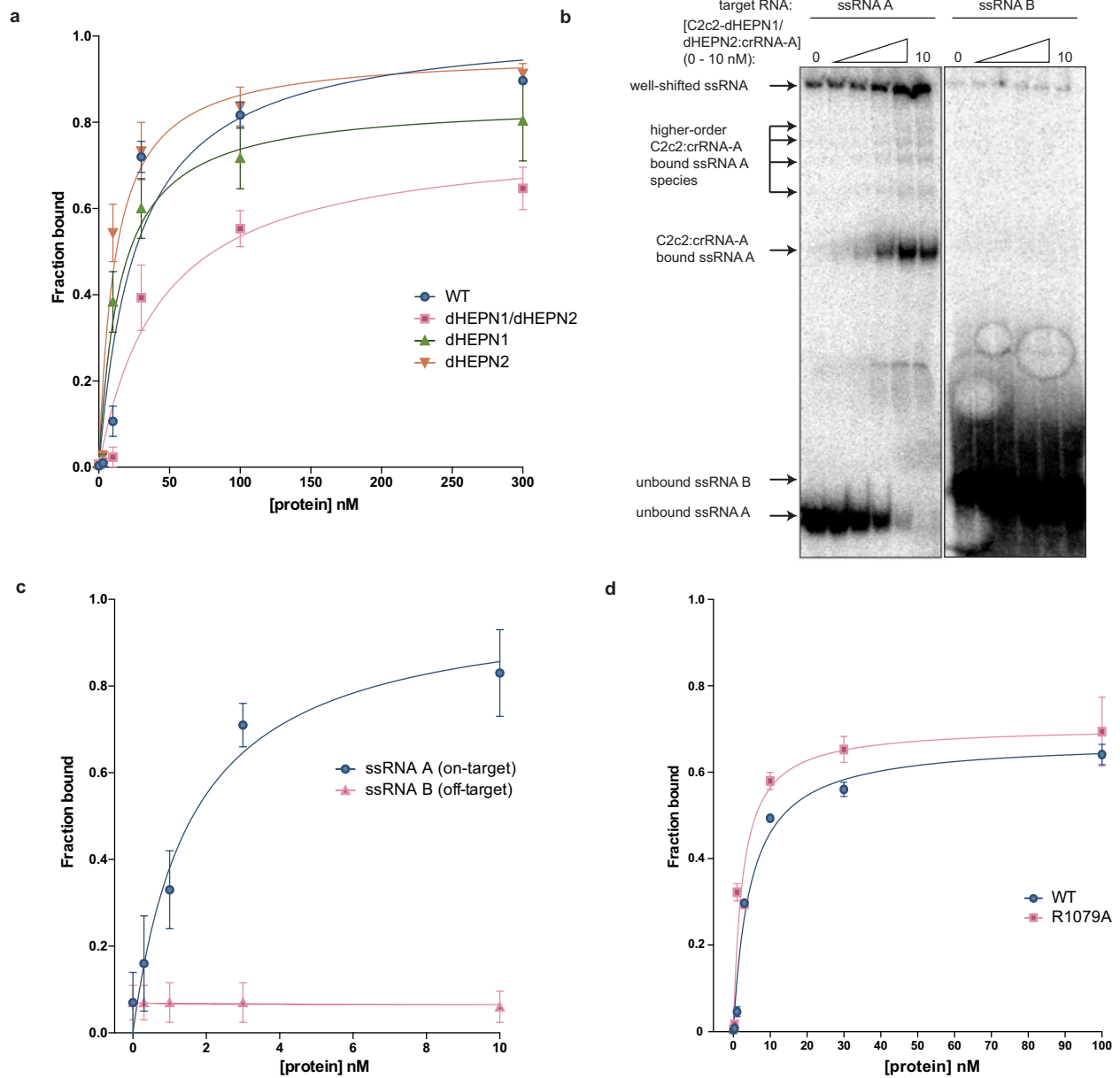
Notably, crRNA-mediated cleavage of target ssRNA occurs at an ~80-fold faster rate than pre-crRNA processing (Fig. 3.11a), and in contrast to pre-crRNA processing, RNA-guided target cleavage is abolished in the presence of EDTA, indicating that this activity is divalent metal ion-dependent (Fig. 3.8c, 3.10 and 3.11a). Given these clear differences, we reasoned that C2c2 might possess two orthogonal RNA cleavage activities: one for crRNA maturation, and the other for crRNA-directed, non-specific RNA degradation. To test this hypothesis, we systematically mutated several residues within the conserved HEPN motifs of LbuC2c2 (Abudayyeh et al., 2016; Anantharaman et al., 2013; Niewoehner and Jinek, 2016; Sheppard et al., 2016), and assessed pre-crRNA processing and RNA-guided RNase activity of the mutants (Fig. 3.10d and 3.11). Double and quadruple mutants of conserved HEPN residues (R472A, R477A, R1048A and R1053) retained robust pre-crRNA cleavage activity (Fig. 3.11c). By contrast, all HEPN mutations abolished RNA-guided cleavage activity while not affecting crRNA or ssRNA-binding ability (Fig. 3.10d and 3.12) (Abudayyeh et al., 2016).



**Figure 3.11 LbuC2c2 contains two distinct ribonuclease activities**

**a**, Quantified time-course data of *cis* ssRNA target (black) and pre-crRNA (teal) cleavage by LbuC2c2 performed at 37°C. Exponential fits are shown as solid lines (n=3), and the calculated pseudo-first-order rate constants ( $k_{obs}$ ) (mean  $\pm$  s.d.) are  $9.74 \pm 1.15 \text{ min}^{-1}$  and  $0.12 \pm 0.02 \text{ min}^{-1}$  for *cis* ssRNA target and pre-crRNA cleavage, respectively. **b**, LbuC2c2 architecture depicting the location of HEPN motifs and processing deficient point mutant **c,d** Ribonuclease activity of LbuC2c2 mutants for pre-crRNA processing in **c** and ssRNA targeting in **d** and Fig 3.10d.

Next we sought mutations that would abrogate pre-crRNA processing activity without disrupting target RNA cleavage. Given that we were unable to predict any other potential RNase motifs beyond the HEPN motifs, and that C2c2 proteins bear no homology to Cpf1, we opted to systematically mutate the charged residues throughout LbuC2c2. We identified an arginine residue (R1079A) that upon mutation resulted in severely attenuated pre-crRNA processing activity (Fig. 3.11c). This C2c2 mutant enzyme retained crRNA-binding ability as well as RNA target cleavage activity (Fig. 3.11d and 3.12d). Taken together, our results show that distinct active sites within the C2c2 protein catalyze pre-crRNA processing and RNA-directed RNA cleavage.



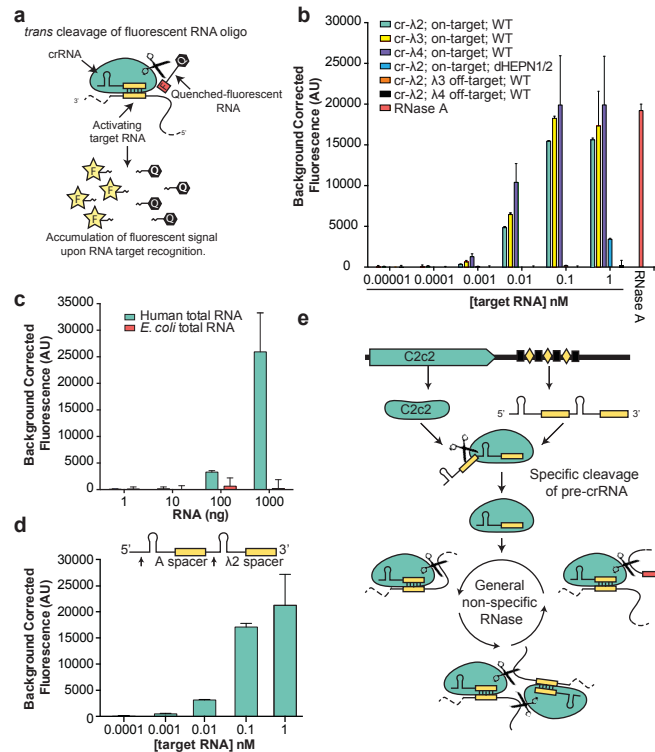
**Figure 3.12 Binding data for LbuC2c2 to mature crRNA and target ssRNA**

**a**, Filter binding assays were conducted as described in the Methods to determine the binding affinity of mature crRNA-A<sub>GG</sub> to LbuC2c2-WT, LbuC2c2-dHEPN1, LbuC2c2-dHEPN2, or LbuC2c2-dHEPN1/dHEPN2. The quantified data were fit to standard binding isotherms. Error bars represent the standard deviation from three independent experiments. Measured dissociation constants from three independent experiments (mean  $\pm$  sd) were  $27.1 \pm 7.5$  nM (LbuC2c2-WT),  $15.2 \pm 3.2$  nM (LbuC2c2-dHEPN1),  $11.5 \pm 2.5$  nM (LbuC2c2-dHEPN2), and  $43.3 \pm 11.5$  nM (LbuC2c2-dHEPN1/dHEPN2). **b**, Representative electrophoretic mobility shift assay for binding reactions between LbuC2c2-dHEPN1/dHEPN2: crRNA-A<sub>GG</sub> and either 'on-target' A ssRNA or 'off-target' B ssRNA, as indicated. Three independent experiments were conducted as described in the Methods. The gel was cropped for clarity. **c**, Quantified binding data from (b) were fitted to standard binding isoforms. Error bars represent the standard deviation from three independent experiments. Measured dissociation constants from three independent experiments (mean  $\pm$  sd) were  $1.62 \pm 0.43$  nM for ssRNA A and N.D ( $\gg 10$  nM) for ssRNA B. **d**, Filter binding assays were conducted as described in the Methods to determine the binding affinity of mature crRNA-A<sub>GA</sub> to LbuC2c2-WT and LbuC2c2-R1079A. The quantified data were fit to standard binding isotherms. Error bars represent the standard deviation from three independent experiments. Measured dissociation constants from three independent experiments (mean  $\pm$  sd) were  $4.65 \pm 0.6$  nM (LbuC2c2-WT) and  $2.52 \pm 0.5$  nM (LbuC2c2-R1079A). It is of note that these binding affinities differ from panel a. This difference is accounted for in a slight difference in the 5' sequence of the guide with panel a guides beginning with a 5'-GGCCA... and panel d 5'-GACCA. While the native sequence guide (5'-GACCA) binds tighter to LbuC2c2, no difference is seen in the RNA targeting efficiencies of these guide variants (Fig. 10c).

### 3.3.5 RNA detection by C2c2 harness for biotechnological applications

We recognized that C2c2's robust RNA-stimulated cleavage of *trans* substrates might be employed as a means of detecting specific RNAs within a pool of transcripts. While many polymerase-based methods have been developed for RNA amplification and subsequent detection, few approaches are able to directly detect the target RNA without significant engineering or stringent design constraints for each new RNA target (Cordray and Richards-Kortum, 2012; Rohrman et al., 2012; Yan et al., 2014; Yang, 2011). As a readily-programmable alternative, we tested whether C2c2's RNA-guided *trans* endonuclease activity could be harnessed to cleave a fluorophore-quencher-labeled reporter RNA substrate, thereby resulting in increased fluorescence upon target RNA-triggered RNase activation (Fig. 3.13a). LbuC2c2 was loaded with bacteriophage  $\lambda$ -targeting crRNAs and tested for its ability to detect the corresponding  $\lambda$  ssRNA targets spiked into HeLa cell total RNA. We found that upon addition of as little as 1-10 pM complementary  $\lambda$  target-RNA, a substantial crRNA-specific increase in fluorescence occurred within 30 min (Fig. 3.13b and 3.14a). Control experiments with either C2c2:crRNA complex alone or in the presence of crRNA and a non-complementary target RNA resulted in negligible increases in fluorescence relative to an RNase A positive control (Fig. 3.13b and 3.14a). We note that at the 10 pM concentration of a  $\lambda$  target RNA, only  $\sim 0.02\%$  of the C2c2:crRNA complex is predicted to be in the active state, yet the observed fluorescent signal reflected  $\sim 25-50\%$  cleavage

of the reporter RNA substrate, depending on the RNA target. Fragment size resolution of the background RNA in these reactions revealed significant degradation, even on highly structured tRNAs (Fig. 3.14b). Since reporter RNA cleavage occurs in the presence of a vast excess of unlabeled RNA, we conclude that LbuC2c2 is a robust multiple-turnover enzyme capable of at least  $10^4$  turnovers per target RNA recognized. Thus, in contrast to previous observations (Abudayyeh et al., 2016), crRNA-directed *trans* cleavage is potent and detectable even at extremely low levels of activated protein.

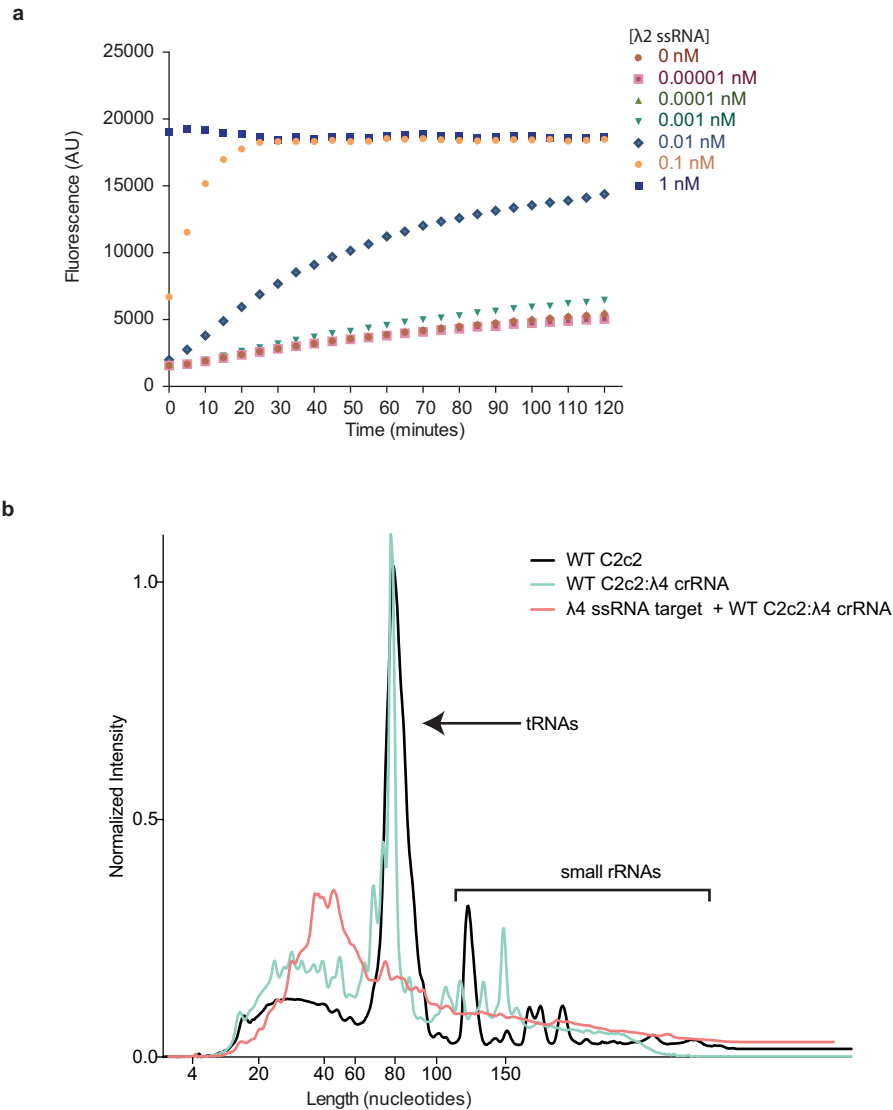


**Figure 3.13 C2c2 provides sensitive detection of transcripts in complex mixtures**

a, Illustration of LbuC2c2 RNA detection approach using a quenched fluorescent RNA reporter. b, Quantification of fluorescence signal generated by LbuC2c2 after 30 min for varying concentrations of target RNA in the presence of human total RNA. RNase A shown as positive RNA degradation control. (mean  $\pm$  s.d.,  $n = 3$ ) c, Quantification of fluorescence signal generated by LbuC2c2 loaded with a  $\beta$ -actin targeting crRNA after 3h for varying amounts of human total RNA or bacterial total RNA (as a  $\beta$ -actin null negative control). (mean  $\pm$  s.d.,  $n = 3$ ) d, Tandem pre-crRNA processing also enables RNA detection. (mean  $\pm$  s.d.,  $n = 3$ ) e, Model of the Type VI CRISPR pathway highlighting both of C2c2's ribonuclease activities.

To extend this LbuC2c2 RNA detection system, we designed a crRNA to target endogenous beta-actin mRNA. We observed a measurable increase in fluorescence in the presence of human total RNA relative to *E. coli* total RNA, demonstrating the specificity of this method (Fig. 3.13c). Furthermore, given that C2c2 processes its own guide, we combined pre-crRNA processing and RNA detection in a single reaction by

designing tandem crRNA-repeat containing spacers complementary to target RNAs A and  $\lambda 2$ . LbuC2c2 incubated with this unprocessed tandem guide RNA in the detection assay generated a significant increase in fluorescence similar in magnitude and sensitivity to experiments using mature crRNAs (Fig. 3.13b, d). Taken together, these data highlight the exciting opportunity to take advantage of C2c2's two distinct RNase activities for a range of biotechnological applications (Fig. 3.13e).



**Figure 3.14 RNase detection assay  $\lambda 2$ -ssRNA time-course and background RNA cleavage**

**a**, LbuC2c2:crRNA- $\lambda 2$  was incubated with RNAase-Alert substrate (Thermo-Fisher) and 100 ng HeLa total RNA in the presence of increasing amounts of  $\lambda 2$  ssRNA (0-1 nM) for 120 min at 37°C. Fluorescence measurements were taken every 5 min. The 1 nM  $\lambda 2$  ssRNA reaction reached saturation before the first time point could be measured. Error bars represent the standard deviation from three independent experiments. **b**, LbuC2c2:crRNA- $\lambda 4$  or apo LbuC2c2 was incubated in HeLa total RNA for 2 hours in the

presence or absence of on-target activating  $\lambda$ 4 ssRNA. Degradation of background small RNA was resolved on a small RNA chip in a Bioanalyzer 2100 as per Methods. Small differences are seen in the fragment profile of between apo LbuC2c2 and LbuC2c2:crRNA- $\lambda$ 4. In contrast, upon addition of the on-target ssRNA to the reaction, a drastic broadening and shifting of the tRNA peak reveals extensive degradation of other structured and nonstructured RNA's present in the reaction upon activation of LbuC2c2 *trans* activity.

### 3.4 Discussion

In bacteria, C2c2 likely operates as a sentinel for viral RNAs (Abudayyeh et al., 2016). We propose that when invasive transcripts are detected within the host cell via base pairing with crRNAs, C2c2 is activated for promiscuous cleavage of RNA in *trans* (Fig. 3.13e). As a defense mechanism, this bears striking similarity to RNase L and caspase systems in eukaryotes, whereby a cellular signal triggers promiscuous ribonucleolytic or proteolytic degradation within the host cell, respectively, leading to apoptosis (Choi et al., 2015; McIlwain et al., 2013). While the RNA targeting mechanisms of Type III CRISPR systems generally result in RNA cleavage within the protospacer-guide duplex (Samai et al., 2015; Zhang et al., 2016), recent examples of associated nucleases Csx1 (Sheppard et al., 2016) and Csm6 (Niewoehner and Jinek, 2016) provide compelling parallels between the Type VI systems and the multi-component Type III inference complexes.

Our data show that CRISPR-C2c2 proteins represent a new class of enzyme capable of two separate RNA recognition and cleavage activities. Efficient pre-crRNA processing requires sequence and structural motifs within the CRISPR repeat which prevent non-endogenous crRNA loading and helps to reduce the potential toxicity of this potent RNase. The entirely different pre-crRNA processing mechanisms of C2c2 and the Type V CRISPR effector protein Cpf1 indicate that each protein family has converged upon independent activities encompassing both the processing and interference functions of their respective CRISPR pathways. Furthermore, the two distinct catalytic capabilities of C2c2 can be harnessed in concert for RNA detection, as the activation of C2c2 to cleave thousands of *trans*-RNAs for every target RNA detected enables potent signal amplification. The capacity of C2c2 to process its own guide RNAs from arrays could also allow the use of tissue-specific Pol II promoters for guide expression, in addition to target multiplexing for a wide range of applications. The C2c2 enzyme is unique within bacterial adaptive immunity for its dual RNase activities, and highlights the utility of harnessing CRISPR proteins for precise nucleic acid manipulation in cells and cell-free systems.

# Chapter 4

## Delving into the Cas13a Family Tree

A portion of the content presented in this chapter has been previously published as part of the following research article: Alexandra East-Seletsky, Mitchell R. O'Connell, David Burstein, Gavin J. Knott and Jennifer A. Doudna, (2017). RNA targeting by functionally orthogonal Type VI-A CRISPR-Cas enzymes. *Mol Cell (in press)*

Alexandra Seletsky and Mitchell O'Connell conceived the study and designed experiments with input from Jennifer Doudna. Alexandra Seletsky executed all experimental work with assistance from Mitchell O'Connell. David Burstein advised on bioinformatic analyses and Gavin Knott assisted with data analysis of homolog activities. All authors discussed the data and wrote the manuscript.



