THE PARADOXICAL ROLE OF ATP

IN VIRAL INFECTIONS AND CELLULAR DEFENSE

by

Trevor Ronald Zahl

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Microbiology and Immunology

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2024

©COPYRIGHT

by

Trevor Ronald Zahl

2024

All Rights Reserved

TABLE OF CONTENTS

1. INTRODUCTION 1

References 6

2. CRISPR-CAS, ARGONAUTE PROTEINS AND THE EMERGING LANDSCAPE OF AMPLIFICATION-FREE DIAGNOSTICS 7

Manuscript Information 9
Abstract 10
Introduction 11
Next-Generation Programmable Diagnostics 15
Target-Activated Collateral Nucleases 15
Digital Enzymology 16
Amperometric Biosensors 17
Surface-Enhanced Raman Scattering Analysis of Lateral Flow assays 18
CRISPR Feedback Amplification Circuits 18
Target-Activated Signal Amplification 21
CRISPR-Catalyzed ATP Polymerization 22
Engineered CRISPR Signaling Cascades 23
Target-Guided Nucleases 26
Target Binding 28
Graphene Field Effect Transistor Sensing of DNA 28
Reconstitution of Split Reporter Enzyme 29
Suggested Reporting Metrics for Molecular Diagnostics 31
Outlook 33
Manuscript References 36

3. SEQUENCE-SPECIFIC CAPTURE AND CONCENTRATION OF VIRAL RNA BY TYPE III CRISPR SYSTEM ENHANCES DIAGNOSTIC 49

Manuscript Information 51
Abstract 52
Introduction 53
Results 56
Discussion 71
Methods 101
Ethical Statement 101
Human Clinical Sample Collection and)reparation 101
Nucleic Acids 101
Non-Targeting Control (NTC) 102

TABLE OF CONTENTS CONTINUED

Plasmids 102
Protein Expression and Purification 103
^{32}P-Labeling of RNA Oligos 104
Binding and Pull-Down of RNA Oligos with TtCsm 104
Complexing of TtCsm with Magnetic Beads 105
Thin-Layer Chromatography (TLC) 105
Type III-Based RNA Detection 106
Type III-Based RNA Pull-Down and Detection 107
Pull-Down and Detection from RNA Samples and Nasopharyngeal Swabs 107
RT-qPCR 108
Nanopore Sequencing of DNA Cleavage Fragments 109
RNA and DNA Reporter's Libraries 110
In vitro DNA and RNA Cleavage Assays 111
Phylogenetic Analysis of Can1 and Can2 Proteins 111
Phylogenetic Analysis of NucC 113
Statistics & Reproducibility 113
Manuscript References 115

4. CRISPR-BASED ENGINEERING OF RNA VIRUSES 118

Manuscript Information 119
Abstract 120
Introduction 121
Results 122
Coupling CRISPR-Based RNA Cleavage and Splint Ligation for RNA Editing 122
Programmed Deletions in the RNA Virus Genome 126
Sequencing and Isolation of Edited Viruses 129
Programmable Insertion of New RNA in the Viral Genome 131
Discussion 133
Materials and Methods 135
Plasmids 135
Nucleic Acids 136
Protein Expression and Purification 137
SINV Genomic RNA Purification 138
Editing of in vitro Transcribed RNA 138
Deletion of Viral RNA 139
Substitution of Viral RNA 140
Electroporation of Edited Viral RNA 140
Fluorescence Imaging 141
Plaque Purification 141

TABLE OF CONTENTS CONTINUED

Reverse Transcription qPCR 142
Nanopore Sequencing 142
Sequencing Data Analysis 144
Statistical Analysis 145
Acknowledgments 146
Manuscript References 165

5. VIRAL PROTEINS ACTIVATE PARIS-MEDIATED TRNA DEGRADATION AND VIRAL TRNAS RESCUE INFECTION 170

Manuscript Information 173
Abstract 175
Introduction 176
Results 177
PARIS Assembles into a Propeller Shaped Macromolecular Complex 177
AriA is an ABC ATPase that Sequesters the AriB TOPRIM Nuclease 180
AriA Binds to the Ocr Anti-Restriction Protein Leading to AriB Release and Activation 183
PARIS Activation Results in Nucleoid Compaction and Translation Inhibition 188
Phage T5 Encodes Two Protein Triggers and a Suppressor of PARIS 191
AriB Cleaves the Host tRNALys and Phage T5 Encodes a Non-Cleavable tRNALys Substitute 192
Evolutionary History of the PARIS Defense System 195
Discussion 197
Methods 199
Bacterial Strains, Phages, and Plasmids 199
Expression and Purification of the PARIS Complex 200
In vivo Protein Pull-Down 200
Glutaraldehyde Cross-Linking Assay 201
In vitro Protein Pull-Down 201
Cryo-EM Sample Preparation and Data Collection 202
Cryo-EM Data Processing of Inactive PARIS 203
Cryo-EM Data Processing of Ocr-Pulldown 204
Model Building and Refinement 205
ATPase Assays 206
Efficiency of Plaquing Assay 207
Liquid Culture Phage Infection 207
Toxicity Assay 208

TABLE OF CONTENTS CONTINUED

Fluorescence Microscopy 208
Extraction of Total DNA 209
TUNEL Assay 209
Hi-C Procedure and Sequencing 209
Metabolic Labeling 210
In vitro Translation 212
Extraction of Total RNA 213
Northern Blot 214
Defense System Detection 215
Defense System ATPase AAA15/21 Tree 216
Acknowledgements 216
Manuscript References 218

6. CONCLUSION 223

Current Landscape of CRISPR Diagnostics 223
In vitro RNA Editing 224
Mechanistic Understanding of Bacterial Antigen Sensing 225
Outlook 226
References 227