## 4.1 Summary

CRISPR adaptive immunity pathways protect prokaryotic cells against foreign nucleic acids using CRISPR RNA (crRNA)-guided nucleases. In Type VI-A CRISPR-Cas systems, the signature protein Cas13a (formerly C2c2) contains two separate ribonuclease activities that catalyze crRNA maturation and ssRNA degradation, respectively. The Cas13a protein family occurs across different bacterial phyla and varies widely in both protein sequence and corresponding crRNA sequence conservation. Although defined phylogenetically as a single enzyme family, we show that Cas13a enzymes comprise two distinct functional groups that recognize orthogonal sets of crRNAs and possess different ssRNA cleavage specificities. These functional distinctions could not be bioinformatically predicted, suggesting more subtle co-evolution of Cas13a enzymes. Additionally, we find that Cas13a pre-crRNA processing is not essential for ssRNA cleavage, although it enhances ssRNA targeting for crRNAs encoded internally within the CRISPR array. We define two Cas13a protein subfamilies that can operate in parallel for RNA detection both in bacteria and for diagnostic applications.

## 4.2 Introduction

Competition for survival within microbial communities drives the evolution of diverse pathways for anti-viral defense (Dy et al., 2014). Unique among such defense mechanisms, CRISPR-Cas (clustered regularly interspaced short palindromic repeats - CRISPR associated) systems provide adaptive immunity by means of CRISPR genomic loci that contain integrated viral DNA sequences (spacers) flanked by conserved palindromic sequences (repeats) (Charpentier et al., 2015; Mohanraju et al., 2016; Wright et al., 2016). CRISPR RNAs (crRNAs) expressed from these loci assemble with Cas proteins to form surveillance complexes that recognize and cleave nucleic acids matching the virally-derived segment of the crRNA. Three steps ensure ongoing adaptive immunity: 1) integration of new virus spacer sequences into the CRISPR locus; 2) Cas protein expression and crRNA biogenesis; and 3) surveillance complex assembly and target interference (Barrangou et al., 2007; Brouns et al., 2008; Garneau et al., 2010).

Although double-stranded DNA (dsDNA) is the typical CRISPR-Cas target, Type III and VI CRISPR-Cas systems instead recognize single-stranded RNA (ssRNA). Unlike the DNA-targeting CRISPR effector complexes or the type III CRISPR-Cas effector complexes, which catalyze site-specific cleavage of DNA or RNA sequences specifically matching the crRNA, respectively (reviewed in (Charpentier et al., 2015; Mohanraju et al., 2016; Wright et al., 2016)), the Type VI systems catalyze complete degradation of any ssRNA present upon activation (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Smargon et al., 2017). The signature protein of Type VI-A CRISPR-Cas systems, Cas13a (formerly C2c2), is a dual nuclease responsible for both crRNA maturation and RNA-activated ssRNA cleavage (East-Seletsky et al., 2016). Cas13a binds to precursor crRNA (pre-crRNA) transcripts and cleaves them within the repeat region to produce mature crRNAs. This cleavage reaction is presumed to be

single-turnover, resulting in a functional Cas13a:crRNA surveillance complex. Binding to a ssRNA with sequence complementarity to the crRNA activates Cas13a for *trans*-ssRNA cleavage, potentially triggering cell death or dormancy of the host organism (Abudayyeh et al., 2016; East-Seletsky et al., 2016; Liu et al., 2017b). While previous work demonstrated that these two enzymatic activities of Cas13a are chemically distinct and occur within separate active sites (East-Seletsky et al., 2016; Liu et al., 2017b), the interdependence, if any, between active site substrate specificities is unknown.

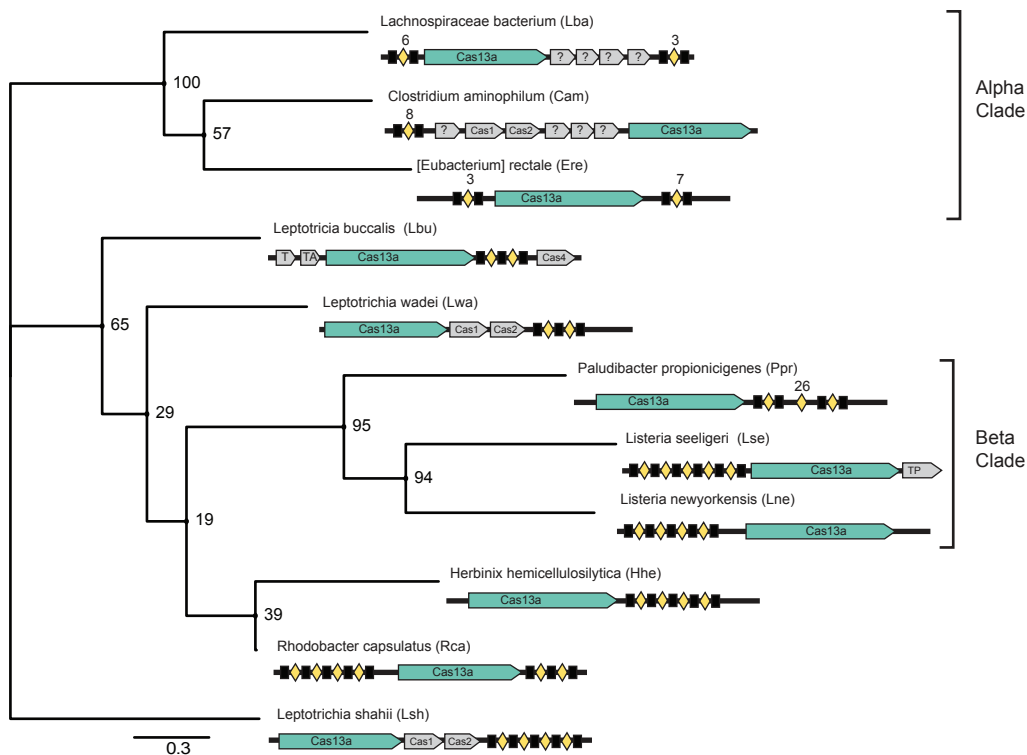
Multiple CRISPR-Cas loci, whether from redundant or distinct subtypes, commonly co-exist within host genomes, but the relationships between these systems are unclear. Maturation of crRNAs from different CRISPR loci is sometimes catalyzed by the same processing enzyme (Nickel et al., 2013; Niewoehner et al., 2014), while other organisms require distinct enzymes for each crRNA repeat species (Scholz et al., 2013; Sokolowski et al., 2014). Once generated, crRNAs can sometimes assemble into targeting complexes that include Cas proteins from other loci, but the functional consequences of such CRISPR-loci crosstalk are not known (Staals et al., 2014; Zhang et al., 2012). Cas13a provides a unique opportunity to study pathway interdependence within a CRISPR subtype for both pre-crRNA processing and target cleavage since the same protein catalyzes both activities. Furthermore, understanding the extent of orthogonality between distinct type VI-A CRISPR-Cas systems (i.e. are there subsets of type VI-A systems that recognize orthogonal sets of crRNAs and/or possess different ssRNA cleavage specificities), will advance the development of these proteins as biotechnological tools.

CRISPR proteins are classified according to sequence and phylogenetic conservation in attempts to simplify the broad diversity found within these microbial defense systems (Makarova et al., 2011; 2015; Shmakov et al., 2017). In turn, representative operons are investigated with the subsequent conclusions applied across the entire subtype protein family. But it is unclear how appropriate these generalizations are given the extremely fast evolution of CRISPR enzymes driven by bacteria-phage interactions. Within the Type VI-A protein family, homologs only share 15-20% residue identity and exist within diverse host genomes. This suggested that we could investigate divergent Cas13a homologs, which presumably each contain two well-defined nuclease activities, to test the functional diversity within a phylogenetically determined CRISPR-Cas subtype. Working with ten Cas13a homologs that span the major branches of the Cas13a family, we find that both bioinformatic analyses and biochemical results define two functional enzyme subfamilies within the Type VI-A subtype. Orthogonal pre-crRNA and ssRNA-target cleaving activities distinguish these groups, such that each enzyme subfamily recognizes distinct crRNAs yielding targeting complexes with subfamily-distinct ssRNA-degradation patterns. Furthermore, these results revealed that pre-crRNA processing is not essential for ssRNA targeting in Type VI systems, despite strong conservation of this function within the Cas13a family. Instead, pre-crRNA processing enhances *trans*-ssRNA targeting by liberating crRNAs from the long pre-crRNA transcript, leading to a revised model for Type VI-A CRISPR-Cas systems. The two orthogonal subfamilies of Cas13a enable potential applications in RNA detection, or imaging for the parallel detection of distinct RNA species.

## 4.3 Methods

### 4.3.1 Phylogenic and repeat conservation analysis

Cas13a maximum-likelihood phylogenies were computed as previously described (East-Seletsky et al., 2016) using RAXML (Stamatakis, 2014a) with the PROTGAMMALG evolutionary model and 100 bootstraps. Protein clades alpha and beta were defined as branch points with bootstrap values greater than 90, suggesting high confidence in having a common ancestor (Fig. 4.1). The remaining proteins were labeled as ambiguous ancestry, as the phylogenetic relationships between them were low confidence, reflected in bootstrap values less than 90. Sequences were aligned by MAFFT with the 'einsi' method (Kato and Standley, 2013a). Candidates were selected to represent each of the major branches of the Cas13a protein tree (Table 4.1). Alignments were performed for all non-redundant homologs and candidate proteins. Comparison of the CRISPR-RNA (crRNA) repeats was carried out by calculating pairwise similarity scores using the Needleman-Wunsch algorithm through the Needle tool on EMBL-EBI (McWilliam et al., 2013). Hierarchical clustering of CRISPR crRNA was performed in R using the similarity score matrix.



**Figure 4.1 CRISPR loci architecture for Cas13a homologs used in this study**

(A) Maximum-likelihood phylogenetic tree of Cas13a proteins with diagrams of Type VI-A loci adapted from (Shmakov et al., 2015). Cas13a ORFs shown in teal. CRISPR arrays depicted as black boxes (repeats), yellow diamonds (spacers), and spacer array

Prefix	Species	Strain	Cas13a Accession	Genome Accession	CRISPR array start	CRISPR array stop	CRISPR Repeat Consensus	Mature crRNA handle	Use in Study
Lwa	<i>Leptotrichia wadei</i>	F0279	ERK53440.1	AWW/M010000026.1	24602	24774	CATATAGACCACCCCAATAATCGAAGGGGACTTAAACTTT	GACCACCCAAUUCGAAAGGGGACUAAAACUU	Yes
Rca	<i>Rhodobacter capsulatus</i>	R121	ETD76934.1	AYQC01000019.1	321133	321440	TCACATCACCGGCAAGACGCGGGGACTGAACC	CAUACCCGCAAGACGACGGCGGACUAGAACC	Yes
Hhe	<i>Herbinix hemiellulosilytica</i>	T3/65T	ORZ35554.1	CVTD020000026.1	325840	326009	<b>CGAC</b> CATCACCCTCCCAAATCGACGGGACTGAACCT	not tested	Yes
Lbu	<i>Leptotrichia buccalis</i>	DSM 1135	WP_015770004.1	NC_013192.1	59561	59859	GTAAACATCCCGTAGACAGGGGAACTGCAAC	not cleaved	No
Lse	<i>Listeria seeligeri</i>	SLCC3954	WP_012985477.1	NC_013891.1	1174057	1174422	GATTTAGACCCCCCAAAAATGAAGGGGACTTAAACA	GACCACCCAAAAUUAAGAAAGGGGACUAAAACA	Yes
Ppr	<i>Paludibacter propionigenes</i>	WB4	WP_013443710.1	NC_014734.1	223912	225653	CTTGTGCAATTTCCCAATTTGAAGGAACTACAAC	not tested	Yes
Lba	<i>Lachnospiraceae bacterium</i>	NK4A179	WP_022765443.1	NZ_ATWC01000054.1	10469	10613	CTTGTGCAATTTCCCAATTTGAAGGAACTACAAC	not tested	No
Ere	[ <i>Eubacterium</i> ] <i>rectale</i>	T1-815	WP_055061018.1	NZ_CVRO010000008.1	49	290	CCTGGAGAATAGCCCAAGAAAGAGGGGCAATAAC	AGAUAGCCCAAGAAAGAGGGGCAUAUAC	Yes
Lny	<i>Listeria newyorkensis</i>	FSL M6-0635	WP_036091002.1	NZ_JNFB01000012.1	63967	64487	GTGAAATGACCCCGATATAGAGGGCAATAAC	GAAGAGCCCAAGAUAGAGGGGCAUAUAC	No
Cam	<i>Clostridium aminophilum</i>	DSM 10710	WP_031473346.1	NZ_JONJ01000012.1	58872	59187	GTGAAATGACCCCGATATAGAGGGGCAATAAC	AUACAGCUUAUUAAGAAAGGGGCAUAUAG	Yes
Lsh	<i>Leptotrichia shahii</i>	DSM 19757	WP_018451595.1	NZ_KB890278.1	169592	169890	GATTTAGAGTACTCTCAAAACAAAAGAGGACTTAAAC	GAGUACCUCAAAAAGAAAGGGGCAUAUAC	Yes
					1521	2181	GTTTGGAGAAGCCCGATATAGAGGGGCAATAAC	GAACAGCCGAUUAAGAGGGGCAUAUAG	Yes
					34127	34365	CATATAGACCACCCCAATAATCGAAGGGGACTTAAACA	CACCCCAAUUCGAAAGGGGCAUAUAC	Yes

size for larger arrays noted above. ORFs of interest surrounding the loci are noted with the following abbreviations: T-Toxin, AT- antitoxin and TP-transposase.

### Table 4.1 List of Cas13a homologs used in this study

**Prefix:** Cas13a abbreviation used in this study.

**CRISPR repeat consensus sequences:** For homologs with multiple CRISPR loci within 10 kb of the Cas13a containing operon (RcaCas13a, LbaCas13a and EreCas13a), or long arrays with repeat variations (PprCas13a), multiple crRNA repeat sequences are listed with mutations highlighted in red text. For PprCas13a, the first crRNA repeat at the leader side of the array was chosen for this study. For RcaCas13a, LbaCas13a and EreCas13a, the crRNA repeat sequences analyzed for two factors to choose a representative crRNA for this study (1): the length of the array, and (2) capacity to direct trans-ssRNA cleavage by the cognate Cas13a protein (Data not shown). Sequences used in the main text are noted in the last column.

### 4.3.2 Cas13a protein expression and purification

Expression vectors for protein purification were assembled using synthetic gBlocks ordered from Integrated DNA Technologies (IDT). The codon-optimized Cas13a genomic sequences were *N*-terminally tagged with a His<sub>6</sub>-MBP-TEV cleavage site sequence, with expression driven by a T7 promoter. Mutant proteins were cloned via site-directed mutagenesis of wild-type Cas13a constructs. Purification of all homologs was based off of a previously published protocol (East-Seletsky et al., 2016). Briefly, expression vectors were transformed into Rosetta2 DE3 or BL21 *E. coli* cells grown in 2xYT broth at 37 °C, induced at mid-log phase with 0.5 mM IPTG, and then transferred to 16 °C for overnight expression. Cell pellets were resuspended in lysis buffer (50 mM Tris-Cl pH 7.0, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 0.5 mM PMSF, and EDTA-free protease inhibitor (Roche)), lysed by sonication, and clarified by centrifugation at 15,000g. Soluble His<sub>6</sub>-MBP-TEV-Cas13a was isolated over metal ion affinity chromatography, and in order to cleave off the His<sub>6</sub>-MBP tag, the protein-containing eluate was incubated with TEV protease at 4 °C overnight while dialyzing into ion exchange buffer (50 mM Tris-Cl pH 7.0, 250 mM KCl, 5% glycerol, 1 mM TCEP). Cleaved protein was loaded onto a HiTrap SP column (GE Healthcare) and eluted over a linear KCl (0.25-1.5M) gradient. Cas13a containing fractions were pooled, concentrated, and further purified via size-exclusion chromatography on a S200 column (GE Healthcare) in gel filtration buffer (20 mM Tris-Cl pH 7.0, 200 mM KCl, 5% glycerol, 1 mM TCEP) and were subsequently stored at -80°C. All homologs were purified using this protocol except LwaCas13a which was bound to a HiTrap Heparin column instead of a SP column, and the size-exclusion chromatography step was omitted due to sufficient purity of the sample post ion-exchange. All expression plasmids are deposited with Addgene.

### 4.3.3 *In-vitro* RNA transcription

All pre-crRNAs, mature crRNAs, and targets were transcribed *in vitro* using previously described methods (East-Seletsky et al., 2016; Sternberg et al., 2012). Briefly, all substrates were transcribed off a single-stranded DNA oligonucleotide template (IDT), except for mature crRNAs requiring a non-GR 5' terminus. For these mature crRNAs, T7 polymerase templates containing a Hammerhead Ribozyme sequence immediately upstream of the mature crRNA sequence were generated using overlap PCR, and then purified for use as the template for T7 transcription (see Table 4.2 for sequences). Lbu six-mer CRISPR array *in vitro* transcription template was synthesized by GeneArt (ThermoFisher) as a plasmid. The T7 promoter-CRISPR array region was PCR amplified and purified prior to use as the template for T7 transcription. All transcribed RNAs were purified using 15% Urea-PAGE, except for the array which was purified using 6% Urea-PAGE. All RNAs were subsequently treated with calf alkaline phosphatase to remove 5' phosphates. Radiolabeling was performed as previously described (East-Seletsky et al., 2016; Sternberg et al., 2012). A, C, G and U homopolymers, and fluorescently-labeled RNA reporters for *trans* -ssRNA cleavage were synthesized by IDT. Homopolymers were purified using 25% Urea-PAGE after radiolabeling to reduce substrate heterogeneity.



#### 4.3.4 Radiolabeled ssRNA nuclease assays

pre-crRNA processing assays were performed at 37 °C in RNA processing buffer (20 mM HEPES-Na pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10 µg/mL BSA, 100 µg/mL tRNA, 0.01% Igepal CA-630 and 5% glycerol) with a 100-fold molar excess of Cas13a relative to 5'-labeled pre-crRNA (final concentrations of 100 nM and <1 nM, respectively). Unless otherwise indicated, reactions were quenched after 60 min with 1.5X RNA loading dye (100% formamide, 0.025% bromophenol blue, and 200 µg mL<sup>-1</sup> heparin). After quenching, reactions were denatured at 95 °C for 5 min prior to resolving by 15% denaturing PAGE (0.5x TBE buffer). Target cleavages assays were performed at 37 °C in cleavage buffer (20 mM HEPES-Na pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, and 5% glycerol). Generally, Cas13a:crRNA complex formation was performed in cleavage buffer, at a molar ratio of 2:1 protein to crRNA at 37 °C for 60 min, prior to adding 5'-labeled target and/or other non-radiolabeled RNA target substrates. Unless otherwise indicated, final concentrations were 100 nM Cas13a, 50 nM crRNA or pre-crRNA, 50 nM crRNA-complementary target ssRNA (henceforth referred to as 'activator') and <1 nM *trans*-ssRNA target.

All bands were visualized by phosphorimaging (Typhoon, GE Healthcare) and quantified with ImageQuant (GE Healthcare). For pre-crRNA processing, the percent cleavage was determined as the ratio of the product band intensity to the total intensity of both the product and uncleaved pre-crRNA bands and normalized for background within each measured substrate. For *trans*-ssRNA cleavage reactions, the percentage cleavage was determined as the ratio of all fragments smaller than the target to the total intensity within the lane and normalized for background within each substrate. These data were subsequently fit to a single-exponential decay using Prism7 (GraphPad) and cleavage rates are reported in figure legends.

#### 4.3.5 Fluorescent ssRNA nuclease assays.

Cas13a:crRNA complexes were assembled in cleavage buffer, as described above. 150 nM of RNase Alert reporter (IDT) and various final concentrations (0-1 µM) of ssRNA-activator were added to initiate the reaction. Notably these reactions are in the absence of competitor tRNA or total RNA compared to our previous work (East-Seletsky et al., 2016) to more accurately measure *trans*-cleavage activity. These reactions were incubated in a fluorescence plate reader (Tecan Infinite Pro F2000) for up to 120 min at 37°C with fluorescence measurements taken every 5 min ( $\lambda_{\text{ex}}$ : 485 nm;  $\lambda_{\text{em}}$ : 535 nm). For fluorescent homopolymer ssRNA reporter studies, Cas13a:crRNA complexes were pre-incubated at 37°C for 60 mins using standard conditions. Activator ssRNA and 200 nM fluorescent ssRNA reporter were added to initiate the reaction immediately before placing reaction in plate reader. For Lbu- and Lba- Cas13a containing samples with fluorescent homopolymer ssRNA reporters, 10 pM and 1 nM activator was used, respectively. For pre-crRNA array experiments, 300 nM Cas13a was first incubated with 50 nM pre-crRNA array for 1 hr in cleavage buffer to enable binding and processing of the array. 100 pM of each ssRNA activator was added along with 150 nM of RNase Alert reporter (IDT) to initiate the reaction, in biological triplicate for each spacer sequence. Apparent rates were calculated using one single-exponential decay using Prism7 (GraphPad) and calculated rates are plotted with their associated standard deviations.

Background-corrected fluorescence values were obtained by subtracting fluorescence values obtained from reactions carried out in the absence of target ssRNA activator. For determining homolog sensitivities and array processing effects, background corrected values were fit to a single-exponential decay using Prism7 (GraphPad) and the calculated rates were plotted with their associated standard deviations from n=3. For comparing non-cognate crRNA directed *trans*-ssRNA cleavage, initial reaction rates were instead calculated due to discrepancies in fluorescence plateau values across the dataset. Rates were then scaled relative to the cognate crRNA to normalize rates across the homologs. See Tables S4-6 for normalized values.

#### **4.3.6 crRNA filter-binding assays**

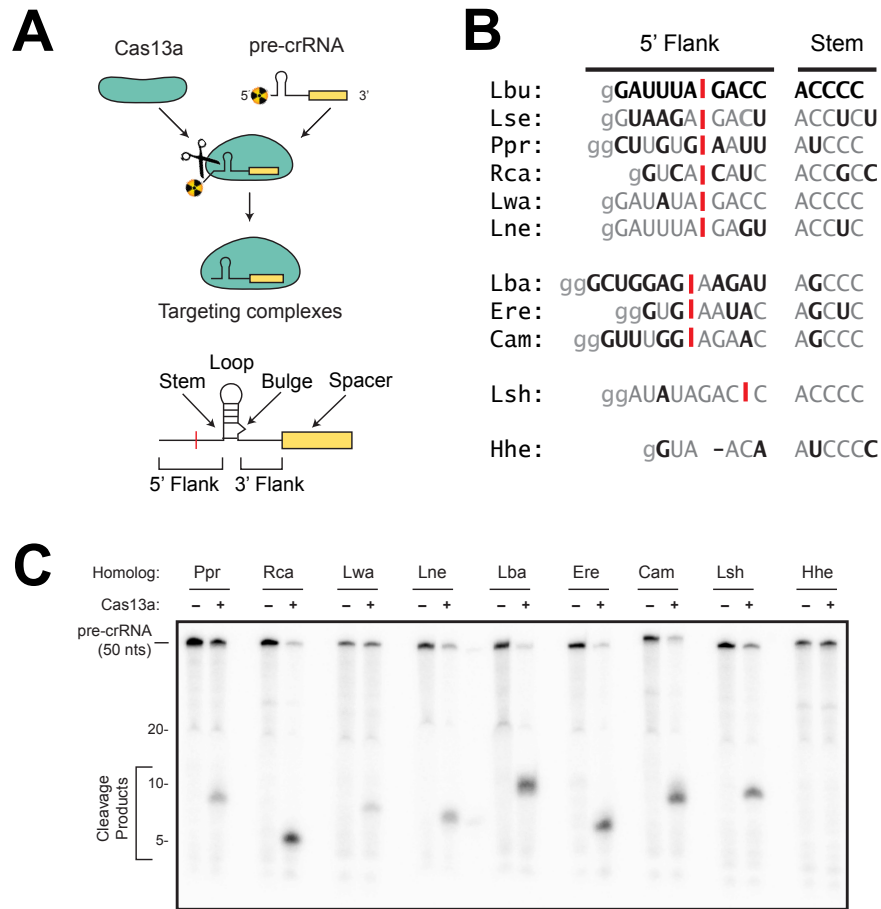
Filter binding assays was carried out as previously described (East-Seletsky et al., 2016). Briefly Cas13a and radiolabeled crRNA were incubated for 1 hr at 37°C in RNA processing buffer (20 mM HEPES-Na pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub>, 10 µg/mL BSA, 100 µg/mL yeast tRNA, 0.01% Igepal CA-630 and 5% glycerol). Tufryn, Protran and Hybond-N+ were assembled onto a dot-blot apparatus in the order listed above. The membranes were washed twice with 50µL Equilibration Buffer (20 mM HEPES-Na pH 6.8, 50 mM KCl, 5 mM MgCl<sub>2</sub> and 5% glycerol) before the sample was applied to the membranes. Membranes were again washed with 50 µL Equilibration Buffer, dried and visualized by phosphorimaging. Data were quantified with ImageQuant TL Software (GE Healthcare) and fit to a binding isotherm using Prism (GraphPad Software). All experiments were carried out in triplicate. Dissociation constants and associated errors are reported in the Figure legends.

### **4.4 Results**

#### **4.4.1 Most Cas13a homologs possess pre-crRNA processing activity**

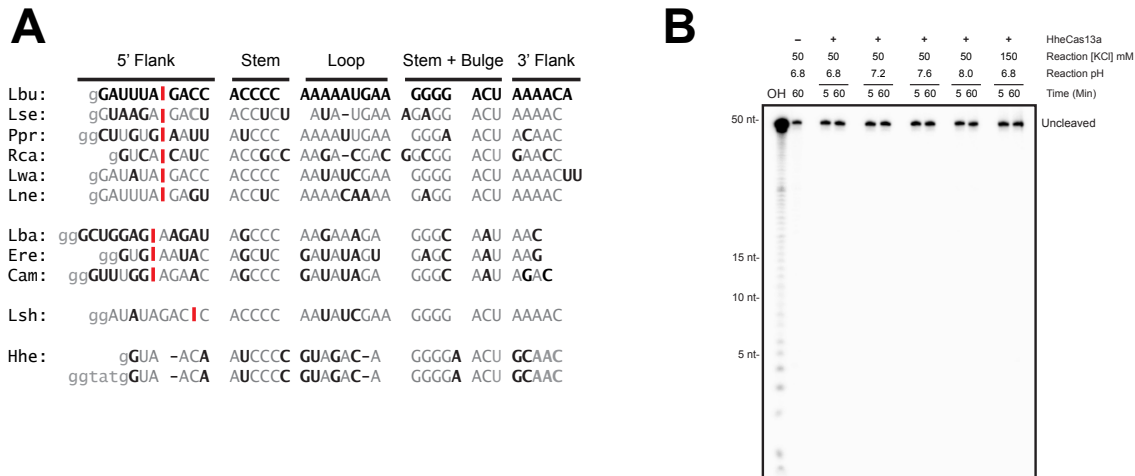
To explore the functional diversity of Cas13a proteins, we compared the pre-crRNA processing activities of ten homologs from across the protein family tree for their capacity to produce mature crRNAs from cognate pre-crRNAs (Fig. 4.1- 4.3). Similar to the three homologs studied previously (LbuCas13a, LseCas13a and LshCas13a) (East-Seletsky et al., 2016), seven additional Cas13a enzymes possess crRNA maturation activity (Fig. 4.2). Only one of the eleven Cas13a proteins tested to date exhibited no detectable cleavage of its cognate pre-crRNA across a wide range of assay conditions (HheCas13a) (Fig. 4.2B-C and 4.3B). Of the homologs that processed their native crRNAs, all but LshCas13a cleaved at the phosphodiester bond four nucleotides upstream of the conserved crRNA-repeat hairpin (Fig. 4.2B-C). These results show strong conservation of Cas13a-mediated crRNA biogenesis activity within Type VI-A CRISPR-Cas systems.





## Figure 4.2 Pre-crRNA processing is broadly conserved within the Cas13a protein family

(A) Schematic of crRNA processing reaction catalyzed by Cas13a. pre-crRNA substrates are cleaved by Cas13a to generate mature crRNAs. Below, a schematic of a pre-crRNA highlighting important functional features. (B) Alignment of the 5' portion of CRISPR repeat sequences from the studied type VI CRISPR systems highlighting the pre-crRNA cleavage site. Mapped cut cleavage sites are shown as red bars. Deviations from the Lbu crRNA-repeat sequence are noted in black text. Lowercase g's were required for transcription purposes and are not part of the native crRNA repeat sequences. Full CRISPR repeat sequence is diagrammed in Fig. 4.3A. Cleavage sites mapped in previous studies for LbuCas13a, LshCas13a and LseCas13a (East-Seletsky et al., 2016). (C) Representative gel of Cas13a-mediated pre-crRNA cleavage by nine Cas13a homologs after 60 min incubation with 5'-radiolabelled pre-crRNA substrates. Cleavage by LbuCas13a has been previously demonstrated (East-Seletsky et al., 2016) and can be observed in Fig. 4.4.



### Figure 4.3 Expanded CRISPR repeat structures and HheCas13a testing

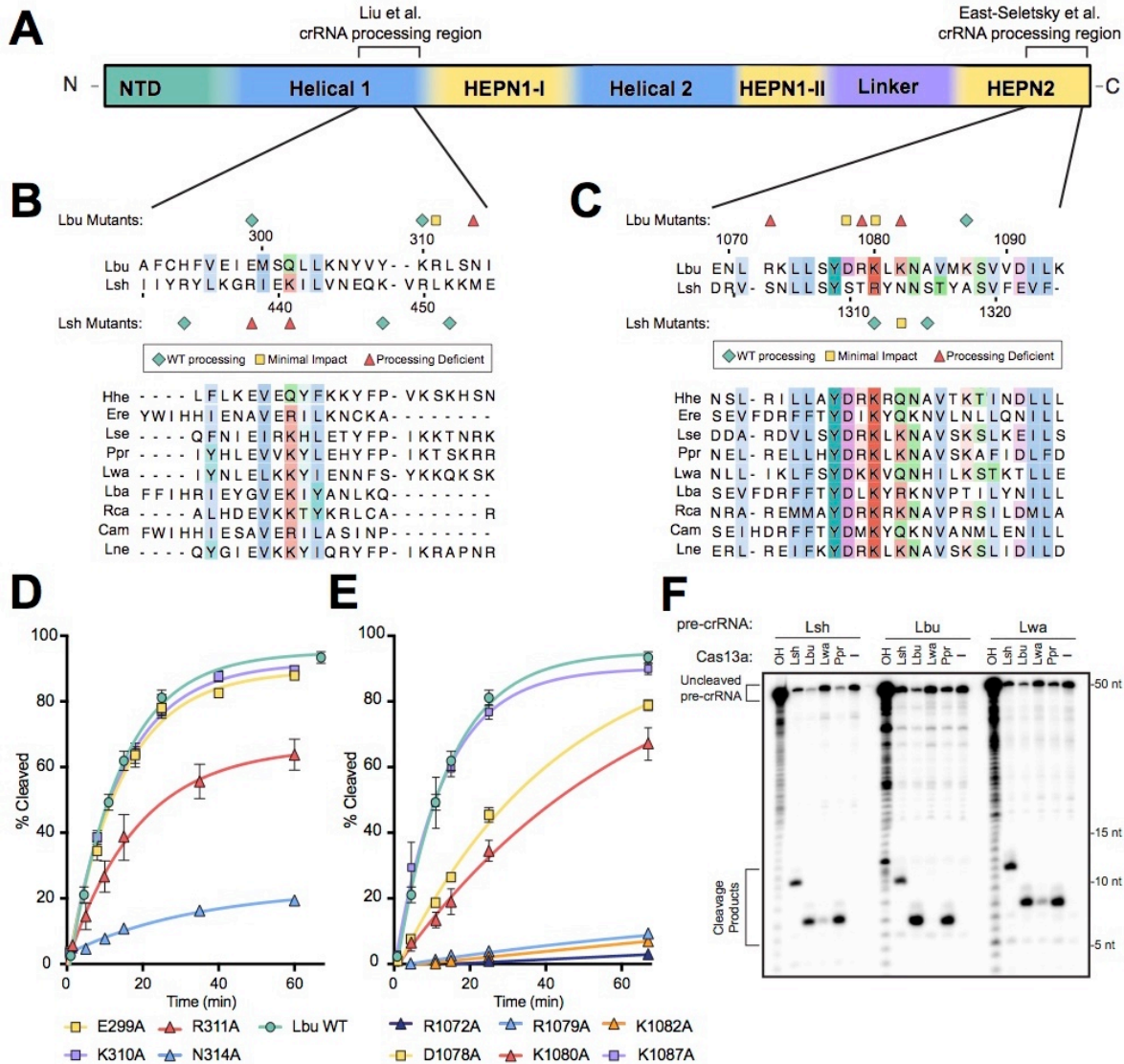
(A) Manual alignment of CRISPR repeat sequences from homologs used. pre-crRNA processing cleavage sites noted by red lines. Deviations from the Lbu crRNA-repeat sequence are noted in black text. Lowercase g's were required for transcription purposes and are not part of the native crRNA repeat sequences. Two separate Hhe crRNA sequences were tested, the first containing the native sequence and a second with four nucleotide extension to extend the atypically short native repeat. Neither crRNA repeat was cleaved by HheCas13a under any of the studied conditions (data not shown). (B) pre-crRNA processing assay with HheCas13a on native crRNA repeat sequence across variable salt and pH conditions. No cleavage products were observed.

#### 4.4.2 A conserved crRNA maturation center within most Cas13a enzymes

Previous studies have implicated two distinct regions of Cas13a as responsible for pre-crRNA processing. The general Cas13a protein architecture established by the LshCas13a crystal structure consists of an N-terminal domain and two HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domains separated by two helical domains (Fig. 4.4A) (Liu et al., 2017b). For LbuCas13a, mutation of a single residue (R1079) within the HEPN2 domain was sufficient to substantially reduce pre-crRNA processing activity (East-Seletsky et al., 2016). By contrast, mutations at two positions located in the helical 1 domain, R438 and K441, were shown to diminish pre-crRNA cleavage by LshCas13a (Liu et al., 2017b). While both of these regions may be involved in pre-crRNA processing, it is unclear whether processing by LbuCas13a is inhibited by helical 1 domain mutations, and which domain is primarily responsible for crRNA maturation across the Cas13a protein family.

First we examined the conservation of both the helical 1 and HEPN2 domains across nineteen Cas13a homologs. Previously reported alignments (Liu et al., 2017b; Shmakov et al., 2015) conflict in the helical 1 domain region, suggesting high ambiguity in the relationship between homologs in this domain. We observed minimal conservation within our alignment of the helical 1 domain implicated in pre-crRNA processing, while in contrast, consistent levels of conservation across the HEPN2 domain (Fig. 4.4B-C). The only pre-crRNA processing-defective homolog, HheCas13a, maintains a majority of the conserved charged residues throughout both domains,

suggesting that other parts of the protein or the repeat sequence may be preventing pre-crRNA cleavage. Among the homologs used in this study, LshCas13a is the most divergent across the HEPN2 domain, potentially explaining the alternative catalytic domain and atypical cleavage site selection by this homolog.



**Figure 4.4 Identification of residues important for pre-crRNA cleavage by LbuCas13a**

(A) General Cas13a domain organization schematic with annotated based off a LshCas13a crystal structure (Liu et al., 2017b). Multiple-sequence amino acid alignment of (B) the local region in Cas13a's helical 1 domain implicated in pre-crRNA processing by studies on LshCas13a, and (C) the region within in Cas13a's HEPN2 domain implicated in pre-crRNA processing by studies on LbuCas13a. Residues whose mutation severely affect pre-crRNA processing are marked by red triangles, minimal impacts on processing by yellow squares and residues whose mutation did not affect pre-crRNA processing marked with teal diamonds. Symbols above the LbuCas13a

sequences correspond to mutations made to LbuCas13a, and symbols below the LshCas13a sequence correspond to mutations made to LshCas13a by Liu et al, 2017. Coloration of the matrix alignment denotes residue conservation using the ClustalX scheme, with darker hues indicating the strength of the conservation. pre-crRNA processing under single turnover conditions measured for mutants in (D) the helical 1 domain, and (E) the HEPN2 domain. Quantified data were fitted with single-exponential decays with calculated pseudo-first-order rates constants ( $k_{\text{obs}}$ ) (mean  $\pm$  s.d.,  $n = 3$ ) as follows: Lbu WT  $0.074 \pm 0.003 \text{ min}^{-1}$ , E299A  $0.071 \pm 0.005 \text{ min}^{-1}$ , K310A  $0.071 \pm 0.003 \text{ min}^{-1}$ , R311A  $0.054 \pm 0.007 \text{ min}^{-1}$ , N314A  $0.029 \pm 0.008 \text{ min}^{-1}$ , R1079A  $0.009 \pm 0.007 \text{ min}^{-1}$ , D1078A  $0.023 \pm 0.002 \text{ min}^{-1}$ , K1080A  $0.016 \pm 0.004 \text{ min}^{-1}$ , and K1087A  $0.076 \pm 0.007 \text{ min}^{-1}$ , while R1072A and K1082A could not be fitted. (F) Representative gel of pre-crRNA processing of LshCas13a, LwaCas13a, and LbuCas13a pre-crRNAs by LbuCas13a, LshCas13a, LwaCas13a and PprCas13a proteins using standard conditions. Hydrolysis ladders for each pre-crRNA substrate demonstrate subtle differences in the migration of these fragments from differing sequences.

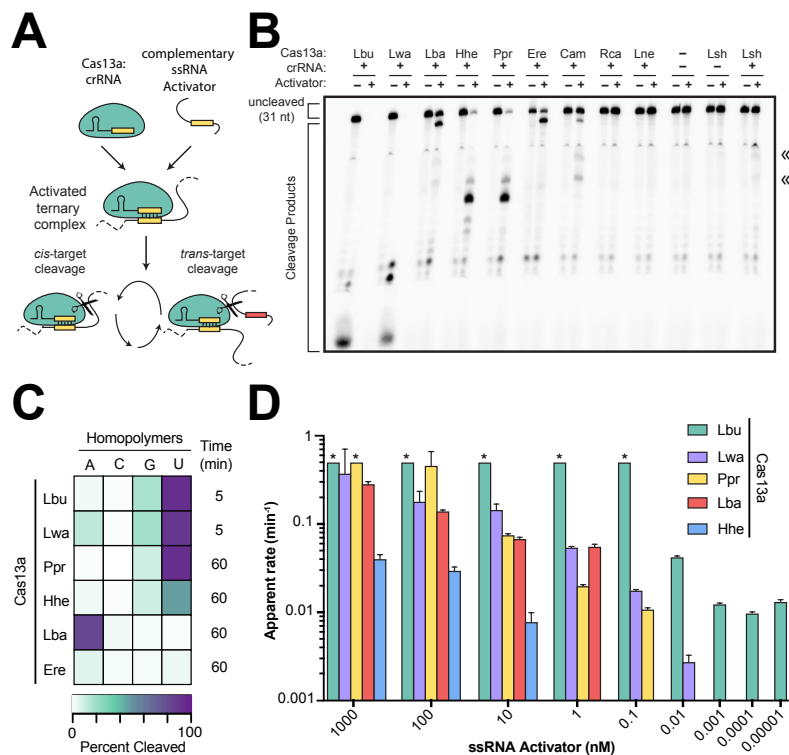
Next, we tested whether the helical 1 domain residues critical for LshCas13a pre-crRNA processing affect LbuCas13a's pre-crRNA processing activity. Due to discrepancies between alignments presented in the previous published studies, we tested four residues (E299, K310, R311 and N314) for their role in pre-crRNA cleavage (Fig. 4.4B) (East-Seletsky et al., 2016; Liu et al., 2017b; Shmakov et al., 2015). Mutation of these residues to alanine revealed a range of impacts on pre-crRNA processing efficiencies: N314A significantly reduced the cleavage rate, R311A minimally impaired activity, and E299A and K310 had no effect on pre-crRNA processing (Fig. 4.4D). In parallel, we performed mutagenesis within the HEPN2 domain of LbuCas13a. Alanine substitutions at R1072 and K1082 significantly reduced pre-crRNA cleavage, while other mutations in the same region (D1078A, K1080A and K1087A) had minimal impacts on pre-crRNA processing (Fig. 4.4E). These results suggest that the HEPN2 and to a lesser extent the helical 1 domains play significant roles in crRNA biogenesis for LbuCas13a, although the residues directly responsible for catalyzing hydrolysis remain unknown. The difference between the regions implicated in pre-crRNA processing across LbuCas13a (this study) and LshCas13a (Liu et al., 2017b) do not necessarily contradict each other, as the 5' terminus of the crRNA is held between HEPN2 and helical 1 domains in the LshCas13a structure (Liu et al., 2017b). Residues from both domains might play pivotal roles in pre-crRNA processing by either stabilizing substrate binding, promoting proper substrate orientation and/or catalyzing hydrolysis.

The lack of conservation within the HEPN2 domain of LshCas13a, its putative pre-crRNA processing active site, and this enzyme's atypical pre-crRNA cleavage site led us to hypothesize that LshCas13a may utilize a different region within the helical 1 domain to catalyze pre-crRNA processing. In the absence of a three-dimensional structure of a pre-crRNA-bound Cas13a homolog, we tested this hypothesis by mapping the LshCas13a cleavage sites on non-cognate pre-crRNAs (Fig. 4.4F). LshCas13a was able to process pre-crRNAs from LwaCas13a and LbuCas13a, generating a shifted cleavage site one nucleotide from the predicted hairpin base. In concordance with this observation, processing of the Lsh pre-crRNA by LbuCas13a, LwaCas13a and

PprCas13a occurs at the standard four-nucleotide interval from the repeat stem, differing from the cognate LshCas13a site. This supports our previous observations that the distinct LshCas13a processing site depends on the protein architecture, not the pre-crRNA sequence, and that LshCas13a is an outlier within the Cas13a tree with regard to pre-crRNA processing.