APPENDICES 229

IN SILICO MODELLING OF TYPE III CRISPR DIAGNOSTIC 230
References 234

PARIS MEDIATED DEPLETION OF INTRACELLULAR ATP 235

REFERENCES CITED 238

LIST OF TABLES

Table Page

1. Table S1. Recommended Information for publication of CRISPR-dx and pAgo-dx Assays checklist. 32
2. Supplementary table 3.1. Fluorescent RNA and DNA reporters. 95
3. Supplementary Table 3.2. RNase P RNA levels in nasopharyngeal swab samples and diluted total RNA from HEK 293T cells used as negative control in the manuscript. 99
4. Supplementary Table 3.3. Oligonucleotide primers used to generate DNA templates for in vitro transcription with T7 polymerase 100
5. Table 4.S1. Oligonucleotides used for cloning anti-GFP and anti-BGP CRISPR loci 160
6. Table 4.S2. DNA splints and RNA inserts used for RNA editing. 161
7. Table 4.S3. Oligonucleotide primers used to generate DNA templates for in vitro transcription of RNA. 162
8. Table 4.S4. Oligonucleotide primers used for PCR and qPCR. 163
9. Table 4.S5. PCR primers used for amplicon sequencing of SINV-GFP and SINV-BFP genomes. 164

LIST OF FIGURES

Figure Page

1. Figure 2.1 Evolution of molecular diagnostics. 14
2. Figure 2.3. CRISPR-mediated signalling inspires new CRISPR-dx applications. 25
3. Figure 2.4. Target-guided nucleases and sensitive methods to detect target binding 30
4. Figure 5. The emerging landscape of pre-amplification-independent CRISPR-dx and pAgo-dx. 35
5. Fig 3.1: Type III CRISPR-based RNA concentration enhances sequence-specific detection of RNA..... 58
6. Fig. 3.2: Can1 and Can2 ancillary nucleases function as cA4-activated nucleases on plasmid DNA, ssRNA and ssDNA 62
7. Fig.3.3: Incorporation of Can2 nuclease improves sensitivity of Csm-based RNA detection..... 64
8. Fig.3.4: Incorporation of cA3-activated nucleases into Csm-based RNA detection assay. 67
9. Fig.3.5: TtCsm-based RNA capture directly detects SARS-CoV-2 in clinical samples..... 70
10. Supplementary Figure 3.1: Type III CRISPR-based RNA concentration enhances detection. 75
11. Supplementary Figure 3.2: Biochemical characterization of Thermus thermophilus Can1 (TtCan1) nuclease activities. 76
12. Supplementary Figure 3.3: Purification of Can2 from Archaeoglobi archaeon JdFR-42. 77
13. Supplementary Figure 3.4: Biochemical characterization of Can2 from Archaeoglobi archaeon JdFR-42. 78
14. Supplementary Figure 3.5: TtCan1 and AaCan2 linearize plasmid DNA in unwound regions. 79

LIST OF FIGURES CONTINUED

15. Supplementary Figure 3.6: Sequence cleavage preference for AaCan2 and TtCan1 81

16. Supplementary Figure 3.7: TtCan1 and AaCan2 nuclease activities at different temperatures and salt concentrations. 82

17. Supplementary Figure 3.8: TtCan1 and AaCan2 do not degrade cA3 or cA4. 83

18. Supplementary Figure 3.9: Decreasing TtCsm-complex concentration in RNA detection assay reduces the background signal. 84

19. Supplementary Figure 3.10: Purification and biochemical characterization of thermostable NucC orthologs 86

20. Supplementary Figure 3.11: Thermostable NucC homologs Amtb01NucC, EsNucC, and CtNucC demonstrate sequence cleavage preference of plasmid DNA 87

21. Supplementary Figure 3.12: CtNucC nuclease activity at different temperatures and salt concentrations. 88

22. Supplementary Figure 3.13: Integration of cA3-activated nucleases into Csm-based RNA detection assay does not improve sensitivity. 89

23. Supplementary Figure 3.14: Optimization of TtCsm-based RNA capture and detection assay. 90

24. Supplementary Figure 3.15: Testing effect of detergents and chelating agents on TtCsm-based RNA capture and detection assay. 91

25. Supplementary Figure 3.16: Direct detection of SARS-CoV-2 RNA in clinical samples using TtCsm-based RNA capture. 92

26. Supplementary Figure 3.17: The evolution of type III-based diagnostics. 93

27. Supplementary Figure 3.18: Alignment of N gene sequences targeted by TtCsm 94

28. Fig. 4.1. RNA editing with CRISPR-based cleavage and RNA ligation. 125

29. Fig. 4.2. Programmed deletion of specific RNA sequence from the SINV-GFP genome. 128

LIST OF FIGURES CONTINUED

30. Fig. 4.3. Plaque purification of edited viruses. 130

31. Fig. 4.4. Programmed substitutions in viral genomes. 132

32. Fig. 4.S1. Purification of type III CRISPR complex of Streptococcus thermophilus (SthCsm). 147

33. Fig. 4.S2 Architecture of recombinant Sindbis GFP virus (SINV-GFP) and tittering viral RNA. 149

34. Fig. 4.S3. Purification and characterization of SthCsm-complexes targeting gfp gene in Sindbis virus genome. 150

35. Fig. 4.S4. CRISPR-based programmed deletions in SINV-GFP. 151

36. Fig. 4.S5. Direct cDNA sequencing with Nanopore. 152

37. Fig. 4.S6. Plaque purification of the edited SINV-dGFP virus. 153

38. Fig. 4.S7. Amplicon sequencing of plaque-purified viruses to confirm programmed deletions in the viral genome. 154

39. Fig. 4.S9. Genome sequencing of bulk SINV-BFP-to-GFP viruses (A) Direct cDNA sequencing of bulk SINV-BFP-to-GFP edited or unedited viruses. 158

40. Fig. 4.S10. Genotyping of plaque-purified SINV-BFP-to-GFP viruses 159

41. Figure 5.1: PARIS is a two-component system that assembles into a propeller-shaped supramolecular complex. 179

42. Figure 5.2: Interaction between the AriA ATPase and AriB are essential for PARIS-mediated defense. 182

43. Figure 3: AriA interacts with the Ocr trigger in an ATPase-dependent manner, leading to the release of AriB from the PARIS complex. 186

44. Figure 5.4: The activation of PARIS leads to cell death and inhibition of translation. 190

45. Figure 5: PARIS cleaves the *E. coli* tRNALys and T5 bypasses immunity by encoding its own non-cleavable tRNALys.. 194

LIST OF FIGURES CONTINUED

46. Figure 5.6: Phylogenies of PARIS and related ABC ATPase powered defense systems. 196

47. Figure A1: Mathematical model of type III CRISPR based diagnostic 232

48. Figure A2: All steps of diagnostic contribute equally to theoretical LoD 233

ABSTRACT

Conflict between prokaryotes and viruses has driven the evolution of diverse defense systems. The study of these systems has not only expanded our knowledge of the mechanisms underlying genetic conflict, but has also revolutionized biotechnologies.

I joined the Wiedenheft lab in early 2021 during the height of the SARS-CoV-2 pandemic. As such, the lab was focused on developing technologies to aid in the identification and study of SARS-CoV-2. I joined a team that was actively repurposing a type III CRISPR system as a molecular diagnostic. Type III CRISPRs are RNA targeting systems that unleash antiviral signaling cascades upon recognition of a target. Previous work in the lab had shown that this activity could be repurposed as a molecular diagnostic, but the first generation technology was not sensitive enough for clinical use. To address this need, we developed a technology termed Capture and Concentrate, which utilized 'CRISPR magnets' to concentrate target RNAs from complex mixtures. Using this technology, we increased diagnostic sensitivity by 1,000 fold. In addition to my work at the bench, I developed an *in silico* model of the diagnostic to inform future development efforts.

While type III CRISPRs are unique due to their ability to generate antiviral signaling cascades, they also contain intrinsic endonuclease activity. In 2023 we demonstrated that this activity could be repurposed as an RNA editing technology. Using the type III system from *Streptococcus thermophilus*, we introduced precise genomic modifications in a recombinant sindbis virus. This technology removed the need for reverse genetics systems, which hampered the study of the rapidly evolving SARS-CoV-2 virus.

In the final year of my graduate studies, I worked with a team focused on determining how, PARIS, a bacterial innate immune system, induced host death upon viral infection. PARIS is a two-component system comprised of the ATPase, AriA, and a nuclease, AriB. Previous studies had shown that PARIS induced abortive infection in response to viral proteins, but the mechanism was not known. We determined that upon antigen recognition, AriB is activated and cleaves host tRNAs to induce cell death and prevent community spread.

CHAPTER ONE

INTRODUCTION

Viruses that infect bacteria, phage, are the most abundant biological agent on Earth, with an estimated 10^{31} virions present at any given time. Their staggering titers play critical roles in nutrient cycling in almost every environment. In the oceans alone, viral lysis of microbes turns over billions of tons of biomass daily[1]. However, while phage are the most abundant biological agent, they are also the simplest. Phage are requisite parasites that require a living host to reproduce. Viruses do not contain the molecular machinery to perform energy metabolism or protein synthesis, they must utilize the host's ribosomes and energy stores to create progeny virions. However, despite their lack of essential metabolic genes, phage contain sophisticated molecular machinery to subvert host defenses and gain control of essential host functions.

However, host cells are not defenseless in this conflict. Like their eukaryotic counterparts, Bacteria and Archaea have correspondingly equipped themselves with an ever-expanding arsenal of immune systems. Prokaryotes have both innate immune systems, which recognize fixed antigenic signatures, and an adaptive immune system[2]. The study of prokaryotic defense systems has not only expanded our understanding of the natural world, but has also led to major biotechnological developments, like restriction modification-based cloning methods, and CRISPR Cas9-derived treatments for genetic disease[3].

My research efforts have focused on determining the function and applications of ATP-dependent immune responses. ATP is the fourth most abundant intracellular metabolite, with concentrations reaching up to ~10 mM, and a key energy storage compound. While the hydrolysis of ATP powers energetically unfavorable reactions within the cell, previous work on

type III CRISPRs and other prokaryotic immune systems has revealed that ATP plays an important role in sensing and eliminating viral infections[4,5]. We hypothesized that interrogation of ATP based defense systems would not only increase our understanding of prokaryotic viral defense but would also allow for the development of noevel biotechnologies.

In early 2020, the onslaught of COVID-19 cases revealed an urgent need for new decentralized diagnostics that provide users with immediate results in any environment without complex procedures, or sophisticated equipment. Polymerase Chain Reaction (PCR) is the gold standard for molecular diagnostics but suffers from the need for laboratories and trained experts to carry out testing. Rapid antigen tests have proven to be a quick alternative to PCR-based molecular tests, but at the cost of decreased sensitivity. Thus, there is an urgent need for inexpensive new technologies that enable fast, reliable, and scalable detection of viruses.