#### 4.4.3 Cas13a enzymes initiate ssRNA cleavage at either uridines or adenosines

Previous studies established that Cas13a:crRNA complexes recognize and bind complementary ssRNA targets, hereby referred to as ssRNA activators, to trigger general RNase activity at exposed uridine residues (Fig. 4.5A) (Abudayyeh et al., 2016; East-Seletsky et al., 2016). We wondered if the panel of homologs within this study retained the non-specific degradation activity previously demonstrated by LbuCas13a and LshCas13a and if the uridine preference within the HEPN active site is universal within the family. To systematically test general RNase activity, we monitored the ability of a ternary complex comprising Cas13a:crRNA with a bound ssRNA activator to degrade a *trans*-ssRNA target (Fig. 4.5A). We detected *trans*-ssRNA cleavage activity for eight of the ten homologs over the course of one hour (Fig. 4.5B, 4.6A). Most notable of the Cas13a homologs active for *trans*-ssRNA cleavage is HheCas13a, which possesses no detectable pre-crRNA processing activity, yet catalyzed complete degradation of substrates guided by its cognate pre-crRNA. While the previously characterized homologs LshCas13a (Abudayyeh et al., 2016) and LbuCas13a (East-Seletsky et al., 2016) both exhibit a preference for uridine 5' to the scissile bond, products of different lengths generated by the other homologs suggests that different active site nucleotide preferences may exist within this protein family (Fig. 4.6B).



**Figure 4.5 Members of the Cas13a protein family cleave ssRNA with a range of efficiencies**

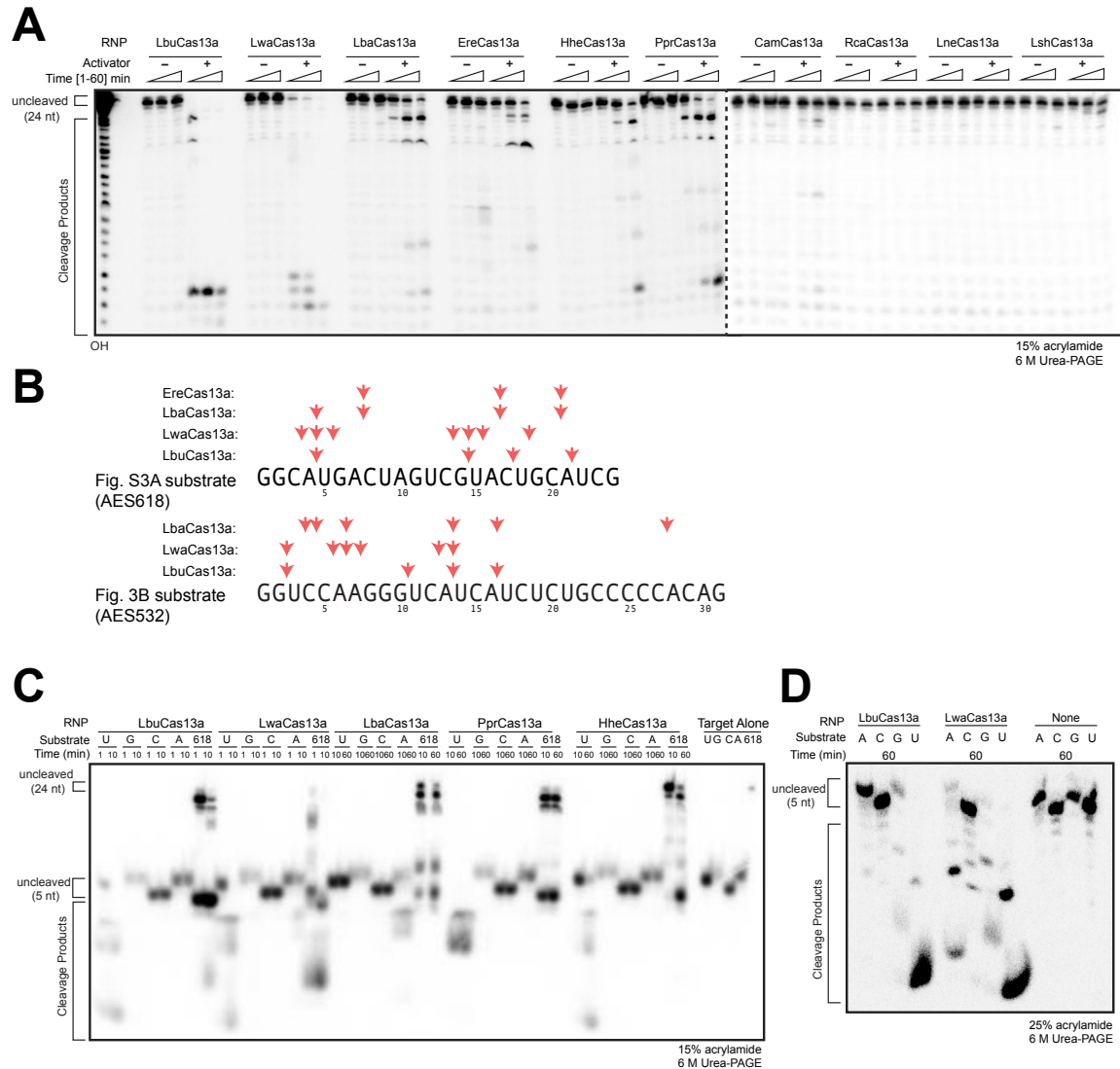
(A) Schematic of ssRNA-targeting by Cas13a. For simplicity, *trans*-ssRNA cleavage was the focus of study. (B) Representative gel of Cas13a mediated *trans*-ssRNA cleavage by all ten homologs after 60 min incubation. Cas13a:crRNA complexes were formed as described in the methods using mature crRNA products with a final RNP complex concentration of 50 nM. <1 nM radiolabeled *trans*-ssRNA target was added to initiate reaction in the presence and absence of 50 nM unlabeled, crRNA-complementary ssRNA activator. Weak *trans*-ssRNA cleavage activity was observed by LshCas13a with product bands noted by double arrows to the right. (C) Heat map reporting Cas13a-catalyzed *trans*-ssRNA cleavage percentages for each 5-mer homopolymer ssRNA substrate, for six different Cas13a. Assay conditions were identical to part (B), except LbuCas13a and LwaCas13a which were incubated for 5 min instead of 60 min. (n=3, values with associated errors presented in Table S3). (D) Apparent cleavage rates of a fluorescent ssRNA reporter by five homologs across a range of ssRNA activator concentrations. Cas13a:crRNA complexes were pre-incubated at a 2:1 ratio respectively with a final active complex concentration of 50 nM. Complementary ssRNA activator and fluorescent ssRNA cleavage reporter were added to initiate reactions. Normalized reporter signal curves timecourses were fitted with single-exponential decays and the apparent rates are plotted (n =3). Some conditions plateaued before first measured time-point therefore their rates are minimally assumed to be 0.5 min<sup>-1</sup> and are labeled with a \* in the chart.

Cas13a Homolog	rArArArArA		rCrCrCrCrC		rGrGrGrGrG		rUrUrUrUrU	
	Avg	SD	Avg	SD	Avg	SD	Avg	SD
Lbu	2.66	0.15	0.96	0.09	17.83	0.79	92.69	2.10
Lwa	13.67	0.92	0.87	0.10	18.97	1.31	90.87	1.76
Lba	82.34	5.35	3.01	1.16	2.23	0.75	1.73	0.70
Ppr	1.19	0.26	0.50	0.14	10.21	6.63	93.20	4.78
Hhe	3.56	0.51	1.01	0.12	10.34	8.55	48.04	9.07
Ere	6.32	0.18	2.42	0.03	2.37	1.58	4.25	1.99

**Table 4.3 Homopolymer cleavage percentages for six Cas13a homologs with associated errors.**

To further probe the *trans*-ssRNA cleavage nucleotide preferences of the Cas13a homologs, we measured the *trans*-cleavage capacity of these enzymes using 5-mer homopolymers of A, C, G, and U as substrates. Six homologs were able to cleave these short substrates (Fig. 4.5C, 4.6C, and Table 4.3). Four of the homologs, LbuCas13a, LwaCas13a HheCas13a and PprCas13a, exhibited preferred cleavage of the homouridine substrate, although secondary preferences were observed for the homologs with the highest activities (Fig. 4.5C and 4.6D). In contrast, LbaCas13a, and EreCas13a preferred homo-adenosine in agreement with biochemically mapped cleavage sites on longer targets (Fig. 4.6B). Identical product generation from these long substrates by CamCas13a is consistent with adenosine preference by this clade of the Cas13a family tree (Figs. 4.1, 4.5 and 4.6).





### Figure 4.6 *trans*-ssRNA cleavage by Cas13a homologs

(A) Time course analysis *trans*-ssRNA cleavage by ten different Cas13a homologs in the presence and absence of ssRNA activator. Time points were taken at 1, 10 and 60 min. ssRNA-activator specific cleavage products are noted for LbuCas13a, LwaCas13a, LbaCas13a, EreCas13a, HheCas13a, PprCas13a, CamCas13a, and LshCas13a. Dotted line denotes boundary between separate PAGE gels. (B) Mapped *trans*-ssRNA cleavage sites across multiple cleavage reactions for four Cas13a homologs. Different cleavage patterns are noted by red arrows. It appears LbaCas13a and EreCas13a may have an adenosine preference, while LwaCas13a appears to be more promiscuous with respect to nucleotide preference. (C-D) Representative *trans*-ssRNA cleavage gels of homopolymer ssRNA substrates by five Cas13a homologs.

One notable difference between these enzymes was the rate at which *trans*-ssRNA cleavage reaches saturation under the tested conditions of equimolar ssRNA activator and Cas13a:crRNA interference complex. These enzymatic differences could

have dramatic effects on Cas13a's biological role, so we next aimed to quantify the variance in *trans*-ssRNA cleavage within the Cas13a homolog family. We developed a high-throughput screen utilizing a short fluorescent ssRNA reporter for RNA cleavage to account for both ssRNA activator binding and *trans*-ssRNA cleavage, the two core properties of Cas13a enzymes that contribute to total enzymatic output. To interrogate the sensitivity of each Cas13a homolog, decreasing amounts of complementary ssRNA activator were added to initiate the reaction, and the apparent rate of fluorescent ssRNA reporter cleavage was calculated from each of the resulting timecourses. While the calculated rates are a convolution of the ssRNA activator binding affinity and the catalytic turnover rate for each of the enzymes, they give a relative measure of cleavage activity that is comparable across homologs.

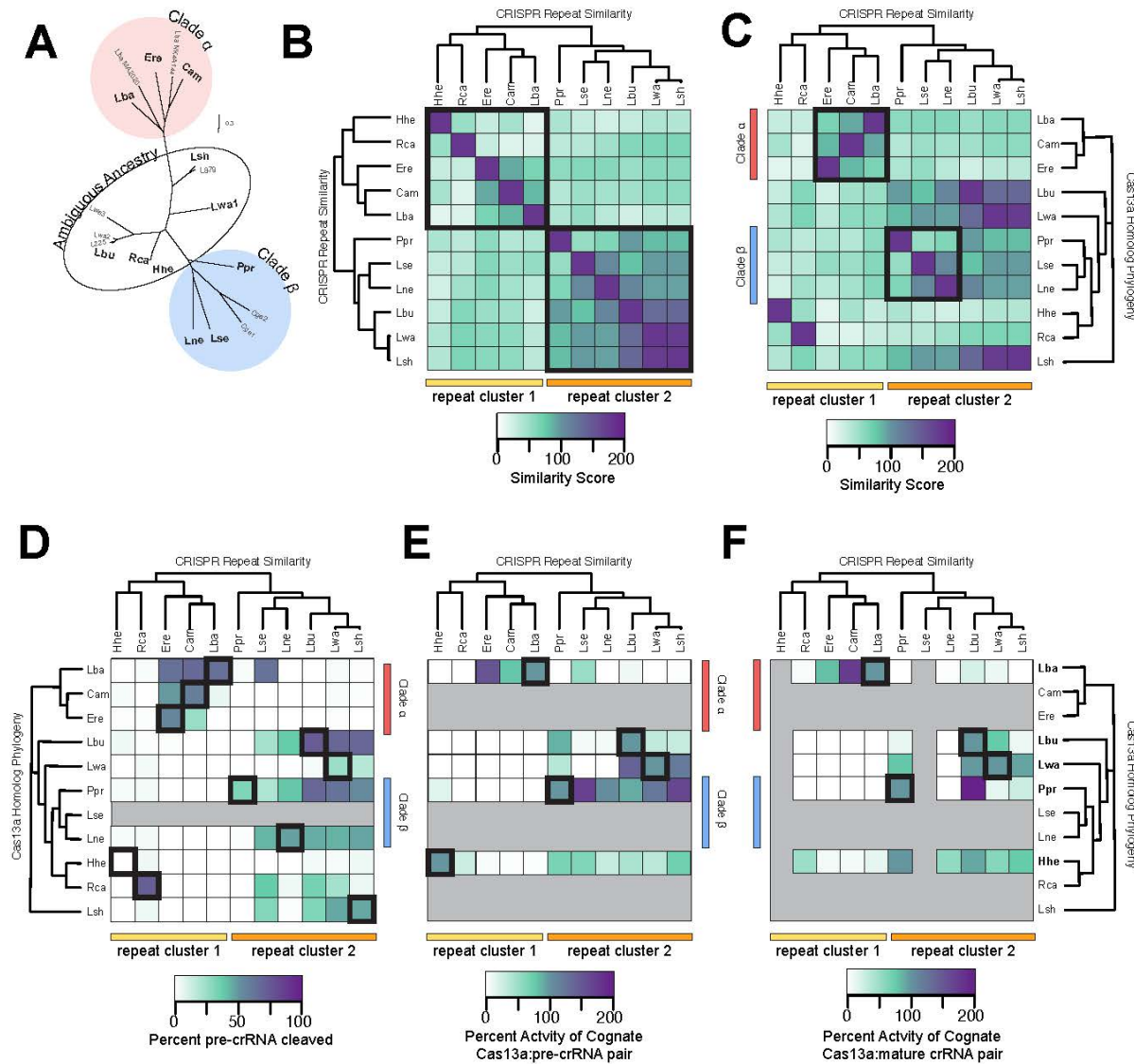
Five homologs (LbuCas13a, LwaCas13a, LbaCas13a, HheCas13a and PprCas13a) demonstrated sufficiently detectable cleavage activity within this assay for reproducible analysis. Of these five homologs, LbuCas13a exhibited the most sensitivity, with detectable reporter cleavage in the presence of only 10 fM complementary activator (Fig. 4.5D). Only two homologs, LwaCas13a and PprCas13a, displayed enough activity to detect the activator in the picomolar range with sensitivities of 10 pM and 100 pM, respectively. LbaCas13a and HheCas13a were much less sensitive, only becoming active at nanomolar levels of reporter, which is close to equimolar relative to the Cas13a complex. Since this assay relies on a substantial number of *trans*-cleavage events to produce detectable fluorescence, we can assume that the three homologs unable to produce detectable signal above background despite similar cleavage site preferences (EreCas13a, CamCas13a, and LshCas13a) possess even less sensitive complementary target sensitivities. The remarkably broad range in sensitivities ( $\sim 10^7$ -fold) suggests a diverse capacity of Cas13a enzymes to protect a host organism from foreign RNA.

#### **4.4.4 CRISPR repeat sequence determines non-cognate pre-crRNA processing**

The dual activities of Cas13a provide an opportunity to study the interdependence of pre-crRNA processing and targeting between distinct Type VI-A CRISPR-Cas operons. To determine the substrate requirements for both activities, we asked to what extent different homologs can recognize non-cognate crRNAs for guide processing and targeting. Initially, we hoped to predict bioinformatically the likely crRNA exchangeability through phylogenetic analysis of the Cas13a family and crRNA similarities. This analysis suggested that two distinct clades of homologs exist, termed alpha and beta for clarity (Stamatakis, 2014a) (Fig. 4.7A). Five of our purified homologs exist outside of these clades with ambiguous ancestral relationships, leading us to wonder if the pre-crRNA (CRISPR repeat) sequence might dictate functional orthogonality, or incompatibility of these sequences crRNA sequences to facilitate *trans*-ssRNA cleavage when paired with a non-cognate Cas13a protein. Due to the short and structured content of the CRISPR repeats, we used a pairwise sequence alignment score matrix to build a hierarchical clustering relationship between the CRISPR repeats to score the variation across the family (Burstein et al., 2016). Surprisingly, this analysis pointed to the existence of two crRNA clusters, overlapping but distinct from the protein clades determined by the amino acid sequences (Figs. 4.7B-C). While cluster 1 crRNAs correlate well with a subset of the alpha-clade proteins and all the beta-clade associated



crRNAs are within cluster 2, the homologs with ambiguous phylogenetic relationships are split across the two clusters (Fig. 4.7C).



**Figure 4.7 crRNA exchangeability within the Cas13a family**

(A) Maximum-likelihood phylogenetic tree of all Cas13a family members. Homologs used in this study are bolded and clades are highlighted. Bootstrapped values are located in previous work (East-Seletsky et al. 2016). (B) Symmetrical similarity score matrix for CRISPR repeats from homologs used in this study. Rows and columns are ordered by CRISPR repeat clustering. (C) Asymmetrical similarity score matrix for CRISPR repeats from homologs used in this study. The same pairwise scores are presented here as in (B), except the rows are reordered to correspond to the Cas13a phylogenetic tree of the subset of homologs used in this study. Bootstrap values for this smaller tree are in Fig. 4.1. (D-F) Functional activity matrix for (D) pre-crRNA processing by non-cognate proteins, (E) *trans*-ssRNA cleavage directed by pre-crRNAs, and (F)

*trans*-ssRNA cleavage directed by mature crRNAs. Processing assays were performed using standard conditions and 60 min reaction endpoints were analyzed. *trans*-ssRNA cleavage assays were performed using the fluorescent ssRNA reporter assay with fitted initial rates. ssRNA activator concentrations were as follows: LbuCas13a:100 pM, LwaCas13a:100 pM, PprCas13a:100 nM, LbaCas13a:10 nM, and HheCas13a:100 nM. Initial rates were fit across three replicates to account for differences in fluorescence plateau values and normalized to each Cas13a:crRNA cognate pair. See Tables 4.4-4.6 for numerical values and associated errors (n=3).

Unable to easily predict Cas13a:crRNA orthogonality using bioinformatic analysis alone, we tested the extent of functional exchangeability between non-cognate crRNAs for both processing and *trans*-ssRNA target cleavage by each of the Cas13a homologs (Fig. 4.7D-F). For pre-crRNA processing, we found that the crRNA clusters defined by the pairwise sequence comparisons predicted their ability to be processed by their associated Cas13a proteins (Fig. 4.7D, Table 4.4). For example, pre-crRNAs from cluster 1 are only processed by the proteins of clade alpha and vice versa for cluster 2 and clade beta. In contrast, the protein classification is less predictive, as most of the ambiguously classified proteins could process sequences from repeat cluster 2, independent of where their repeat sequences were clustered.

Descriptive ID	RNA ID	Lbu WT		CamCas13a		EreCas13a		LbaCas13a		LwaCas13a		RcaCas13a		HheCas13a		PprCas13a		LshCas13a		LnyCas13a	
		Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD
pre-crLbu	AES634	82.0	3.1	0.0	3.2	0.3	0.7	0.2	0.7	3.6	2.1	44.0	29.7	4.7	3.1	78.5	5.5	40.8	9.7	58.8	38.5
pre-crLse	AES625	29.1	11.9	0.8	1.3	0.3	1.7	74.7	7.6	0.9	0.9	49.3	28.4	-0.5	2.2	27.0	4.7	39.9	7.8	55.9	21.1
pre-crLba1	AES555	1.7	2.7	15.3	6.8	2.5	1.6	75.9	7.0	2.5	3.1	2.5	2.6	0.8	2.0	1.3	3.0	-0.2	1.8	2.8	2.2
pre-crLba2	AES556	0.1	1.9	7.6	2.5	-0.1	3.7	75.4	7.1	0.4	3.4	-0.7	4.1	-1.9	5.0	-2.2	5.5	-2.5	4.1	0.3	7.3
pre-crEre2	AES557	1.6	1.4	63.9	6.0	69.7	1.7	76.2	6.4	0.7	1.4	-1.4	3.8	-2.2	3.7	-1.4	3.5	-2.7	8.3	-1.8	6.3
pre-crEre1	AES558	-0.9	2.3	68.5	9.5	66.4	6.8	74.5	9.7	-0.3	2.7	0.4	3.4	-1.7	5.1	0.2	4.9	-0.5	7.1	2.5	3.2
pre-crCam	AES559	-0.1	1.2	72.1	4.4	33.6	5.0	77.7	3.4	0.7	0.9	-0.2	0.4	-1.2	0.6	-1.3	1.1	-0.6	0.7	0.6	0.3
pre-crLwa	AES553	78.4	9.4	2.8	1.8	1.1	2.6	2.6	2.0	32.7	10.1	29.4	8.6	3.5	2.3	74.4	9.3	59.9	9.5	56.6	5.8
pre-crLsh	AES566	79.0	7.0	1.3	2.7	0.8	1.7	1.9	2.4	24.0	9.4	6.1	1.8	5.4	8.1	69.0	11.5	59.9	9.6	60.9	16.6
pre-crRca	AES568	1.6	1.4	0.8	2.0	1.5	2.8	2.1	2.3	4.6	2.7	80.7	3.4	5.4	3.6	4.4	2.0	5.4	5.0	4.4	3.5
pre-crPpr	AES597	-1.0	2.2	-0.8	1.2	-1.0	1.8	2.1	2.3	-1.0	2.2	-0.2	2.8	-1.5	2.3	41.3	4.9	-1.0	2.5	1.5	1.9
pre-crHhe	AES620	5.8	4.5	3.9	14.3	1.9	10.4	-1.0	6.9	0.5	4.4	-2.8	9.7	-6.6	5.2	3.5	9.4	-6.5	6.6	3.8	10.6
pre-crLne	AES619	48.6	1.7	0.7	2.3	-0.5	1.3	-0.1	0.3	-0.9	3.2	5.6	3.6	-2.4	1.5	50.6	4.4	1.3	2.9	61.4	1.1

**Table 4.4 pre-crRNA processing cleavage percentages for ten Cas13a homologs with associated errors**

Three homologs (PprCas13a, HheCas13a, and Rca13a) were pre-crRNA processing outliers with respect to their position within the crRNA clusters. HheCas13a was unable to cleave any non-cognate pre-crRNAs, and conversely, no homolog, including HheCas13a, processed the Hhe pre-crRNA. This suggests that the inability of HheCas13a to process its cognate pre-crRNA reflects not just the divergent repeat sequence (Fig. 4.2 and 4.3), but also the loss of pre-crRNA processing activity within the protein. In contrast, the deviating activity of PprCas13a and RcaCas13a is explained by crRNA repeat sequence divergence. PprCas13a and RcaCas13a process their own crRNAs, yet both also process the crRNA repeat cluster 2 non-cognate sequences. The PprCas13a crRNA repeat sequence differs from the other cluster 2 crRNA repeats across the 5' flanking region cleavage site, suggesting greater substrate flexibility for pre-crRNA processing by PprCas13a. Similarly, a distinguishing sequence feature of the RcaCas13a crRNA is an extended six-base pair stem-loop relative to the standard

five-base pair stem-loop present in crRNAs of the rest of the family. It is worth noting that the positional substitution tolerance within each crRNA repeat for Cas13a pre-crRNA processing is consistent with our previous mutation studies of the LbuCas13a:crRNA complex, and recent structural insights obtained of the LshCas13a:crRNA complex (East-Seletsky et al., 2016; Liu et al., 2017b). Overall, these results suggest that the sequence of a Type VI-A CRISPR repeat dictates its capacity for pre-crRNA processing by the Cas13a family. However, homologs that evolved in the presence of divergent repeats (PprCas13a and RcaCas13a) retain the capacity to process other cluster 2 sequences.