To address these challenges, we repurposed a type III CRISPR system as a molecular diagnostic. Although CRISPR-Cas9 has attracted the most attention for its applications as a programmable endonuclease, there are five other subtypes with diverse biochemical and mechanistic properties that present new opportunities for innovation[6]. To date, the type V and VI effector proteins Cas12 and Cas13 have been the most widely repurposed CRISPR-Cas enzymes for molecular diagnostics. Both Cas12 and Cas13-based diagnostics rely on their collateral nuclease activities. Specifically, when CRISPR RNA guided Cas12 or Cas13 bind to a complementary target sequence, they activate a non-sequence specific ssDNase or RNase activity, respectively. The target activated non-sequence specific nuclease activity has been leveraged to cleave a nucleic acid tether that links a fluorophore to a quencher. Cleavage of the tether, liberates the fluorophore, resulting in a fluorescent signal. While Cas12 and Cas13 based diagnostics have

been the most widely adopted systems for diagnostic use, they lack the sensitivity for direct detection of nucleic acid from clinical samples. To combat this deficiency, current Cas12 and Cas13 based diagnostic utilize an isothermal nucleic acid preamplification step[7,8]. In 2021 the Santiago group demonstrated that Cas12 and Cas13 based diagnostics will always need a preamplification step to reach relevant sensitivities. The rate of Cas12 or Cas13 mediated reporter cleavage is simply not fast enough to generate a measurable signal in a rapid diagnostic[9]. To combat the need for pre-amplification-driven CRISPR diagnostics, we turned to using the type III CRISPR effector from *Thermus thermophilus* (TtCsm) for direct detection of SARS-CoV-2 RNAs from patient samples.

Like type VI, type III systems also recognize complementary RNA. However, unlike any other CRISPR system, target recognition by type III complexes simultaneously activates polymerase and HD-nuclease domains in the Cas10 subunit. The polymerase domain has been estimated to generate ~1000 cyclic oligoadenylates per bound RNA, which trans-activate and allosterically regulate diverse multi-turnover ancillary nucleases that provide defense from invading genetic parasites. This biochemical cascade exponentially amplifies the signal when a type III complex detects target RNA, suggesting that these systems have the potential to enhance the sensitivity of CRISPR-based diagnostics[5,10].

As the global impacts of the SARS-CoV-2 pandemic began to subside, my interests shifted. I began working on a project that aimed to determine how the prokaryotic immune system PARIS defends against phage infection. PARIS (Phage Anti Restriction Induced System) was discovered on integrated prophages that contained hotspots of genetic diversity. PARIS is a multimeric protein complex comprised of two subunits, AriA and AriB. AriA is an ABC ATPase while AriB is a

putative nuclease. PARIS is a secondary defense system that responds to phage-derived immunomodulatory proteins[11]. Many phages produce proteins that inhibit a bacteria's first line of defense like CRISPR and Restriction modification systems[12]. Through a previously unknown mechanism, PARIS recognizes these viral proteins and attenuates the infection. We hypothesized that upon viral antigen recognition, AriB would indiscriminately cleave viral and host nucleic acids while AriA would increase the rate of ATP hydrolysis to kill the host and provide an altruistic form of protection for the immediate bacterial community.

Chapter 2- This chapter focuses on a review that I coauthored in 2021 that examines the current landscape of CRISPR-Cas and argonaute based diagnostics. While CRISPR-Cas has gained notoriety for its role in genome engineering, there has been an explosion of diagnostic technologies based on CRISPR-Cas and other RNA guided nucleases like Agos. We comprehensively identified current diagnostics that rely on CRISPR-Cas and Agos while also looking at technologies that have the potential to increase the sensitivity of these diagnostics and remove the need for target preamplification.

Chapter 3- When I joined the Wiedenheft lab in 2021 amid the SARS-Cov-2 pandemic, I immediately started developing an alternative technology for rapid identification of RNA viruses. We hypothesized that we could reprogram type III CRISPRs for the *in vitro* detection of clinically relevant RNAs. While previous work in the lab had demonstrated that it was possible to reprogram type III interference complexes to recognize SARS-CoV-2 genomic RNA, diagnostic sensitivity was still insufficient for clinical use. To combat low sensitivity, I contributed to the development of a technique we termed Capture and Concentrate. This innovation increased diagnostic sensitivity four orders of magnitude by using 'CRISPR magnets'

to concentrate target RNAs from large sample volumes. I also developed a mathematical model of the diagnostic to identify kinetic checkpoints that have the most significant effect on diagnostic sensitivity. My model showed that all steps of the diagnostic essentially contribute equally to overall sensitivity.

Chapter 4- This chapter outlines a CRISPR based RNA editing approach we developed in response to the lack of direct RNA editing technologies. We hypothesized that Type III CRISPRs could be used as precise RNA editing technologies. To test this hypothesis, we repurposed the Type III CRISPR effector from *Streptococcus thermophilus* (SthCsm) to make programmable cuts in a recombinant sindbis virus. Using SthCsm, an RNA ligase and custom DNA splints we were able introduce, stop and alternative codons in a GFP gene fused to the sindbis genome.

Chapter 5- In 2023 I began work aimed at determining the role of ATP turnover in an innate immune system termed PARIS. PARIS is a secondary line of defense that provides cellular defense against viruses that shut down the host's primary defense systems. We showed that ATP turnover is necessary for antigen recognition by PARIS but is does not deplete intracellular ATP. Additionally, we demonstrated that antigen recognition releases the AriB nuclease to cleave lysine tRNAs to induce host senescence and prevent viral spread to neighboring cells. Specifically, I showed that while PARIS ATP hydrolysis is not fast enough to deplete cellular ATP stocks and antigen recognition slows ATP hydrolysis. These results contrast with our original hypothesis that PARIS provides an abortive infection phenotype by depleting host ATP stores upon antigen recognition.

References

1. Keen, E. C. A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **37**, 6–9 (2015).

2. Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556-1569.e5 (2022).

3. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **5**, 593–599 (2009).

4. Rousset, F. *et al.* A conserved family of immune effectors cleaves cellular ATP upon viral infection. *Cell* **186**, 3619-3631.e13 (2023).