#### 4.4.5 Two subfamilies of functionally orthogonal Cas13a enzymes

Next we wondered whether the pre-crRNA processing exchangeability clusters defined in Fig. 4.7D were competent for directing *trans*-ssRNA cleavage by non-cognate Cas13a homologs. To study Cas13a-mediated *trans*-ssRNA cleavage directed by non-cognate pre-crRNAs and mature crRNAs, we modified the previously described fluorescence assay and limited our analysis to the five homologs that exhibited significant cleavage activity in previous ssRNA-cleavage experiments (see Fig. 4.5D). Broadly, these results mirrored the pre-crRNA processing results, with the crRNA repeat cluster identity determining functional groups, but with some striking contrasts consistent with processing and targeting being independent enzymatic activities (Fig. 4.7E-F and Table 4.5-4.6). For instance, the Ppr pre- and mature crRNA can direct ssRNA cleavage by non-cognate proteins LwaCas13a and LbuCas13a, despite their inability to process these pre-crRNAs. Another surprise is the promiscuity of HheCas13a, which is directed by all cluster 2 pre- and mature crRNAs for *trans*-ssRNA cleavage, despite lacking pre-crRNA processing activity with any of these guides. This suggests that crRNA maturation is not required for *trans*-ssRNA cleavage, an observation in agreement with our previous finding that LbuCas13a pre-crRNA processing-deficient mutants possess unaffected *trans*-ssRNA cleavage capacity (East-Seletsky et al., 2016).

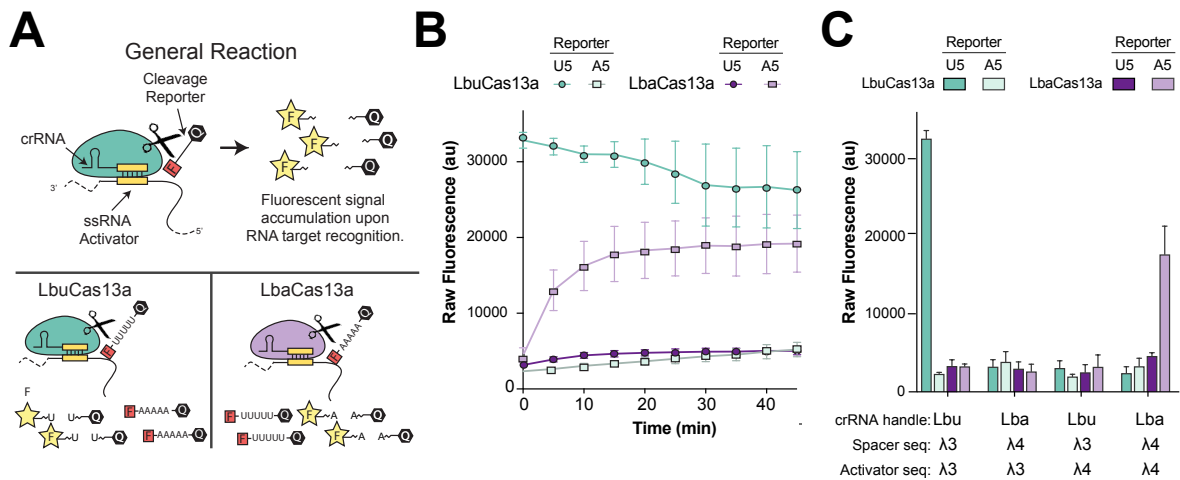
Guide	LwaCas13a		LbaCas13a		LbuCas13a		PprCas13a		HheCas13a	
	slope	SE	slope	SE	slope	SE	slope	SE	slope	SE
Lwa pre-crRNA	100.0	4.2	124.5	37.4	35.6	1.9	31.5	6.4	1.5	0.2
Lba pre-crRNA	-0.3	0.0	0.0	0.0	1.8	0.1	-0.6	-0.1	100.0	15.8
Ere pre-crRNA	-0.2	0.0	-0.1	0.0	3.2	0.2	0.0	0.0	152.5	27.3
Cam pre-crRNA	-0.3	0.0	0.0	0.0	2.7	0.2	0.0	0.0	77.8	10.0
Lsh pre-crRNA	120.7	7.7	166.0	49.8	71.7	4.8	30.7	6.3	2.4	0.3
Rca pre-crRNA	0.0	0.0	0.0	-0.7	26.6	1.4	0.1	0.0	0.8	0.1
Ppr pre-crRNA	37.4	1.4	100.0	42.4	66.4	4.3	86.7	22.8	1.5	0.2
Lne pre-crRNA	-0.2	0.0	104.6	31.4	29.8	1.6	7.1	1.6	1.3	0.2
Hhe pre-crRNA	0.0	0.0	16.2	4.9	100.0	7.3	-0.1	0.0	0.8	0.1
Lse pre-crRNA	1.5	0.0	170.2	51.1	63.7	3.6	6.6	1.8	51.5	6.0
Lbu pre-crRNA	140.0	12.0	90.0	27.0	42.6	2.2	100.0	28.0	7.8	0.9

**Table 4.5 Normalized initial *trans*-target rates for five Cas13a homologs directed by pre-crRNAs with associated errors.**

Guide	LwaCas13a		LbaCas13a		LbuCas13a		PprCas13a		HheCas13a	
	slope	SE	slope	SE	slope	SE	slope	SE	slope	SE
Cam crRNA	0.0	0.1	174.8	50.9	-0.2	0.0	-0.9	-0.1	7.4	0.4
Lba crRNA	-0.4	-0.1	100.0	22.3	0.0	0.0	-1.2	-0.1	12.6	0.7
Ere crRNA	-0.5	-0.1	81.7	23.5	0.1	0.0	-0.8	-0.1	6.5	0.4
Ppr crRNA	77.8	14.3	1.5	0.2	11.6	0.7	100.0	2.6	100.0	7.5
Lne crRNA	0.1	0.0	1.6	0.3	0.1	0.0	-0.2	-0.1	54.7	3.0
Rca crRNA	-0.1	-0.1	3.9	0.6	-0.8	0.0	-0.1	-0.1	50.4	2.7
Lsh crRNA	93.3	21.5	0.9	0.2	10.0	0.5	27.2	1.1	75.9	4.5
Lwa crRNA	100.0	24.3	10.6	1.7	75.7	5.5	12.8	0.4	60.4	3.7
Lbu crRNA	98.2	18.2	21.1	3.4	100.0	6.1	398.3	20.8	92.6	7.0

**Table 4.6 Normalized initial trans-target rates for five Cas13a homologs directed by mature crRNAs with associated errors.**

Comparison of crRNA exchangeability for both pre-crRNA processing and *trans*-ssRNA cleavage defines two functionally orthogonal subfamilies within the Cas13a protein family. The first group (i.e. LbuCas13a) has wide promiscuity for both pre-crRNA processing and *trans*-ssRNA cleavage directed by crRNAs from across the protein family. This polyphyletic group also shows a preference for uridine within the *trans*-ssRNA cleavage active site. The second group (i.e. LbaCas13a), defined both by crRNA and protein sequences, has a distinct crRNA exchangeability profile and preferentially cleaves at adenosines during *trans*-ssRNA target cleavage. To test orthogonal reactivities of these Cas13a subfamilies, we first verified that LbuCas13a and LbaCas13a cleaved homo-A or homo-U ssRNA reporters, respectively (Fig. 4.8A-B). While the different probes generated similar amounts of fluorescent signal, it should be noted that substantially different quantities of ssRNA activator were added due the differential sensitivities of the homologs (10 pM vs 1 nM for LbuCas13a and LbaCas13a, respectively). To verify orthogonality, a panel of control reactions with all possible non-cognate combinations between crRNA, activator and reporter were tested, with no substantial signal detected except for the cognate combinations (Fig. 4.8C). Taken together, these results define distinct Cas13a homologs that can function in parallel within the same system.

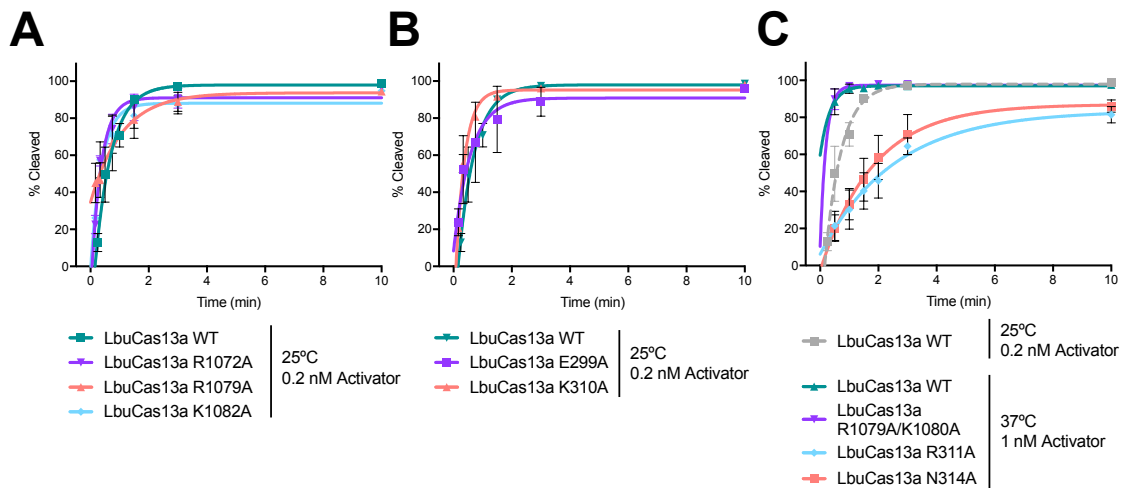


**Figure 4.8 Functional validation of orthogonal Cas13a subfamilies for RNA detection**

(A) Schematic of the RNA detection assay modified to use fluorescent homopolymer ssRNA reporter substrates to assay *trans*-ssRNA cleavage activation by either LbuCas13a or LbaCas13a. (B) Timecourse of raw fluorescence measurements generated by homopolymer reporters incubated with either LbuCas13a: Lbu-crRNA: 10 pM ssRNA activator or LbaCas13a: Lba-crRNA: 1 nM ssRNA activator (mean  $\pm$  s.d., n=3). (C) Raw fluorescence measurements generated by the fluorescent homopolymer ssRNA reporters across a panel of crRNA, ssRNA activator, and Cas13a protein combinations (mean  $\pm$  s.d., n=3).

#### 4.4.6 Pre-crRNA processing enhances targeting efficiencies within the context of a CRISPR array

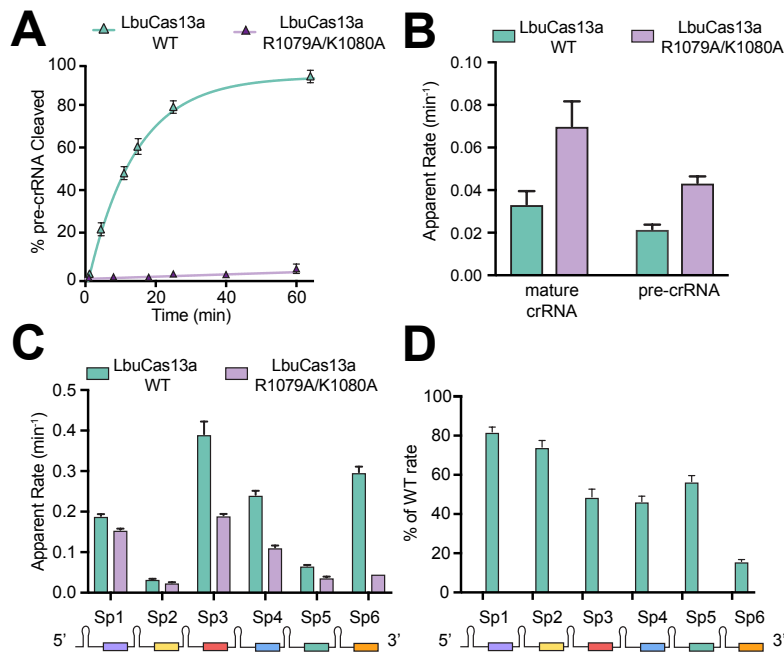
One puzzling finding of this study is the lack of a stringent requirement for mature crRNA to trigger the subsequent *trans*-ssRNA target cleavage reaction by Cas13a. Additionally, processing deficient mutants of LbuCas13a maintain similar efficiencies of *trans*-ssRNA cleavage, even when directed by a pre-crRNA instead of a mature crRNA (Fig. 4.9 and 4.10). This led us to hypothesize that the role of pre-crRNA processing within Type VI CRISPR loci is not necessary for ssRNA targeting but instead serves to liberate each crRNA from the confines of a long CRISPR array transcript. We wondered whether pre-crRNA processing might relieve RNA folding constraints and potential steric hindrance of neighboring Cas13a:crRNA-spacer species during crRNA loading and/or ssRNA targeting. To test this, we compared the efficiency of *trans*-ssRNA cleavage directed by a CRISPR array using either wildtype LbuCas13a or a pre-crRNA processing-inactive mutant.



**Figure 4.9 *trans*-cleavage by LbuCas13a point mutants in regions implicated in crRNA-processing.**

(A-C) *trans*-ssRNA cleavage by various pre-crRNA processing ‘implicated’ point mutants of LbuCas13a. Cleavage reactions were performed with 50 nM Cas13a:crRNA at either 25°C or 37°C with 0.2 nM or 1 nM ssRNA activator as noted. Lower temperature and activator concentrations were used to slow down reaction kinetics for more accurate measurements. Fitted curves are single-exponential decays with

calculated pseudo-first-order rates constants (mean  $\pm$  s.d.,  $n = 3$ ) as follows: 25°C conditions: Lbu WT  $1.76 \pm 0.24 \text{ min}^{-1}$ , K299A  $1.72 \pm 0.65 \text{ min}^{-1}$ , K310A  $3.19 \pm 0.27 \text{ min}^{-1}$ , R1072A  $2.95 \pm 0.53 \text{ min}^{-1}$ , R1079A  $0.93 \pm 0.28 \text{ min}^{-1}$ , and K1082A  $2.64 \pm 0.63 \text{ min}^{-1}$  and 37°C conditions: R311A  $0.39 \pm 0.09 \text{ min}^{-1}$ , and N314A  $0.54 \pm 0.11 \text{ min}^{-1}$ , while the wildtype and processing inactive mutant (R1079A/K1080A) plateaued too quickly for an accurate rate measurement: WT  $2.95 \pm 1.97 \text{ min}^{-1}$ , and R1079A/K1080A  $4.86 \pm 4.99 \text{ min}^{-1}$ .

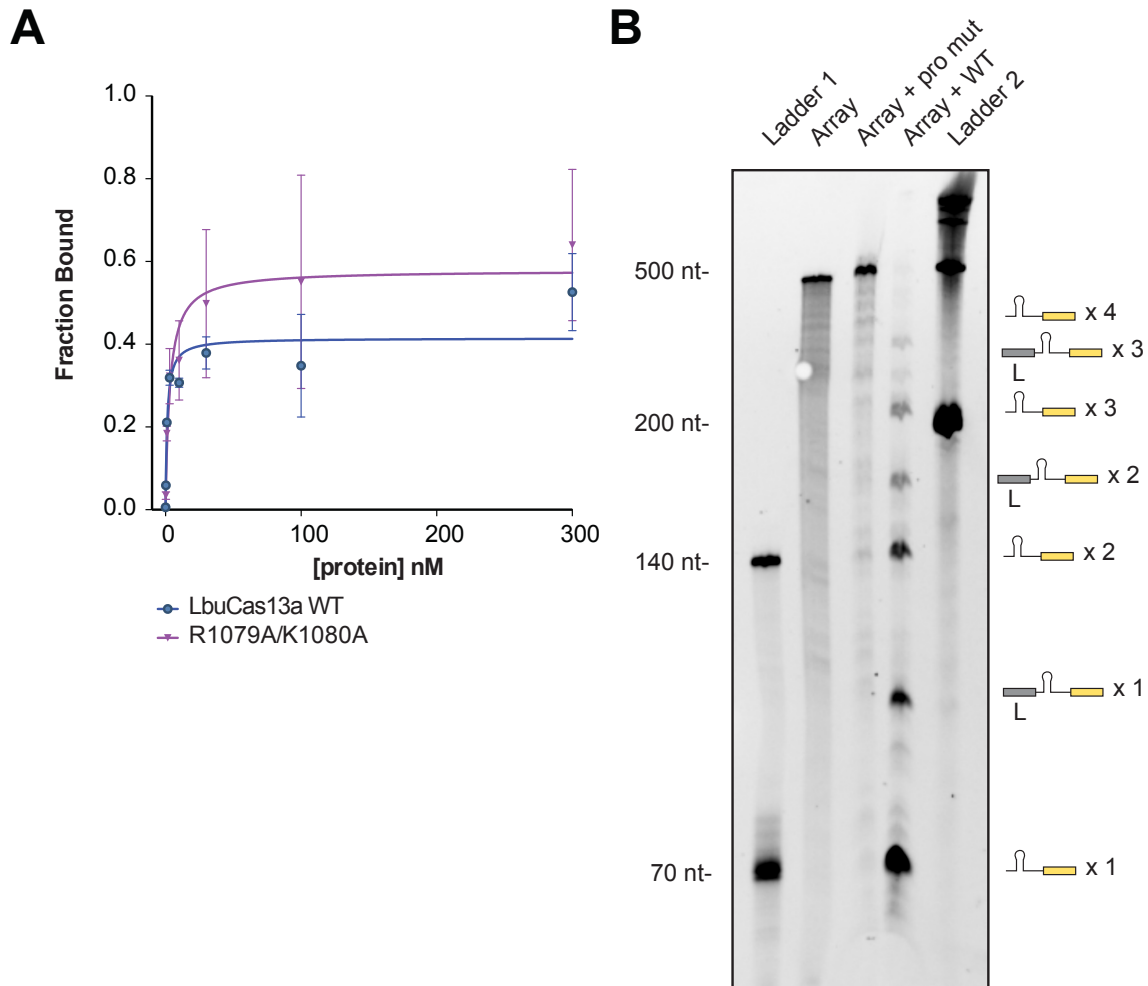


#### Figure 4.10 Deciphering the role of crRNA array processing for LbuCas13a

(A) Quantified timecourse data of pre-crRNA processing assays for R1079A/K1080A mutant compared to wildtype LbuCas13a. Quantified data was fitted to single-exponential decays and pseudo-first-order rate constant ( $k_{\text{obs}}$ ) (mean  $\pm$  s.d.,  $n = 3$ ) for LbuCas13a WT of  $0.074 \pm 0.003 \text{ min}^{-1}$ , while the R1079A/K1080A mutant could not be fit with sufficient confidence to yield a rate constant. (B) Apparent rate of fluorescent reporter by LbuCas13a wildtype and R1079A/K1080A processing inactive mutant as directed by pre-crRNA and mature crRNAs. Cas13a:RNA complexes were pre-incubated for 60 min at a 1:1 ratio, and then 10 pM of activator and 150 nM reporter were added to initiate reaction. (mean  $\pm$  s.d.,  $n = 3$ ) (C) Apparent rates of fluorescent ssRNA reporter cleavage by 300 nM wildtype LbuCas13a or R1079A/K1080A pre-crRNA processing inactive mutant as directed by 50 nM of a CRISPR array containing six crRNA repeat-spacers. Each bar group represents the addition of 100 pM of a distinct ssRNA activator sequence complementary to schematized positions within the CRISPR array indicated below each bar group. Each rate is fitted from data from three biological replicates and the standard deviation of the rate is depicted. Mutant protein rate is statistically different from the wildtype LbuCas13a for all spacer positions (one-sided t-test:  $p < 0.001$ ). (D) Data from (C) depicted as a percentage of wildtype LbuCas13a activity demonstrating the positional effect of the decreased *trans*-ssRNA



targeting efficiencies by the pre-crRNA processing inactive mutant. Processing of the array is shown in Fig. 4.11.