5. Niewoehner, O. *et al.* Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).

6. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).

7. Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O. & Zhang, F. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nat. Protoc.* **14**, 2986–3012 (2019).

8. Broughton, J. P. *et al.* CRISPR–Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* **38**, 870–874 (2020).

9. Ramachandran, A. & Santiago, J. G. CRISPR Enzyme Kinetics for Molecular Diagnostics. *Anal. Chem.* **93**, 7456–7464 (2021).

10. Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *eLife* **7**, e36734 (2018).

11. Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740-753.e5 (2022).

12. Yirmiya, E. *et al.* Phages overcome bacterial immunity via diverse anti-defence proteins. *Nature* **625**, 352–359 (2024).

CHAPTER TWO

CRISPR-CAS, ARGONAUTE PROTEINS AND THE EMERGING LANDSCAPE OF AMPLIFICATION-FREE DIAGNOSTICS

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Co-Author: Andrew Santigo-Frangos

Contributions: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Project Administration, Supervision, Visualization, Writing -original draft, Writing – review and editing

Co-Author: Artem Nemudryi

Contributions: Investigation, Writing -original draft, Writing – review and editing

Author: Anna Nemudraia

Contributions: Investigation, Writing -original draft, Writing – review and editing

Co-Author: Tanner Wiegand

Contributions: Visualization

Co-Author: Joseph E. Nichols

Contributions: Data Curation

Co-Author: Pushya Krishna

Contributions: Data Curation

Co-Author: Andrew M. Scherffius

Contributions: Data Curation

Co-Author: Trevor Zahl

Contributions: Data Curation

Co-Author: Royce A. Wilkinson

Contributions: Writing – review and editing

Co-Author: Blake Wiedenheft

Contributions: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Project Administration, Supervision, Visualization, Writing -original draft, Writing – review and editing

Manuscript Information

Andrew Santiago-Frangos, Artem Nemudryi, Anna Nemudraia, Tanner Wiegand, Joseph Nichols, Pushya Krishna, Andrew M. Scherffius, Trevor Zahl, Royce A. Wilkinson and Blake Wiedenheft

Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717, USA

Methods

Status of Manuscript:

___ Prepared for submission to a peer-reviewed journal

___ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

_X_ Published in a peer-reviewed journal

Methods

Volume 205

11 June 2022

DOI: 10.1016/j.ymeth.2022.06.002

Abstract

Polymerase Chain Reaction (PCR) is the reigning gold standard for molecular diagnostics. However, the SARS-CoV-2 pandemic reveals an urgent need for new diagnostics that provide users with immediate results without complex procedures or sophisticated equipment. These new demands have stimulated a tsunami of innovations that improve turnaround times without compromising the specificity and sensitivity that has established PCR as the paragon of diagnostics. Here we briefly introduce the origins of PCR and isothermal amplification, before turning to the emergence of CRISPR-Cas and Argonaute proteins, which are being coupled to fluorimeters, spectrometers, microfluidic devices, field-effect transistors, and amperometric biosensors, for a new generation of nucleic acid-based diagnostics.

Introduction

Eureka moments are rare in science, but that's how Kary Mullis described his conception of the polymerase chain reaction (PCR) [1]. While the idea for PCR arrived in an instant, the development of this assay required time and optimization. Initially, PCR was slow, insensitive, error-prone, labor-intensive, and did not involve thermal cycling. However, in March of 1985, Mullis and his team demonstrated that PCR could be used to detect mutations responsible for sickle cell anemia [2,3]. This proof-of-principle stimulated a flurry of innovations, including the use of thermostable polymerases, proofreading enzymes, incorporation of reverse transcriptases that enable detection of RNA, and the development of automated thermocyclers, which dramatically increase sensitivity and reduce contamination (**Fig. 2.1**). Now, after almost forty years of development, PCR continues to serve as the backbone for most molecular diagnostics. However, thermocyclers and the technical expertise necessary to perform these reactions have largely confined PCR to centralized testing facilities. Moreover, once samples have been transported to the lab, PCR generally requires about an hour to complete. While these requirements are compatible with diagnostic testing in most situations, the SARS-CoV-2 pandemic reveals an urgent need for innovations that enable fast, reliable, and scalable detection of viruses without technical requirements or time limitations intrinsic to PCR.

Isothermal amplification methods (e.g., LAMP, RPA, NASBA, etc.) use a variety of different strategies to eliminate the need for thermocyclers, and as such, these methods have garnered considerable interest for distributed diagnostics (**Fig. 2.1**) [4,5]. In addition, isothermal amplification methods have been coupled to colorimetric or fluorescent indicators that enable simple and rapid molecular detection. While isothermal amplification methods are sensitive and

fast, they are prone to mispriming or template-independent amplification, which increases the frequency of false positives. Therefore, amplified nucleic acids generated by isothermal amplification often require confirmation by more specific methods, including CRISPR- and Argonaute-based diagnostics or high-throughput sequencing [6,7,16–22,8–15].

CRISPR-based diagnostics (CRISPR-dx) refers to a growing repertoire of CRISPR (clusters of regularly interspaced short palindromic repeats) RNA-guided nucleases that have been creatively repurposed for sequence-specific detection of RNA or DNA targets [23,24]. The sequence specificity of CRISPR RNA-guided nucleases, which naturally provide bacteria and archaea with defense from invading genetic parasites, are emerging as programable new diagnostics [23–27]. Similarly, prokaryotic Argonautes (pAgos) are a phylogenetically and functionally diverse family of RNA- and DNA-guided proteins [28,29] . While some pAgo variants directly cleave bound targets, others lack nuclease activity and recruit trans-acting effectors with diverse enzymatic activities, reminiscent of the diversity observed in CRISPR systems [30–36]. The biological roles of pAgo are diverse, though several have been shown to function as immune systems while others facilitate de-concatenation of circular chromosomes generated during DNA replication [30,33,36–39]. However, like PCR, isothermal amplification, or almost any other emerging technology, the pioneering publications that demonstrate proof-of-principle for CRISPR-dx and pAgo-dx were not immediately competitive with technologies that have been developed over decades (**Fig. 2.1**) [12,14,40,41]. Here, we provide an overview of emerging CRISPR-dx and pAgo-dx technologies and highlight new advances that couple these technologies with innovative detection methodologies that aim to increase bedside accessibility, multiplexable detection, and sensitivity. Finally, we suggest standardized reporting metrics that will facilitate

objective comparisons between diagnostic platforms.

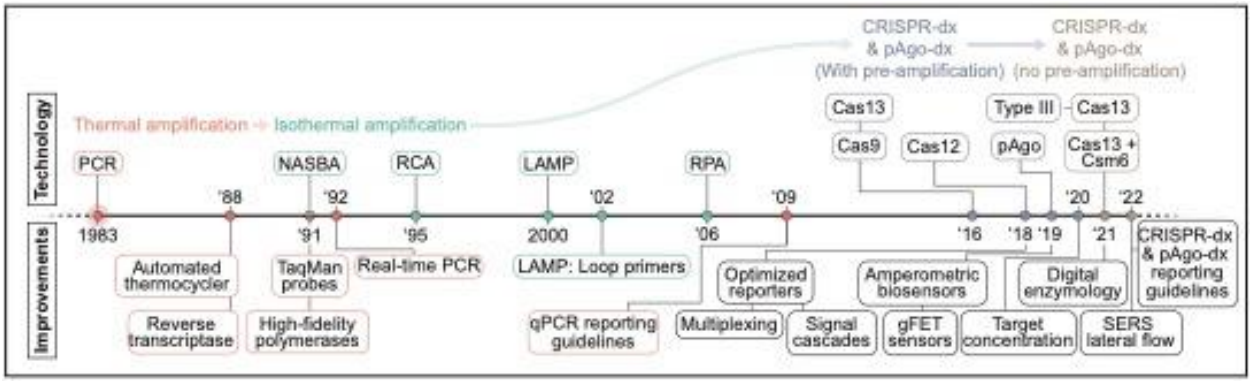


Figure 2.1 Evolution of molecular diagnostics. A decade of technological and methodological improvements to PCR resulted in a sensitive and specific amplification method that is the gold-standard of molecular diagnostics. However, guidelines that facilitated reproducible and reliably interpretable qPCR assays were established over a 16-year period. Isothermal amplification methods developed in the 90s and early 2000s eliminate the need for thermocyclers and reduce reaction times. While CRISPR-dx and pAgo-dx were initially used as secondary methods to confirm results from isothermal amplification, current research aims to combine CRISPR-dx or pAgo-dx, with novel detection methodologies, signaling cascades and protein circuits that do not rely on pre-amplification.

Next-Generation Programmable Diagnostics

Target-Activated Collateral Nucleases

All CRISPR systems rely on RNA guides for target recognition. In most CRISPR systems, the RNA delivers a CRISPR-associated (Cas) nuclease to complementary DNA targets. However, in 2016, Cas13 (formerly C2c2) proteins were shown to be RNA-guided RNases [40,41]. Cas13 cleaves complementary RNA targets, but target recognition also triggers Cas13 to cleave any RNA in solution, which is frequently referred to as "collateral cleavage" (**Fig. 2A**) [41]. In nature, collateral cleavage of cellular RNAs results in senescence of the host cell, which limits the spread of the viral infection [42–44]. Like most CRISPR systems, Cas13 can be easily reprogrammed to target any sequence by design. Therefore, East-Seletsky *et al.,* repurposed Cas13 for sequence-specific detection of RNA by using the target activated collateral nuclease activity to liberate a fluorophore from an RNA tethered quencher, which is a clever adaptation of hydrolysis probes (e.g., TaqMan^{TM}) that were originally designed to increase the specificity of quantitative PCR [40,45]. This initial work on Cas13-based diagnostics led to a burst of programmable diagnostics that all rely on the collateral cleavage of single-stranded RNA or single-stranded DNA by Cas13 or Cas12, respectively (**Fig. 2.2A**) [10–12,46,47]. Recently, TnpB proteins, the evolutionary precursors to Cas12, have been shown to be RNA-guided DNA-binding proteins with collateral nuclease activity [47,48]. Presumably, these nucleases will also be repurposed for molecular diagnostics.

The sensitivity of CRISPR-based diagnostics that rely on collateral cleavage of fluorescent nucleic acid reporters by target-activated Cas12 and Cas13 may be limited to 100 fM ($6x10^4$ copies/μL) of target RNA or DNA detectable within a 1-hour assay [49,50]. These

reactions are constrained by substrate affinity (K_M), reaction rate (k_{cat}), and the background fluorescent signal [49–51]. Several innovations have helped improve sensitivity including: i) pre-amplification of the target nucleic acid via isothermal amplification [10,11,54–57,12,14,18,20–22,52,53], ii) target concentration [22,58,59], iii) more sensitive collateral cleavage detection methods [60,61,70–72,62–69], and iv) design of protein circuits or signaling cascades [11,15,73–79]. Pre-amplification methods improve the sensitivity by six orders of magnitude and this work has been reviewed elsewhere [24–27]. Here we focus on the integration of CRISPR-Cas and Ago-based detection methods with technologies that enable rapid and reliable results. Further, we discuss the design of CRISPR signaling cascades and feedback amplification circuits that drastically increase sensitivity.