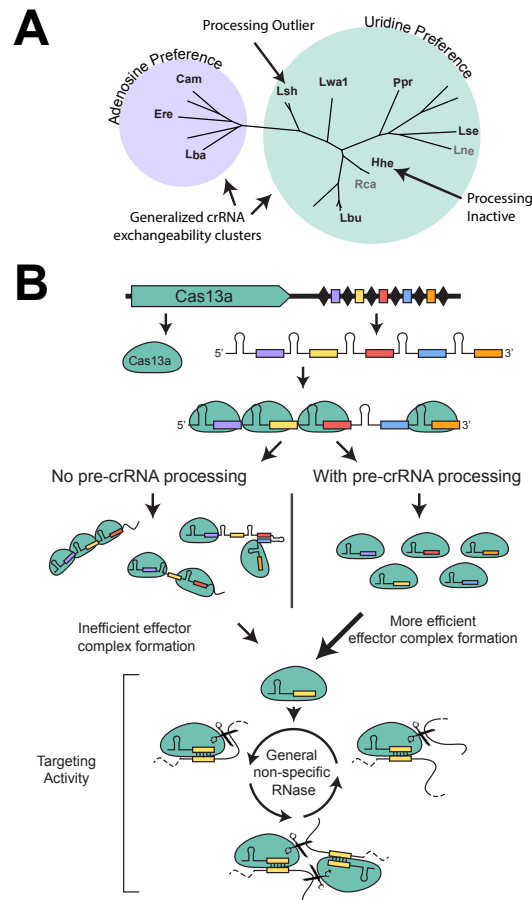


**Figure 4.11 crRNA processing inactive mutant R1079A/K1080A retains similar crRNA binding affinity and does not process a pre-crRNA array**

(A) Filter binding assays were conducted as described in the Methods to determine the binding affinity of mature crRNA to LbuCas13a WT and LbuCas13a R1079A/K1080A. The quantified data were fit to standard binding isotherms. Measured dissociation constants from three independent experiments (mean  $\pm$  sd) were  $1.21 \pm 0.57$  nM (LbuCas13a WT), and  $3.11 \pm 0.89$  nM (LbuCas13a R1079A/K1080A). (B) pre-crRNA processing assay using a six-mer CRISPR array as the substrate with LbuCas13a and LbuCas13a R1079A/K1080A mutant along with various size markers. Product identities are depicted to the right of the gel. Due to an additional leader region that was occasionally not processed, additional sized products occasionally occurred as noted.

Since all processing-defective single point mutants of LbuCas13a (Fig. 4.4) retained low levels of pre-crRNA processing activity, we created a double mutant (R1079A/K1080A) that possessed no detectable processing activity, yet retained *trans-*

ssRNA cleavage efficiencies similar to or greater than wildtype LbuCas13a (Fig. 4.10). Next, this mutant and the wildtype LbuCas13a enzyme were tested for ssRNA cleavage in the presence of a target RNA and a small CRISPR array transcript consisting of six distinct repeat-spacer units (Fig. 4.10C-D and 4.11). The rate of *trans*-ssRNA cleavage by the crRNA-processing inactive mutant was significantly reduced for all spacer sequences within the array (one-sided t-test:  $p < 0.001$  for all pairs). The reduced activity compared to wildtype LbuCas13a is more pronounced with each successive spacer within the array, with the last spacer directing cleavage with a rate that is only 15% of that catalyzed by the wildtype enzyme. This finding suggests that while pre-crRNA processing is not necessary for targeting, it enhances activity by liberating crRNAs from the CRISPR array, leading to a revised model incorporating the interplay between crRNA-processing and ssRNA-targeting within Type VI-A CRISPR-Cas systems (Fig. 4.12).



### Figure 4.12 A revised model for Type VI-A CRISPR-Cas system

(A) Graphical summary of key findings in this study. Homologs used in this study are indicated with abbreviations in bold, with *trans*-ssRNA cleavage inactive homologs depicted in grey. Colored circles highlight the two orthogonal Cas13a enzyme groups, as defined by their generalized crRNA exchangeability and *trans*-ssRNA cleavage



substrate nucleotide preference. (B) Revised model of Type VI-A CRISPR-Cas systems incorporating the interplay between crRNA maturation and RNA targeting efficiency.

#### 4.5 Discussion

Cas13a is a unique dual ribonuclease that catalyzes two of the three steps of Type VI-A CRISPR-Cas based adaptive immunity, enabling both crRNA maturation and target RNA destruction by the same enzyme. Based on the large divergence in protein sequences and corresponding CRISPR arrays within the Cas13a enzyme family, it was bioinformatically unclear if the entire family maintained consistent functionality. Biochemically, however, analysis of ten Cas13a homologs defined two subfamilies of enzymes that recognize distinct crRNA sequences and possess different nucleotide cleavage preferences during *trans*-ssRNA degradation. These data also showed that despite being conserved in most Cas13a enzymes, pre-crRNA processing activity is not strictly required for *trans*-ssRNA cleavage. In addition, we found that the kinetics of Cas13a-catalyzed ssRNA cleavage range over seven orders of magnitude, raising the possibility that these enzymes can operate in both regulatory and cell-destruction pathways.

Beyond sharing respective sets of crRNAs, the two orthogonal subfamilies of Cas13a homologs defined in this study cleave ssRNAs preferentially at Us versus As. The evolutionary basis for these substrate preferences, which pertain to both independent active sites of the enzyme, are not yet known. It is possible that ancestral versions of these enzymes evolved in response to nucleotide composition of the host and phage transcriptomes, resulting in differing active site molecular architectures. We anticipate that other Type VI CRISPR-Cas enzymes, such as Cas13b, may also include subfamilies with divergent nucleotide preferences. Structural studies will be required to elucidate the molecular basis for *trans*-ssRNA substrate recognition to rationalize these preferences, and to potentially engineer alternative cleavage activities.

The observation that pre-crRNAs can support ssRNA targeting suggests that unlike other CRISPR-Cas systems, crRNA maturation is not essential for assembly into functional surveillance complexes. Alternative mechanisms of generating mature crRNAs have been observed (Zhang et al., 2013), but in the absence of processing, these CRISPR-Cas complexes are not competent for phage defense (Brouns et al., 2008; Deltcheva et al., 2011; Hatoum-Aslan et al., 2014; Maier et al., 2015; Semenova et al., 2015). In contrast, our data suggest that Type VI systems can employ pre-crRNAs, even within the context of a CRISPR array, to trigger ssRNA degradation. Current evidence suggests that Cas12a (formerly Cpf1), the only other known dual CRISPR-Cas nuclease and a component of Type V systems, also contains separate active sites for crRNA maturation and target degradation (Fonfara et al., 2016; Zetsche et al., 2016). Whether Cas12a enzymes can employ pre-crRNAs for target DNA recognition and cleavage, and whether Type V systems are consistent with our revised Type VI model is unknown.

The identification of two orthogonal subfamilies of RNA-guided RNA targeting enzymes could enable diverse multiplexed applications of Cas13a. Harnessing orthogonal Cas13a homologs with distinct crRNA specificities within the same application may enable RNA detection or *in vivo* imaging of distinct RNA species in

parallel, expanding the utility of Type VI-A systems. Notably, representatives of these enzyme subfamilies have not yet been found to co-exist within a single host genome. Only one strain of *Leptotrichia wadei* contains multiple Type VI-A CRISPR-Cas operons, all of which appear to be from one subfamily of Cas13a.

We were surprised to find that rates of ssRNA targeting by the Cas13a family varies by up to  $\sim 10^7$ -fold. Although the *in vivo* consequences of this diversity are unknown, we wonder whether Type VI-A CRISPR-Cas systems orchestrate RNA-triggered responses ranging from cell death to subtle modulation of transcript levels. The recent characterization of a Type VI-B CRISPR-Cas system that contains a Cas13b inhibitor within the *cas* operon provides tantalizing evidence that these systems may display much more dynamic behaviors within their respective host organisms (Smargon et al., 2017). Overall our work demonstrates that phylogenetic classification of CRISPR proteins may conceal functional diversity within biotechnologically relevant enzyme families.

# Chapter 5

Fitting Type VI systems into the  
CRISPR-Cas Toolbox

## 5.1 Discovery of Type VI systems

With the increased interest in CRISPR-Cas systems for biotechnological development, many groups have begun searching for new CRISPR systems. Classically CRISPR systems were identified by searching fully sequenced genomes for Cas1-like ORFs and then categorized by their operon structure (Makarova et al., 2011; 2015). Recently groups have taken other approaches to discover novel systems, focusing on identification of Class 2 single-effectors due to the ease of adapting these simpler systems (Burstein et al., 2017; Shmakov et al., 2015; 2017; Smargon et al., 2017). These new methodologies have utilized a variety of mechanisms to identify new Cas genes including genome-resolved metagenomics (Burstein et al., 2017) and searching for CRISPR arrays instead of Cas1 (Smargon et al., 2017). These searches expanded the CRISPR universe, discovering large ORFs (900+ amino acids) that contained two higher eukaryotes and prokaryotes nucleotide-binding motifs (HEPNS), which typically bind RNA substrates. These operons were deemed Type VI CRISPR-Cas systems and predicted to target ssRNA substrates (Shmakov et al., 2015).

## 5.2 Functional studies of Type VI systems

Two foundational papers studied Cas13a (formerly C2c2), the signature protein of Type VI-A systems, deciphering the basic mechanisms and functions of this CRISPR protein (Abudayyeh et al., 2016; East-Seletsky et al., 2016). Initial studies on LshCas13a determined this RNA-guided enzyme protects *E. coli* from MS2 infection, a lytic ssRNA phage, during heterologous expression (Abudayyeh et al., 2016). They further showed that LshCas13a binds and cleaves ssRNA targets. This cleavage activity is dependent on the HEPN domains and activity may be moderated by a protospacer flanking sequence (PFS). Further work presented in Chapter 3 of this dissertation established that the Cas13a family contains a second, distinct catalytic activity (East-Seletsky et al., 2016). This activity is harbored in the C-terminal region of the protein for LbuCas13a. We presented a revised model for Cas13a activity including the two endonuclease activities and the unification of the on-target cleavage and collateral cleavage proposed by Abudayyeh et al.. A crystal structure of LshCas13a validated this model revealing the composite active site of the two HEPN domains on the exterior face of the protein (Liu et al., 2017b). Our more recent work presented in Chapter 4 demonstrates that the Cas13a family harbors concealed diversity with distinct subgroups with incompatible crRNA and nucleotide preferences (East-Seletsky et al., 2017).

Other Type VI subtypes have been described, but beyond phylogenetic predictions only Type VI-B systems have been validated to target ssRNA (Smargon et al., 2017). The signature nuclease of Type VI-B systems, Cas13b, functions analogously to Cas13a, non-specifically cleaving ssRNA upon complementary target activation. Cas13b also tolerates additions within the stem loop to its guide RNA natively, due to repeat length diversity within the CRISPR arrays. This would suggest that like Cas13a, there is flexibility within the guide of the addition of useful tags like MS2 or boxB affinity tags for various applications. This has already been exploited by engineers in Type II systems, as Cas9 tolerates additions to the 5', 3' and internal loops

of the sgRNA with little functional impact (Koner mann et al., 2015; Shechner et al., 2015). Another intriguing feature of the Type VI-B systems is the incorporation of an auxiliary protein, Csx27 or Csx28 within the Cas operon. Overexpression of these proteins with Cas13b yielded either repression (Csx27) or enhancement (Csx28) of Cas13b's activity. Further work is required to understand mechanistically how these proteins modulate the activity Type VI-B systems and the phenotypic outcomes within their native hosts.

### 5.3 Fitting Type VI CRISPR-Cas systems into the prokaryotic defense arsenal

The non-specific nuclease activity of Cas13a is not a unique mechanism with the CRISPR universe and other bacterial immunity pathways. Toxin-Antitoxin systems and Restriction-Modification systems both contain nucleases that can non-specifically degrade nucleic acids upon activation (Gerdes et al., 2005; Ghafourian et al., 2014; Mruk and Kobayashi, 2014; Page and Peti, 2016; Vasu and Nagaraja, 2013). Within CRISPR systems, activation of Cas13a's non-specific degradation upon ternary complex formation is very analogous to Type III CRISPR mechanisms (Elmore et al., 2016; Estrella et al., 2016; Han et al., 2016; Kazlauskienė et al., 2016; Liu et al., 2017c). Even in Type I CRISPR-Cas systems, the non-specific nuclease-helicase Cas3 is loaded onto the target genome and then released for nonspecific degradation (Redding et al., 2015; Sinkunas et al., 2013; Westra et al., 2012). In addition, non-specific nucleases Csx1 and Csm6 still have undetermined roles in Type III CRISPR-Cas immune pathways (Niewoehner and Jinek, 2016; Sheppard et al., 2016). Physiological affects of Cas13a mediated immunity within native hosts have not been studied, yet it is expected that the non-specific degradation of ssRNA within the host will lead to cell death or dormancy.

The precise mechanisms to prevent auto-immunity within RNA-targeting systems still remain elusive across both Type III and Type VI systems. Unlike DNA-targeting mechanisms which all rely on PAM sequences to limit self-targeting, there is no unifying mode of auto-immunity prevention in Type III systems (Mohanraju et al., 2016; Tamulaitis et al., 2017). Most commonly, it is thought that base-pairing between the 5' crRNA handle and the 3' flank of the target inhibits activation of the ssDNA degradation activity. For bona-fide targets, the flank should not have any complementarity to the handle. In contrast, transcripts originating from the CRISPR locus would contain the spacer-repeat junction, therefore the target:guide complementary will extend beyond the spacer sequence into the crRNA handle (Elmore et al., 2016; Estrella et al., 2016; Han et al., 2016; Kazlauskienė et al., 2016; Liu et al., 2017c). Structural studies of Cmr and Csm complexes have supported these biochemical observations, but also suggest that a limited number of the bases are available for base-pairing (Staals et al., 2013; 2014; Taylor et al., 2015). The primarily reliance on nucleotides 5-7<sup>th</sup> of the crRNA handle, which are immediately adjacent to the spacer, was validated by biochemical studies (Kazlauskienė et al., 2016).

Other alternative modes of auto-immunity prevention are also observed in some Type III systems. For example the Cmr complex from *Pyrococcus furiosus* (PfCmr) requires the presence of an RNA PAM (rPAM) sequence (Elmore et al., 2016). Conflicting results from the ability of target substrates lacking a 3' flanking region to

activate DNase activity of Cas10 do not coherently fit with the proposed model (Elmore et al., 2016; Estrella et al., 2016). Additionally, a divergent targeting mode is observed in Type III-B Cmr complexes from *Sulfolobus solfataricus* where at high target RNA concentrations the crRNA will mediate a sequence-dependent, instead of measured, cleavage events (Zhang et al., 2012; 2016). It remains an open question if these alternatives will be compiled into a universal mechanism, or if divergent complexes maintain distinct mechanisms to reduce auto-immunity and mediate cleavage.

The mechanism for Type VI autoimmunity has not been elucidated. For some homologs, the presence of a protospacer flanking sequence (PFS) motif limits the targeting activation of a given spacer (Abudayyeh et al., 2016; Smargon et al., 2017). The molecular basis for this requirement is unknown, but analysis of the crRNA sequence and PFS requirement for currently PFS characterized homologs leads to the observation of the potential for inhibitor sequences to base pair with each other. It is possible that this inhibitory mechanism is similar to the previously described 3' flanking sequence dependence of Type III CRISPR systems activation. If this is a bone fide mechanism to limit auto-immunity for Type VI systems remains to be seen.

The general model of the CRISPR pathway requires precise crRNA processing for targeting, suggesting that Type VI may represent a novel paradigm for CRISPR mechanisms. Alternative mechanisms of generating mature crRNAs have been observed (Zhang et al., 2013), but in the absence of Cas6 (Type I) or RNase III (Type II) these CRISPR complexes are not competent for phage defense (Brouns et al., 2008; Deltcheva et al., 2011; Maier et al., 2015; Semenova et al., 2015). In contrast, our *in vitro* work on Type VI systems suggests that pre-crRNAs, even within the context of a CRISPR array, can direct targeting. The loss of processing only impacts targeting efficiency of crRNAs located beyond the first few repeat units of the array. It is unclear how well this paradigm translates to other single-effector CRISPR processing systems. Little work has been done to characterize the crRNA requirements for Cas12a (formerly Cpf1), the other single-effector CRISPR protein that can self-process its own crRNA. The current evidence suggests that the dual nuclease activities of Cas13a are also separable (Fonfara et al., 2016; Zetsche et al., 2016)

#### **5.4 Open Questions in Type VI CRISPR-Cas systems**

Many open questions remain within Type VI CRISPR systems. Very few Type VI systems have been thoroughly studied, including no Type VI-C proteins, and only two homologs of the Type VI-B effector complex. The precise mechanism by which the two HEPN domains act in concert to perform target cleavage is unknown. The existing crystal structure of LshCas13a reveals the active sites on the external face of the protein, but they are not close enough to form the expected composite active site. Precise mapping of the activator requirements for binding and ssRNA-targeting will further reveal the mechanism of non-specific nuclease activation. How this activation is modulated by other host proteins including Csx27 and Csx28 remains a mystery. Its possible these small proteins are binding to Cas13 to modulated the propensity for nuclease activation or they might be affecting the off-rate of target binding.

Beyond the interference mechanism, how these RNA-targeting systems adaptive to new invader is unclear. A majority of Type VI operons lack the essential Cas1-Cas2

machinery associated with new spacer acquisition (Shmakov et al., 2017). How these systems select for transcribed spacer sequences and the possible requirements for reverse transcription are dissonant with the established mechanisms of DNA targeting acquisition (Sternberg et al., 2016). An alternative possibility is that these systems rely on Cas1 and Cas2 from co-occurring CRISPR systems to expand their arrays.

During our initial study, we provided proof-of-concept that these systems could be applied for RNA detection. While it remains to be seen if Cas13 proteins will be as applicable as Cas9 for a wide range of RNA binding functions, further protein engineering could develop a minimal chassis for a programmable RNA binding protein. Removal of the HEPN domains, or harnessing the conformational change may enable further applications of this protein. Time will tell if Cas13 proteins will revolutionize RNA biology to the same extent as Cas9 for genome editing, but for now, the novel biology of these systems will keep many people busy for many more years.

# Appendix 1

## RNA-programmed genome editing in human cells

A portion of the content presented in this chapter has been previously published as part of the following research article: Martin Jinek, Alexandra East, Aaron Cheng, Steven Lin, Enbo Ma, and Jennifer Doudna, (2013). RNA-programmed genome editing in human cells. *eLife* 2.

Martin Jinek, Jennifer Doudna, Aaron Cheng, and Alexandra Seletsky conceived the study and designed experiments. Martin Jinek, and Alexandra Seletsky, executed experimental work with assistance from Enbo Ma for northern blots, and Steven Lin for sequencing analysis. All authors discussed the data and wrote the manuscript.



## I.1 Summary

Type II CRISPR immune systems in bacteria use a dual RNA-guided DNA endonuclease, Cas9, to cleave foreign DNA at specific sites. We show here that Cas9 assembles with hybrid guide RNAs in human cells and can induce the formation of double-strand DNA breaks (DSBs) at a site complementary to the guide RNA sequence in genomic DNA. This cleavage activity requires both Cas9 and the complementary binding of the guide RNA. Experiments using extracts from transfected cells show that RNA expression and/or assembly into Cas9 is the limiting factor for Cas9-mediated DNA cleavage. In addition, we find that extension of the RNA sequence at the 3' end enhances DNA targeting activity *in vivo*. These results show that RNA-programmed genome editing is a facile strategy for introducing site-specific genetic changes in human cells.

## I.2 Introduction

Methods for introducing site-specific double-strand DNA (dsDNA) breaks (DSBs) in genomic DNA have transformed the ability to engineer eukaryotic organisms by initiating DNA repair pathways that lead to targeted genetic re-programming. Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have proved effective for such genomic manipulation but their use has been limited by the need to engineer a specific protein for each dsDNA target site and by off-target activity (Bogdanove and Voytas, 2011; Urnov et al., 2010). Thus, alternative strategies for triggering site-specific DNA cleavage in eukaryotic cells are of great interest.

Research into genome defense mechanisms in bacteria showed that CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) loci encode RNA-guided adaptive immune systems that can destroy foreign DNA (Bhaya et al., 2011; Terns and Terns, 2011; Wiedenheft et al., 2012). The Type II CRISPR/Cas systems require a single protein, Cas9, to catalyze DNA cleavage (Sapranauskas et al., 2011). Cas9 generates blunt DSBs at sites defined by a 20-nucleotide guide sequence contained within an associated CRISPR RNA (crRNA) transcript (Gasiunas et al., 2012; Jinek et al., 2012). Cas9 requires both the guide crRNA and a trans-activating crRNA (tracrRNA) that is partially complementary to the crRNA for site-specific DNA recognition and cleavage (Deltcheva et al., 2011; Jinek et al., 2012). Recent experiments showed that the crRNA:tracrRNA complex can be redesigned as a single transcript (single-guide RNA or sgRNA) encompassing the features required for both Cas9 binding and DNA target site recognition (Jinek et al., 2012). Using sgRNA, Cas9 can be programmed to cleave double-stranded DNA at any site defined by the guide RNA sequence and including a GG protospacer-adjacent (PAM) motif (Jinek et al., 2012; Sapranauskas et al., 2011). These findings suggested the exciting possibility that Cas9:sgRNA complexes might constitute a simple and versatile RNA-directed system for generating DSBs that could facilitate site-specific genome editing. However, it was not known whether such a bacterial system would function in eukaryotic cells.