Digital Enzymology

Digital enzymology measures absolute quantities by encapsulating nucleic acids in discrete droplets (**Fig. 2.2B**). This approach has been used to enhance the sensitivity of Cas13-based collateral nuclease assays from 50 pM to 5 fM (a 10,000-fold improvement) and reduces the time to detection from 30 minutes to 5 minutes [66]. Separation of the sample into "micro-containers" results in a statistical distribution of the target nucleic acids among the droplets. This method offers three advantages over ensemble experiments [80,81]. First, the instrumentation for digital measurements must only distinguish between a high (>1) or low (0) signal, while analog measurements of ensemble reactions must accurately measure the magnitude. Second, a small reaction volume increases the effective concentration of the target molecule, which improves the sensitivity. Third, effective partitioning of the target excludes compounds that could inhibit the reaction (i.e., iron or immunoglobulin in blood samples) [82]. Digital enzymology combined

with Cas12- and Cas13-reliant collateral nuclease assays results in highly sensitive and rapid diagnostics that detect 17.5 copies/μL of DNA in 60 minutes (Cas12), or 3,400 copies/μL of RNA in 5 minutes (Cas13) [66–68].

Amperometric Biosensors

Amperometric sensors measure the current generated by oxidation or reduction reactions on a conductive surface (**Fig. 2.2C**) [83,84]. Amperometric biosensors are attractive because they are inexpensive to manufacture, simple to design, amenable to small volumes (0.1-5 μL), and compatible with complex samples (e.g., human serum). To detect collateral nuclease activity on an amperometric sensor, a redox-active molecule (e.g., methylene blue) is tethered to a gold electrode by a nucleic acid tether (e.g., ssDNA). An intact tether oxidizes the gold electrode (high current). Cas12-mediated detection of a complementary target activates the collateral nuclease activity, which cleaves the ssDNA tether, and liberates the methylene blue from the gold electrode (low current). The first generation of Cas12-based amperometric sensors detected $3x10^7$ copies/μL of target DNA in 60 minutes [64,65]. This approach has recently been improved by using a stem-loop DNA structure that brings the redox-active molecule closer to the gold electrode. This simple innovation boosts the signal and further increases the sensitivity to $1.8x10^7$ copies/μL [64]. Coincidentally, Cas13 has been recently reported to preferably cleave tRNA anticodon loops, halting protein translation during phage infection [44]. Therefore, the stem-loop structure of an RNA tether may optimally position the redox-active molecule near the gold electrode, while simultaneously mimicking the natural substrate for Cas13 cleavage. Future integration of scalable and disposable amperometric biosensors with electronic handheld devices may enable point-of-care detection of both RNA and DNA.

Surface-Enhanced Raman Scattering Analysis of Lateral Flow Assays

SERS (surface-enhanced Raman scattering) is a "surface-sensitive" spectroscopy technique that has been used to improve the sensitivity of lateral flow assays (**Fig. 2.2D**). In specific applications, SERS can detect single molecules [85,86]. While lateral flow assays are a low-cost, paper-based platform that provides a visible result within 5-30 minutes (e.g., home pregnancy test strips) [87], they are frequently less sensitive than other diagnostics (**Fig. 2D**) [87]. SERS enhances the sensitivity by coupling a Raman-active reporter molecule (5,5'-dithiobis(2-nitrobenzoic acid); DTNB) to a lateral flow strip that is read by a portable Raman spectrometer [88]. In this assay, the Raman-enhanced reporter is tethered to biotin by a ssDNA. Detection of a DNA target by Cas12, activates the collateral DNase activity, which cleaves the ssDNA tether, and allows the reporter to aggregate at a specific location on the strip. SERS-analyzed lateral flow assays are roughly 1,000-10,000 times more sensitive than a visual readout, and recent advancements enable SERS analysis in a portable and multiplex format [69,70].

CRISPR Feedback Amplification Circuits

The sensitivity of PCR and isothermal amplification techniques are based on exponential amplification of the nucleic acid target. In these assays, each round of DNA amplification doubles the amount of template for further replication, which leads to an exponential "chain reaction". Similarly, "feedback amplification" circuits of CRISPR enzymes increase sensitivity and transform Cas-based detection into an exponential process [73]. A Cas12-based feedback amplification circuit has been designed for sensitive detection of DNA without prior polymerase-based amplification (**Fig. 2.2E**). A complementary DNA target activates the Cas12-mediated collateral ssDNase activity, which leads to cleavage of the fluorescent DNA reporter in a DNA-

crRNA duplex, causing an increase in fluorescent signal (**Fig. 2E**). Cleavage of the DNA reporter releases the crRNA, which is designed to function as a guide for "empty" Cas12 enzymes. These Cas12 proteins bind to complementary dsDNA, which results in additional collateral cleavage, resulting in the release of more RNA guides and more signal (**Fig. 2E**). This single enzyme circuit results in 1,000,000-fold improved sensitivity and 5- to 10-fold improved specificity over direct Cas12-based detection [73].