We show here that Cas9 can be expressed and localized to the nucleus of human cells, and that it assembles with sgRNA *in vivo*. These complexes can generate double stranded breaks and stimulate non-homologous end joining (NHEJ) repair in

genomic DNA at a site complementary to the sgRNA sequence, an activity that requires both Cas9 and the sgRNA. Extension of the RNA sequence at its 3' end enhances DNA targeting activity *in vivo*. Further, experiments using extracts from transfected cells show that sgRNA assembly into Cas9 is the limiting factor for Cas9-mediated DNA cleavage. These results demonstrate the feasibility of RNA-programmed genome editing in human cells.

### **I.3 Methods**

#### **I.3.1 Plasmid design and construction**

The sequence encoding *Streptococcus pyogenes* Cas9 (residues 1-1368) fused to an HA epitope (amino acid sequence DAYPYDVPDYASL), a nuclear localization signal (amino acid sequence PKKKRKVEDPKKKRKVD) was codon optimized for human expression and synthesized by GeneArt. Ligation-independent cloning (LIC) was used to insert this sequence into a pcDNA3.1-derived GFP and mCherry LIC vectors (vectors 6D and 6B, respectively, obtained from the UC Berkeley MacroLab), resulting in a Cas9-HA-NLS-GFP and Cas9-HA-NLS-mCherry fusions expressed under the control of the CMV promoter. Guide sgRNAs were expressed using expression vector pSilencer 2.1-U6 puro (Life Technologies) and pSuper (Oligoengine). RNA expression constructs were generated by annealing complementary oligonucleotides to form the RNA-coding DNA sequence and ligating the annealed DNA fragment between the BamHI and HindIII sites in pSilencer 2.1-U6 puro and BglII and HindIII sites in pSuper.

#### **I.3.2 Cell culture conditions and DNA transfections**

HEK293T cells were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) in a 37°C humidified incubator with 5% CO<sub>2</sub>. Cells were transiently transfected with plasmid DNA using either X-tremeGENE DNA Transfection Reagent (Roche) or Turbofect Transfection Reagent (Thermo Scientific) with recommended protocols. Briefly, HEK293T cells were transfected at 60-80% confluency in 6-well plates using 0.5 µg of the Cas9 expression plasmid and 2.0 µg of the RNA expression plasmid. The transfection efficiencies were estimated to be 30-50% for Turbofect (Figs. 1E and 2A-B) and 80-90% for X-tremegene (Fig. 3B), based on the fraction of GFP-positive cells observed by fluorescence microscopy. 48 hours post transfection, cells were washed with phosphate buffered saline (PBS) and lysed by applying 250 µl lysis buffer (20 mM Hepes pH 7.5, 100 mM potassium chloride (KCl), 5 mM magnesium chloride (MgCl<sub>2</sub>), 1 mM dithiothreitol (DTT), 5% glycerol, 0.1% Triton X-100, supplemented with Roche Protease Inhibitor cocktail) and then rocked for 10 min at 4°C. The resulting cell lysate was divided into aliquots for further analysis. Genomic DNA was isolated from 200 µl cell lysate using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol.

#### **I.3.3 Western blot analysis of Cas9 expression**

HEK293T, transfected with the Cas9-HA-NLS-GFP expression plasmid, were harvested and lysed 48 hours post transfection as above. 5 ul of lysate were electrophoresed on a 10% SDS polyacrylamide gel, blotted onto a PVDF membrane and probed with HRP-conjugated anti-HA antibody (Sigma, 1:1000 dilution in 1x PBS).

### **I.3.4 Surveyor assay**

The Surveyor assay was performed as previously described (Doyon et al., 2011) (Miller et al., 2007; Qiu et al., 2004). Briefly, the human clathrin light chain A (CLTA) locus was PCR amplified from 200 ng of genomic DNA using a high fidelity polymerase, HerculaSe II Fusion DNA Polymerase (Agilent Technologies) and forward primer 5'-GCAGCAGAAGAAGCCTTTGT-3' and reverse primer 5'-TTCCTCCTCTCCCTCCTCTC-3'. 300 ng of the 360 bp amplicon was then denatured by heating to 95°C and slowly reannealed using a heat block to randomly rehybridize wild type and mutant DNA strands. Samples were then incubated with Cel-1 nuclease (Surveyor Kit, Transgenomic) for 1 hour at 42°C. Cel-1 recognizes and cleaves DNA helices containing mismatches (wild type:mutant hybridization). Cel-1 nuclease digestion products were separated on a 10% acrylamide gel and visualized by staining with SYBR Safe (Life Technologies). Quantification of cleavage bands was performed using ImageLab software (Bio-Rad). The percent cleavage was determined by dividing the average intensity of cleavage products (160-200 bps) by the sum of the intensities of the uncleaved PCR product (360 bp) and the cleavage product.

### **I.3.5 *In vitro* transcription**

Guide RNA was *in vitro* transcribed using recombinant T7 RNA polymerase and a DNA template generated by annealing complementary synthetic oligonucleotides as previously described (Sternberg et al., 2012). RNAs were purified by electrophoresis on 7M urea denaturing acrylamide gel, ethanol precipitated, and dissolved in DEPC-treated water.

### **I.3.6 Northern blot analysis**

RNA was purified from HEK293T cells using the mirVana small-RNA isolation kit (Ambion). For each sample, 800 ng of RNA were separated on a 10% urea-PAGE gel after denaturation for 10 min at 70°C in RNA loading buffer (0.5X TBE (pH7.5), 0.5 mg/ml bromophenol blue, 0.5 mg xylene cyanol and 47% formamide). After electrophoresis at 10W in 0.5X TBE buffer until the bromophenol blue dye reached the bottom of the gel, samples were electroblotted onto a Nytran membrane at 20 volts for 1.5 hours in 0.5X TBE. The transferred RNAs were cross-linked onto the Nytran membrane in UV-Crosslinker (Stratagene) and were pre-hybridized at 45°C for 3 hours in a buffer containing 40% formamide, 5X SSC, 3X Denhardt's (0.1% each of ficoll, polyvinylpyrrolidone, and BSA) and 200 µg/ml Salmon sperm DNA. The pre-hybridized membranes were incubated overnight in the prehybridization buffer supplemented with 5'-<sup>32</sup>P-labeled antisense DNA oligo probe at 1 million cpm/ml. After several washes in SSC buffer (final wash in 0.2X SCC), the membranes were imaged phosphorimaging.

### **I.3.7 *In vitro* cleavage assay**

Cell lysates were prepared as described above and incubated with CLTA-RFP donor plasmid (Doyon et al., 2011). Cleavage reactions were carried out in a total volume of 20 µl and contained 10 µl lysate, 2 µl of 5x cleavage buffer (100 mM HEPES pH 7.5, 500 mM KCl, 25 mM MgCl<sub>2</sub>, 5 mM DTT, 25% glycerol) and 300 ng plasmid. Where indicated, reactions were supplemented with 10 pmol of *in vitro* transcribed CLTA1

sgRNA. Reactions were incubated at 37°C for one hour and subsequently digested with 10 U of XhoI (NEB) for an additional 30 min at 37°C. The reactions were stopped by the addition of Proteinase K (Thermo Scientific) and incubated at 37°C for 15 min. Cleavage products were analyzed by electrophoresis on a 1% agarose gel and stained with SYBR Safe. The presence of ~2230 and ~3100 bp fragments is indicative of Cas9-mediated cleavage.

#### **I.4 Results**

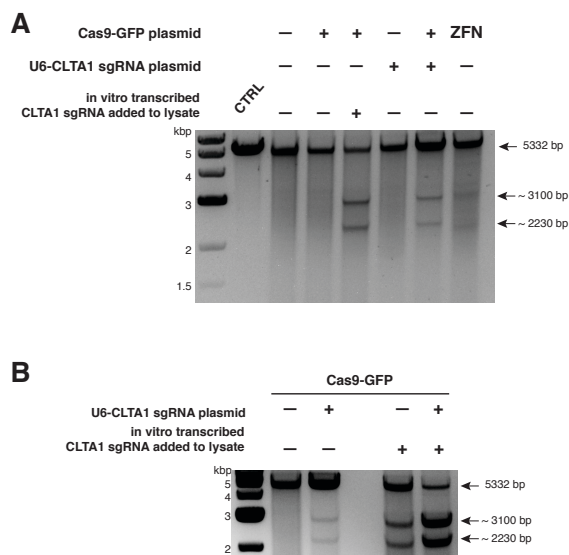
To test whether Cas9 could be programmed to cleave genomic DNA *in vivo*, we co-expressed Cas9 together with an sgRNA designed to target the human clathrin light chain (CLTA) gene. The CLTA genomic locus has previously been targeted and edited using ZFNs (Doyon et al., 2011). We first tested the expression of a human-codon-optimized version of the *Streptococcus pyogenes* Cas9 protein and sgRNA in human HEK293T cells. The 160 kDa Cas9 protein was expressed as a fusion protein bearing an HA epitope, a nuclear localization signal (NLS), and green fluorescent protein (GFP) attached to the C-terminus of Cas9 (Fig. I.1). Analysis of cells transfected with a vector encoding the GFP-fused Cas9 revealed abundant Cas9 expression and nuclear localization (Fig. 1B). Western blotting confirmed that the Cas9 protein is expressed largely intact in extracts from these cells (Fig. I.1A). To program Cas9, we expressed sgRNA bearing a 5'-terminal 20-nucleotide sequence complementary to the target DNA sequence, and a 42-nucleotide 3'-terminal stem loop structure required for Cas9 binding (Fig. I.1C). This 3'-terminal sequence corresponds to the minimal stem-loop structure that has previously been used to program Cas9 *in vitro* (Jinek et al., 2012). The expression of this sgRNA was driven by the human U6 (RNA polymerase III) promoter. Northern blotting analysis of RNA extracted from cells transfected with the U6 promoter-driven sgRNA plasmid expression vector showed that the sgRNA is indeed expressed, and that their stability is enhanced by the presence of Cas9 (Fig. I.1D).



unwinding of the target DNA and formation of a duplex between the guide segment of the sgRNA and the target DNA. This is dependent on the presence of a GG dinucleotide PAM motif downstream of the target sequence in the target DNA. Note that the target sequence is inverted relative to the upper diagram. (D) Northern blot analysis of sgRNA expression in HEK293T cells. (E) Surveyor nuclease assay of genomic DNA isolated from HEK293T cells expressing Cas9 and/or CLTA sgRNA. A ZFN construct previously used to target the CLTA locus (Doyon et al., 2011) was used as a positive control for detecting DSB-induced DNA repair by non-homologous end joining.

Next we investigated whether site-specific DSBs are generated in HEK293T cells transfected with Cas9-HA-NLS-mCherry and the CLTA1 sgRNA. To do this, we probed for minor insertions and deletions in the locus resulting from imperfect repair by DSB-induced NHEJ using the Surveyor nuclease assay (Qiu et al., 2004). The region of genomic DNA targeted by Cas9:sgRNA is amplified by PCR and the resulting products are denatured and reannealed. The rehybridized PCR products are incubated with the mismatch recognition endonuclease Cel-1 and resolved on an acrylamide gel to identify Cel-1 cleavage bands. As DNA repair by NHEJ is typically induced by a DSB, a positive signal in the Surveyor assay indicates that genomic DNA cleavage has occurred. Using this assay, we detected cleavage of the CLTA locus at a position targeted by the CLTA1 sgRNA (Fig. 1.1E). A pair of ZFNs that target a neighboring site in the CLTA locus provided a positive control in these experiments (Doyon et al., 2011).

To determine if either Cas9 or sgRNA expression is a limiting factor in the observed genome editing reactions, lysates prepared from the transfected cells were incubated with plasmid DNA harboring a fragment of the CLTA gene targeted by the CLTA1 sgRNA. Plasmid DNA cleavage was not observed upon incubation with lysate prepared from cells transfected with the Cas9-HA-NLS-GFP expression vector alone, consistent with the Surveyor assay results. However, robust plasmid cleavage was detected when the lysate was supplemented with *in vitro* transcribed CLTA1 sgRNA (Fig. 1.2A). Furthermore, lysate prepared from cells transfected with both Cas9 and sgRNA expression vectors supported plasmid cleavage, while lysates from cells transfected with the sgRNA-encoding vector alone did not (Fig. 1.2A). These results suggest that a limiting factor for Cas9 function in human cells could be assembly with the sgRNA. We tested this possibility directly by analyzing plasmid cleavage in lysates from cells transfected as before in the presence and absence of added exogenous sgRNA. Notably, when exogenous sgRNA was added to lysate from cells transfected with both the Cas9 and sgRNA expression vectors, a substantial increase in DNA cleavage activity was observed (Fig. 1.2B). This result indicates that the limiting factor for Cas9 function in HEK293T cells is the expression of the sgRNA or its loading into Cas9.



### Figure I.2 Cell lysates contain active Cas9:sgRNA and support site-specific DNA cleavage

(A) Lysates from cells transfected with the plasmid(s) indicated at left were incubated with plasmid DNA containing a PAM and the target sequence complementary to the CLTA1 sgRNA; where indicated, the reaction was supplemented with 10 pmol of *in vitro* transcribed CLTA1 sgRNA; secondary cleavage with XhoI generated fragments of ~2230 and ~3100 bp fragments indicative of Cas9-mediated cleavage. A control reaction using lysate from cells transfected with a ZFN expression construct shows fragments of slightly different size reflecting the offset of the ZFN target site relative to the CLTA1 target site. (B) Lysates from cells transfected with Cas9-GFP expression plasmid and, where indicated, the CLTA1 sgRNA expression plasmid, were incubated with target plasmid DNA as in (A) in the absence or presence of *in vitro*-transcribed CLTA1 sgRNA.

As a means of enhancing the Cas9:sgRNA assembly *in vivo*, we next tested the effect of extending the presumed Cas9-binding region of the guide RNA. Two new versions of the CLTA1 sgRNA were designed to include an additional six or twelve base pairs in the helix that mimics the base-pairing interactions between the crRNA and tracrRNA (Fig. I.3A). Additionally, the 3'-end of the guide RNA was extended by five nucleotides based on the native sequence of the *S. pyogenes* tracrRNA (Deltcheva et al., 2011). Vectors encoding these 3' extended sgRNAs under the control of either the U6 or H1 Pol III promoters were transfected into cells along with the Cas9-HA-NLS-GFP expression vector and site-specific genome cleavage was tested using the Surveyor assay (Fig. I.3B). The results confirmed that cleavage required both Cas9 and the CLTA1 sgRNA, but did not occur when either Cas9 or the sgRNA were expressed alone. Furthermore, we observed substantially increased frequencies of NHEJ, as detected by Cel-1 nuclease cleavage, while the frequency of NHEJ mutagenesis obtained with the control ZFN pair was largely unchanged. These results suggest that the 3'-extended





and TALENs (Bogdanove and Voytas, 2011). Our data suggest that sgRNA expression and/or its assembly into Cas9, rather than Cas9 expression, localization or folding, presently limits Cas9 function in human cells. Higher efficiencies of Cas9-mediated genome targeting could be achieved by optimization of the sgRNA construct design, its expression levels or its subcellular localization. We note that the sgRNAs used in this study are not thought to be 5'-capped or 3'-polyadenylated, which may have reduced their stability *in vivo*. This and other 5' and 3' end modifications might provide alternative approaches to enhancing Cas9:sgRNA assembly and activity in cells. Nonetheless, the levels of targeting observed in this study have been obtained with a minimal system that relies on simple base pairing to a guide RNA, in contrast to the ZFN and TALEN proteins, which require a new protein to be engineered for each new cleavage site. RNA-guided genome editing would thus offer distinct advantages due to the simplicity of the sgRNA design.

Our results thus provide the framework for implementing Cas9 as a facile molecular tool for diverse genome editing applications. Although not tested explicitly in this study, a powerful feature of this system is the potential to program Cas9 with multiple sgRNAs in the same cell, either to increase the efficiency of targeting at a single locus, or as a means of targeting several loci simultaneously. Such strategies would find broad application in genome-wide experiments and large-scale research efforts such as the development of multigenic disease models. As an inexpensive and rapid mechanism for triggering site-specific genome modification, the programmable Cas9:sgRNA system could potentially transform next-generation genome-scale studies.

## REFERENCES

- Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., 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–aaf5573.
- Agari, Y., Sakamoto, K., Tamakoshi, M., Oshima, T., Kuramitsu, S., and Shinkai, A. (2010). Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *Journal of Molecular Biology* 395, 270–281.
- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* 14, 67–76.
- 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. *Biology Direct* 8, 1–1.
- Arber, W. (1965). Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* 19, 365–378.
- Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, Ü. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Research* 42, 7884–7893.
- Auyeung, V.C., Ulitsky, I., McGeary, S.E., and Bartel, D.P. (2013). Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing. *Cell* 152, 844–858.
- Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., et al. (2011). A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Molecular Microbiology* 79, 484–502.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* 315, 1709–1712.
- Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., et al. (2008). A Novel Family of Sequence-specific Endoribonucleases Associated with the Clustered Regularly Interspaced Short Palindromic Repeats. *Journal of Biological Chemistry* 283, 20361–20371.
- Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* 45, 273–297.
- Blesa, A., César, C.E., Averhoff, B., and Berenguer, J. (2015). Noncanonical cell-to-cell DNA transfer in *Thermus* spp. is insensitive to argonaute-mediated interference. *Journal of Bacteriology* 197, 138–146.
- Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S.J.J., and Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Molecular Cell* 58, 60–70.

- Bogdanove, A.J., and Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. *Science* 333, 1843–1846.
- Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* 321, 960–964.
- Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A., and Banfield, J.F. (2017). New CRISPR-Cas systems from uncultivated microbes. *Nature* 542, 237–241.
- Burstein, D., Sun, C.L., Brown, C.T., Sharon, I., Anantharaman, K., Probst, A.J., Thomas, B.C., and Banfield, J.F. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications* 7, 10613.
- Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O'Toole, G.A. (2012). The CRISPR/Cas Adaptive Immune System of *Pseudomonas aeruginosa* Mediates Resistance to Naturally Occurring and Engineered Phages. *Journal of Bacteriology* 194, 5728–5738.
- Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development* 22, 3489–3496.
- Chakraborty, S., Mehtab, S., Patwardhan, A., and Krishnan, Y. (2012). Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing. *Rna* 18, 1014–1028.
- Charpentier, E., Richter, H., van der Oost, J., and White, M.F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews* 39, 428–441.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491.
- Chevallet, M., Luche, S., and Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. *Nat Protoc* 1, 1852–1858.
- Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L., and Brüssow, H. (2004). Phage-host interaction: an ecological perspective. *Journal of Bacteriology* 186, 3677–3686.
- Choi, U.Y., Kang, J.-S., Hwang, Y.S., and Kim, Y.-J. (2015). Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases. *Exp. Mol. Med.* 47, e144.
- Chomczynski, P. (1992). One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Anal. Biochem.* 201, 134–139.
- Chou, H.-H. (2010). Shared probe design and existing microarray reanalysis using PICKY. *BMC Bioinformatics* 11, 196.
- Choudhury, N.R., de Lima Alves, F., de Andrés-Aguayo, L., Graf, T., Cáceres, J.F., Rappsilber, J., and Michlewski, G. (2013). Tissue-specific control of brain-enriched miR-7 biogenesis. *Genes & Development* 27, 24–38.
- Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Molecular Cell* 44, 667–678.

- Cociorva, D., L Tabb, D., and Yates, J.R. (2007). Validation of tandem mass spectrometry database search results using DTASelect. *Curr Protoc Bioinformatics Chapter 13*, Unit13.4.
- Cordray, M.S., and Richards-Kortum, R.R. (2012). Emerging nucleic acid-based tests for point-of-care detection of malaria. *Am. J. Trop. Med. Hyg.* 87, 223–230.
- Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications* 3, 945–947.
- Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607.
- Deng, L., Kenchappa, C.S., Peng, X., She, Q., and Garrett, R.A. (2012). Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in *Sulfolobus*. *Nucleic Acids Research* 40, 2470–2480.
- Deng, L., Garrett, R.A., Shah, S.A., Peng, X., and She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular Microbiology* 87, 1088–1099.
- Díez-Villaseñor, C., Guzmán, N.M., Almendros, C., García-Martínez, J., and Mojica, F.J.M. (2013). CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Bio* 10, 0–10.
- Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* 532, 522–526.
- Doyon, J.B., Zeitler, B., Cheng, J., Cheng, A.T., Cherone, J.M., Santiago, Y., Lee, A.H., Vo, T.D., Doyon, Y., Miller, J.C., et al. (2011). Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nature Publishing Group* 13, 331–337.
- Dy, R.L., Richter, C., Salmond, G.P.C., and Fineran, P.C. (2014). Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu Rev Virol* 1, 307–331.
- East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* 538, 270–273.
- Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., and Terns, M.P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes & Development* 30, 447–459.
- Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973–1237973.
- Erdmann, S., and Garrett, R.A. (2012). Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Molecular Microbiology* 85, 1044–1056.
- Erdmann, S., Le Moine Bauer, S., and Garrett, R.A. (2013a). Inter-viral conflicts that exploit host CRISPR immune systems of *Sulfolobus*. *Molecular Microbiology* n/a–

n/a.

- Erdmann, S., Shah, S.A., and Garrett, R.A. (2013b). SMV1 virus-induced CRISPR spacer acquisition from the conjugative plasmid pMGB1 in *Sulfolobus solfataricus* P2. *Biochem. Soc. Trans.* *41*, 1449–1458.
- Estrella, M.A., Kuo, F.-T., and Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes & Development* 1–12.
- Filipovska, A., and Rackham, O. (2011). Designer RNA-binding proteins: New tools for manipulating the transcriptome. *RNA Bio* *8*, 978–983.
- Fineran, P.C., Gerritzen, M.J.H., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A.F.T., Staals, R.H.J., and Brouns, S.J.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A*.
- Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems Acquisition of new information. *Virology* *434*, 202–209.
- Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., and Salmond, G.P.C. (2009). The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc Natl Acad Sci U S A* *106*, 894–899.
- Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lecrivain, A.L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2013). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research* *42*, 2577–2590.
- Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* *532*, 517–521.
- Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Research* *26*, 901–913.
- Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* *468*, 67–71.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* –.
- Gerdes, K., Christensen, S.K., and Løbner-Olesen, A. (2005). Prokaryotic toxin-antitoxin stress response loci. *Nat. Rev. Microbiol.* *3*, 371–382.
- Ghafourian, S., Raftari, M., Sadeghifard, N., and Sekawi, Z. (2014). Toxin-antitoxin Systems: Classification, Biological Function and Application in Biotechnology. *Curr Issues Mol Biol* *16*, 9–14.
- Guil, S., and Cáceres, J.F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol* *14*, 591–596.
- Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* *15*, 509–524.
- Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V.C., III, Graveley, B.R., Terns, R.M., et al. (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Molecular Cell* *45*, 292–302.

- 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.
- Hale, C., Kleppe, K., Terns, R.M., and Terns, M.P. (2008). Prokaryotic silencing ( $\psi$ )RNAs in *Pyrococcus furiosus*. *Rna* **14**, 2572–2579.
- Han, C., Liu, Y., Wan, G., Choi, H.J., Zhao, L., Ivan, C., He, X., Sood, A.K., Zhang, X., and Lu, X. (2014). The RNA-Binding Protein DDX1 Promotes Primary MicroRNA Maturation and Inhibits Ovarian Tumor Progression. *CellReports* **8**, 1447–1460.
- Han, W., Li, Y., Deng, L., Feng, M., Peng, W., Hallstrøm, S., Zhang, J., Peng, N., Liang, Y.X., White, M.F., et al. (2016). A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. *Nucleic Acids Research* gkw1274–11.
- Harms, A., Maisonneuve, E., and Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* **354**, aaf4268–11.
- Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of Bacteriology* **196**, 310–317.
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* **329**, 1355–1358.
- Haurwitz, R.E., Sternberg, S.H., and Doudna, J.A. (2012). Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *Embo J.* **31**, 2824–2832.
- Hazan, R., and Engelberg-Kulka, H. (2004). *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol. Genet. Genomics* **272**, 227–234.
- Heler, R., Marraffini, L.A., and Bikard, D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular Microbiology* **93**, 1–9.
- Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202.
- Heler, R., Wright, A.V., Vucelja, M., Bikard, D., Doudna, J.A., and Marraffini, L.A. (2016). Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. *Molecular Cell* 1–9.
- Hochstrasser, M.L., and Doudna, J.A. (2015). Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences* **40**, 58–66.
- Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A* **111**, 6618–6623.
- Hochstrasser, M.L., Taylor, D.W., Kornfeld, J.E., Nogales, E., and Doudna, J.A. (2016). DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Molecular Cell* **63**, 840–851.
- Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008). Diversity, Activity, and Evolution of CRISPR Loci in *Streptococcus thermophilus*. *Journal of Bacteriology* **190**, 1401–1412.

- Hsia, K.-C., Chak, K.-F., Liang, P.-H., Cheng, Y.-S., Ku, W.-Y., and Yuan, H.S. (2004). DNA binding and degradation by the HNH protein ColE7. *Structure/Folding and Design* *12*, 205–214.
- Høyland-Kroghsbo, N.M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., and Bassler, B.L. (2017). Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A* *114*, 131–135.
- Jackson, R.N., and Wiedenheft, B. (2015). A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Molecular Cell* *58*, 722–728.
- Jackson, R.N., Golden, S.M., van Erp, P.B.G., Carter, J., Westra, E.R., Brouns, S.J.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* *345*, 1473–1479.
- 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.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* *337*, 816–821.
- Ka, D., Lee, H., Jung, Y.-D., Kim, K., Seok, C., Suh, N., and Bae, E. (2016). Crystal Structure of *Streptococcus pyogenes* Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure* *24*, 70–79.
- Katoh, K., and Standley, D.M. (2013a). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* *30*, 772–780.
- Katoh, K., and Standley, D.M. (2013b). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* *30*, 772–780.
- Kaya, E., Doxzen, K.W., Knoll, K.R., Wilson, R.C., Strutt, S.C., Kranzusch, P.J., and Doudna, J.A. (2016). A bacterial Argonaute with noncanonical guide RNA specificity. *Proc Natl Acad Sci U S A* *113*, 4057–4062.
- Kazlauskiene, M., Tamulaitis, G., Kostiuik, G., Venclovas, Č., and Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell* *62*, 295–306.
- Kim, D., and Rossi, J. (2008). RNAi mechanisms and applications. *44*, 613–616.
- Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.-M., Choi, J.W., Woo, E., Koh, H.C., Nam, J.-W., et al. (2017). In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Meth* *14*, 153–159.
- Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., and Joung, J.K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* *34*, 869–874.
- 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.

- Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8, R61.
- Künne, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I.M., Mielliet, W.R., Klein, M., Depken, M., Suarez-Diez, M., and Brouns, S.J.J. (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular Cell* 63, 852–864.
- Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Meth* 9, 357–359.
- Lee, H.Y., Haurwitz, R.E., Apffel, A., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L., and Doudna, J.A. (2013). RNA-protein analysis using a conditional CRISPR nuclease. *Proc Natl Acad Sci U S A* 110, 5416–5421.
- Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drèze, P., and Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic Acids Research* 39, 5513–5525.
- Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U., and Sorek, R. (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* 520, 505–510.
- Li, H. (2015). Structural Principles of CRISPR RNA Processing. *Structure* 23, 13–20.
- Li, M., Wang, R., Zhao, D., and Xiang, H. (2014). Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Research* 42, 2483–2492.
- Lillestøl, R.K., Shah, S.A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R.A. (2009). CRISPR families of the crenarchaeal genus *Sulfolobus*: bidirectional transcription and dynamic properties. *Molecular Microbiology* 72, 259–272.
- Lin, S., Staahl, B., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3.
- Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., and Wang, Y. (2017a). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. *Molecular Cell* 65, 310–322.
- Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., Li, J., Sheng, G., and Wang, Y. (2017b). Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* 168, 121–134.e12.
- Liu, T.Y., Iavarone, A.T., and Doudna, J.A. (2017c). RNA and DNA Targeting by a Reconstituted *Thermus thermophilus* Type III-A CRISPR-Cas System. *PLoS ONE* 12, e0170552.
- Mackay, J.P., Font, J., and Segal, D.J. (2011). The prospects for designer single-stranded RNA-binding proteins. *Nature Publishing Group* 18, 256–261.
- Maier, L.-K., Stachler, A.-E., Saunders, S.J., Backofen, R., and Marchfelder, A. (2015). An active immune defense with a minimal CRISPR (clustered regularly interspaced short palindromic repeats) RNA and without the Cas6 protein. *Journal of Biological Chemistry* 290, 4192–4201.
- Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and



- classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736.
- Makarova, K.S., Wolf, Y.I., van der Oost, J., and Koonin, E.V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology Direct* **4**, 29.
- Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. *Nat Meth* **10**, 957–963.
- Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science* **322**, 1843–1845.
- Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571.
- McDonald, W.H., Tabb, D.L., Sadygov, R.G., MacCoss, M.J., Venable, J., Graumann, J., Johnson, J.R., Cociorva, D., and Yates, J.R. (2004). MS1, MS2, and SQT-three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun. Mass Spectrom.* **18**, 2162–2168.
- McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236.
- Mcllwain, D.R., Mcllwain, D.R., Berger, T., Berger, T., Mak, T.W., and Mak, T.W. (2013). Caspase functions in cell death and disease. *Cold Spring Harb Perspect Biol* **5**, a008656.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research* **41**, W597–W600.
- Medina-Aparicio, L., Rebollar-Flores, J.E., Gallego-Hernandez, A.L., Vazquez, A., Olvera, L., Gutierrez-Rios, R.M., Calva, E., and Hernandez-Lucas, I. (2011). The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in *Salmonella enterica* serovar Typhi. *Journal of Bacteriology* **193**, 2396–2407.
- Miller, J.C., Holmes, M.C., Wang, J., Guschin, D.Y., Lee, Y.-L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., et al. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778–785.
- Moch, C., Fromant, M., Blanquet, S., and Plateau, P. (2016). DNA binding specificities of *Escherichia coli* Cas1-Cas2 integrase drive its recruitment at the CRISPR locus. *Nucleic Acids Research*.
- 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.
- Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740.

- Mruk, I., and Kobayashi, I. (2014). To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Research* 42, 70–86.
- Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345, 1479–1484.
- Nam, K.H., Ding, F., Haitjema, C., Huang, Q., DeLisa, M.P., and Ke, A. (2012a). Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *Journal of Biological Chemistry* 287, 35943–35952.
- Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P., and Ke, A. (2012b). Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *20*, 1574–1584.
- Nickel, L., Weidenbach, K., Jäger, D., Backofen, R., Lange, S.J., Heidrich, N., and Schmitz, R.A. (2013). Two CRISPR-Cas systems in *Methanosarcina mazei* strain Gö1 display common processing features despite belonging to different types I and III. *RNA Bio* 10, 779–791.
- Niewoehner, O., and Jinek, M. (2016). Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *Rna* 22, 318–329.
- Niewoehner, O., Jinek, M., and Doudna, J.A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Research* 42, 1341–1353.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949.
- Nunez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., and Doudna, J.A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell* 62, 824–833.
- Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527, 535–538.
- Nunez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Publishing Group* 21, 528–534.
- Nunez, J.K., Lee, A.S.Y., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193–198.
- Ogura, T., and Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc Natl Acad Sci U S A* 80, 4784–4788.
- Olive, V., Jiang, I., and He, L. (2010). mir-17-92, a cluster of miRNAs in the midst of the cancer network. *Int. J. Biochem. Cell Biol.* 42, 1348–1354.
- Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D.K., and Aravin, A.A. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. *Molecular Cell* 51, 594–605.
- Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K.-I., Stahl, B., Barrangou, R., and Banfield, J.F. (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature Communications* 4, 1430.

- Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* 12, 208–214.
- Park, S.K., Venable, J.D., Xu, T., and Yates, J.R. (2008). A quantitative analysis software tool for mass spectrometry-based proteomics. *Nat Meth* 5, 319–322.
- Patterson, A.G., Chang, J.T., Taylor, C., and Fineran, P.C. (2015). Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Research* 43, 6038–6048.
- Patterson, A.G., Jackson, S.A., Taylor, C., Evans, G.B., Salmond, G.P.C., Przybilski, R., Staals, R.H.J., and Fineran, P.C. (2016). Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Molecular Cell* 64, 1102–1108.
- Peng, J., Elias, J.E., and Thoreen, C.C. (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC– MS/MS) for large-scale protein analysis: the yeast proteome. *Journal of Proteome* ....
- Perez-Rodriguez, R., Haitjema, C., Huang, Q., Nam, K.H., Bernardis, S., Ke, A., and DeLisa, M.P. (2010). Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in *Escherichia coli*. *Molecular Microbiology* 79, 584–599.
- Pommer, A.J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kühlmann, U.C., Cooper, A., Connolly, B.A., Hemmings, A.M., Moore, G.R., et al. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *Journal of Molecular Biology* 314, 735–749.
- Pul, Ü., Wurm, R., Arslan, Z., Geißen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR- caspromoters and their silencing by H-NS. *Molecular Microbiology* 75, 1495–1512.
- Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J., and Gerard, G.F. (2004). Mutation detection using Surveyor nuclease. *BioTechniques* 36, 702–707.
- Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* 1–25.
- Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N.J., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H.J., and Fineran, P.C. (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Research* 42, 8516–8526.
- Rohrman, B.A., Leautaud, V., Molyneux, E., and Richards-Kortum, R.R. (2012). A lateral flow assay for quantitative detection of amplified HIV-1 RNA. *PLoS ONE* 7, e45611.
- Rollie, C., Schneider, S., Brinkmann, A.S., and Bolt, E.L. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife*.
- Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., and Marraffini, L.A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* 161, 1164–1174.
- Sampson, T.R., and Weiss, D.S. (2014). Exploiting CRISPR/Cas systems for biotechnology. *BioEssays* 36, 34–38.
- Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L., and Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497, 254–257.

- Sapranaukas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* *39*, 9275–9282.
- Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular Cell* *46*, 606–615.
- Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Bio* *10*.
- Scholz, I., Lange, S.J., Hein, S., Hess, W.R., and Backofen, R. (2013). CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS ONE* *8*, e56470.
- Seed, K.D. (2015). Battling Phages: How Bacteria Defend against Viral Attack. *PLoS Pathog.* *11*, e1004847.
- Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* *108*, 10098–10103.
- Semenova, E., Kuznedelov, K., Datsenko, K.A., Boudry, P.M., Savitskaya, E.E., Medvedeva, S., Beloglazova, N., Logacheva, M., Yakunin, A.F., and Severinov, K. (2015). The Cas6e ribonuclease is not required for interference and adaptation by the *E. coli* type I-E CRISPR-Cas system. *Nucleic Acids Research* *43*, 6049–6061.
- Shah, S.A., Erdmann, S., Mojica, F.J.M., and Garrett, R.A. (2013). Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Bio* *10*, 891–899.
- Shechner, D.M., Hacısuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Meth* *12*, 664–670.
- Sheppard, N.F., Glover, C.V.C., 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., Savitskaya, E., Semenova, E., Logacheva, M.D., Datsenko, K.A., and Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Research* *42*, 5907–5916.
- Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* *15*, 169–182.
- Silas, S., Mohr, G., Sidote, D.J., Markham, L.M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A.M., and Fire, A.Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* *351*, aad4234–aad4234.
- Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci U S A* *108*, 20497–20502.

- Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *Embo J.* 32, 385–394.
- Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler, P., Shmakov, S., Makarova, K.S., et al. (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular Cell* 65, 618–630.e7.
- Snyder, L. (1995). Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Molecular Microbiology* 15, 415–420.
- Sokolowski, R.D., Graham, S., and White, M.F. (2014). Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. *Nucleic Acids Research* 42, 6532–6541.
- Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Molecular Cell* 52, 146–152.
- Staals, R.H.J., 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*. *Molecular Cell* 52, 135–145.
- Staals, R.H.J., 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*. *Molecular Cell* 56, 518–530.
- Stamatakis, A. (2014a). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Stamatakis, A. (2014b). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Sternberg, S.H., and Doudna, J.A. (2015). Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular Cell* 58, 568–574.
- Sternberg, S.H., Haurwitz, R.E., and Doudna, J.A. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *Rna* 18, 661–672.
- Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* 527, 110–113.
- Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPRRNA-guided endonuclease Cas9. *Nature* 1–17.
- Sternberg, S.H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. *Molecular Cell* 61, 797–808.
- Swarts, D.C., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., and van der Oost, J. (2015). Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Research* 43, 5120–5129.
- Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., et al. (2014). DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507, 258–261.
- Swarts, D.C., Mosterd, C., van Passel, M.W.J., and Brouns, S.J.J. (2012). CRISPR

- Interference Directs Strand Specific Spacer Acquisition. *PLoS ONE* 7, e35888.
- Swarts, D.C., Szczepaniak, M., Sheng, G., Chandradoss, S.D., Zhu, Y., Timmers, E.M., Zhang, Y., Zhao, H., Lou, J., Wang, Y., et al. (2017). Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Molecular Cell* 65, 985–998.e986.
- Tabb, D.L., McDonald, W.H., and Yates, J.R., III (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *Journal of Proteome* ....
- Tamulaitis, G., Venclovas, Č., and Siksnys, V. (2017). Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology* 25, 49–61.
- Taylor, D.W., Zhu, Y., Staals, R.H.J., Kornfeld, J.E., Shinkai, A., van der Oost, J., Nogales, E., and Doudna, J.A. (2015). Structural biology. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science* 348, 581–585.
- Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* 14, 321–327.
- Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends in Genetics* 30, 111–118.
- Thomas, M., Lange-Grünweller, K., Hartmann, D., Golde, L., Schlereth, J., Streng, D., Aigner, A., Grünweller, A., and Hartmann, R. (2013). Analysis of Transcriptional Regulation of the Human miR-17-92 Cluster; Evidence for Involvement of Pim-1. *Ijms* 14, 12273–12296.
- Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11, 636–646.
- van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 12, 479–492.
- van Houte, S., Buckling, A., and Westra, E.R. (2016). Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol. Mol. Biol. Rev.* 80, 745–763.
- Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77, 53–72.
- Walker, F.C., Chou-Zheng, L., Dunkle, J.A., and Hatoum-Aslan, A. (2016). Molecular determinants for CRISPR RNA maturation in the Cas10–Csm complex and roles for non-Cas nucleases. *Nucleic Acids Research* 45, gkw891–12.
- Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* 163, 840–853.
- Wang, R., Li, M., Gong, L., Hu, S., and Xiang, H. (2016). DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Research* 44, 4266–4277.
- Washburn, M.P., Wolters, D., and Yates, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19, 242–247.
- Wei, Y., Chesne, M.T., Terns, R.M., and Terns, M.P. (2015a). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Research*.
- Wei, Y., Terns, R.M., and Terns, M.P. (2015b). Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes & Development* 29, 356–361.

- Wernersson, R., Juncker, A.S., and Nielsen, H.B. (2007). Probe selection for DNA microarrays using OligoWiz. *Nat Protoc* 2, 2677–2691.
- Westra, E.R., Semenova, E., Datsenko, K.A., and Jackson, R.N. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS* ....
- Westra, E.R., Pul, Ü., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Molecular Microbiology* 77, 1380–1393.
- Westra, E.R., van Erp, P.B.G., Künne, T., Wong, S.P., Staals, R.H.J., Seegers, C.L.C., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular Cell* 46, 595–605.
- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338.
- Wright, A.V., and Doudna, J.A. (2016). Protecting genome integrity during CRISPR immune adaptation. *Nature Publishing Group* 23, 876–883.
- 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.
- Wu, H., Lima, W.F., and Crooke, S.T. (1999). Properties of cloned and expressed human RNase H1. *J. Biol. Chem.* 274, 28270–28278.
- Xu, T., Park, S.K., Venable, J.D., Wohlschlegel, J.A., Diedrich, J.K., Cociorva, D., Lu, B., Liao, L., Hewel, J., Han, X., et al. (2015). ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J Proteomics* 129, 16–24.
- Xue, C., Seetharam, A.S., Musharova, O., Severinov, K., Brouns, S.J.J., Severin, A.J., and Sashital, D.G. (2015). CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Research* 43, 10831–10847.
- Yagi, Y., Nakamura, T., and Small, I. (2014). The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.* 78, 772–782.
- Yamaguchi, Y., and Inouye, M. (2011). Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat. Rev. Microbiol.* 9, 779–790.
- Yamaguchi, Y., Park, J.-H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet.* 45, 61–79.
- Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* 165, 949–962.
- Yan, L., Zhou, J., Zheng, Y., Gamson, A.S., Roembke, B.T., Nakayama, S., and Sintim, H.O. (2014). Isothermal amplified detection of DNA and RNA. *Mol Biosyst* 10, 970–1003.
- Yang, C.-D., Chen, Y.-H., Huang, H.-Y., Huang, H.-D., and Tseng, C.-P. (2014). CRP represses the CRISPR/Cas system in *Escherichia coli*: evidence that endogenous CRISPR spacers impede phage P1 replication. *Molecular Microbiology* 92, 1072–1091.
- Yang, H., Gao, P., Rajashankar, K.R., and Patel, D.J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell* 167,

1814–1828.e12.

- Yang, W. (2011). Nucleases: diversity of structure, function and mechanism. *Quart. Rev. Biophys.* *44*, 1–93.
- Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.-W., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* *504*, 168–171.
- Yoganand, K.N.R., Sivathanu, R., Nimkar, S., and Anand, B. (2017). Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Research* *45*, 367–381.
- Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research* *40*, 5569–5576.
- Yosef, I., Shitrit, D., and Goren, M.G. (2013). DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array.
- Young, J.C., Dill, B.D., Pan, C., Hettich, R.L., Banfield, J.F., Shah, M., Fremaux, C., Horvath, P., Barrangou, R., and VerBerkmoes, N.C. (2012). Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS ONE* *7*, e38077.
- Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., et al. (2017). Guide-independent DNA cleavage by archaeal Argonaute from *Methanocaldococcus jannaschii*. *Nat Microbiol* *2*, 17034.
- 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.
- Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., et al. (2016). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat Biotechnol* *35*, 31–34.
- Zhang, J., Graham, S., Tello, A., Liu, H., and White, M.F. (2016). Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Research* *44*, 1789–1799.
- Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.-V., et al. (2012). Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Molecular Cell* *45*, 303–313.
- Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, H.S., Schoen, C., Vogel, J., and Sontheimer, E.J. (2013). Processing-Independent CRISPR RNAs Limit Natural Transformation in *Neisseria meningitidis*. *Molecular Cell* *50*, 488–503.
- Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* *515*, 147–150.