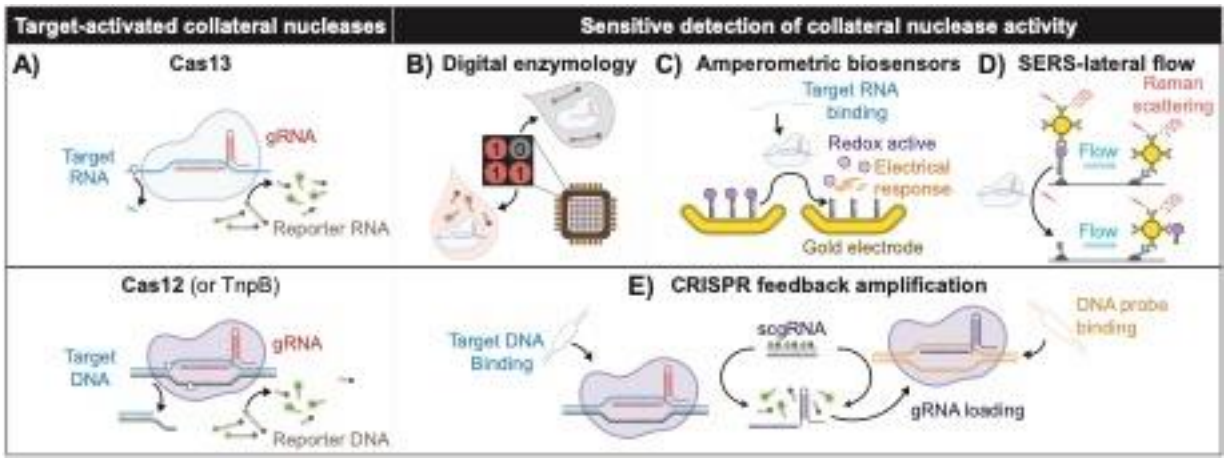


Figure 2.2. Sensitive nucleic acid detection using programmable target recognition and activated collateral nucleases. (A) RNA-guided target recognition by Cas13 (top) or Cas12 (bottom) allosterically activates a collateral nuclease activity, which cleaves a reporter RNA or DNA. Reporter nucleic acids can be labelled with different functional groups that result in a variety of readouts including fluorescent, colorimetric, electrical and Raman scattering. (B) Dilution of the diagnostic reaction into microchambers leads to a digital count (0 or 1) of target-containing chambers, enabling the absolute quantification of targets in the original sample and improving sensitivity 10,000-fold over bulk reactions. (C) Amperometric sensors detect the reduction or oxidation of a gold electrode, which results in a change in current. To detect collateral nuclease activity on an amperometric sensor, a redox-active molecule (purple sphere) is tethered to a gold electrode by a nucleic acid tether. Cleavage of the nucleic acid tether changes the current across an electrode. (D) Surface-enhanced Raman scattering (SERS) improves the sensitivity of lateral flow assays >1,000-fold over visual inspection alone. (E) Feedback amplification circuits of CRISPR enzymes can be designed to mimic the exponential “chain reaction” of DNA amplification that occurs during PCR, leading to a 1,000,000-fold increase in sensitivity.

Target-Activated Signal Amplification

In 2009, Hale *et al.* demonstrated that type III CRISPR systems rely on RNA-guided complexes that bind and cleave RNA targets [89]. This result implied that at least some type III CRISPR systems provide protection by directly cleaving the genome of RNA phages, degrading transcripts from DNA phages, or both. However, these *in vitro* experiments didn't jive with *in vivo* experiments that suggested a role for DNA targeting by these systems [90]. We now know that most type III CRISPR RNA-guided complexes (i.e., Csm or Cmr) cleave the complementary RNA in six nucleotide intervals [91–93], and that RNA binding activates a DNA nuclease in the Cas10 subunit of the complex [94]. In 2017, the versatility and complexity of these systems expanded again when two groups independently reported that the Cas10 subunits of these type III CRISPR complexes are also polymerases that generate cyclic oligonucleotides [95,96]. The cyclic nucleotides generated by the Cas10 polymerase serve as an alarm signal that is recognized by a diverse and growing group of "effectors" that function as RNases [95,96], DNases [97,98], peptidases [99], or a wide array of other downstream effector enzymes [100,101]. The discovery of CRISPR-mediated signaling cascades has inspired several new applications that rely on signaling pathways to improve CRISPR-dx applications [15,76–78].

The type III CRISPR immune signaling pathway has been repurposed for specific and sensitive detection of viral RNA [15,76–78]. RNA-bound Csm complex polymerizes ribo-ATP into cyclic oligoadenylate messengers, including cyclic tetra-adenylate (cA_4), which bind and activate the *Thermus thermophilus* Csm6 effector nuclease to cleave a fluorescent reporter (**Fig. 3A**). To increase the sensitivity of RNA detection, the backbone subunits (Csm3) of the Csm complex were mutated to prevent Csm3-mediated cleavage of the bound target RNA. This mutation preserves, rather than degrades, the RNA target and locks the polymerase into an

activated state. The initial use of these engineered complexes increased sensitivity 3-fold over the wildtype Csm complex [15]. Although the initial iteration of this approach was fast (1-30 minutes), it wasn't especially sensitive (10^7 copies/μL) [15]. Successive iterations have improved the sensitivity by replacing Csm6 with downstream nucleases (i.e., NucC, Can1 or Can2) that do not degrade the cyclic nucleotide activator [78,102,103], and by pairing effector nucleases (e.g., NucC) with type III complexes that predominantly generate the activating cyclic nucleotide (e.g., cA3) [98,101]. These advancements improved the sensitivity of Csm- or Cmr-based diagnostics to ~10^3 copies/μL in a 30-minute assay [76,78]. However, diagnostic parameters are often reported for purified RNA as the starting material, and exclude the time, costs or laboratory equipment typically required for extracting RNA from complex samples. A Histidine-tagged Csm complex from *Thermus thermophilus* was recently repurposed for capturing and concentrating target RNA directly from a patient sample using nickel-derivatized magnetic beads. This approach simultaneously reduces time, cost, sample handling requirements associated with RNA extraction, and the concentration step increases the sensitivity of this assay [103].

While current Csm- and Cmr-based diagnostics rely on cyclic oligoadenylate-activated effector nucleases and the readout of collateral nuclease activity, nucleases are a common contaminant in human samples, which can result in false-positives. The discovery of new enzymatic activities (i.e., peptidases, pore-forming toxins, etc.) activated by these and other CRISPR immune signaling molecules [99,100] could enable novel detection methodologies that are more robust.

CRISPR-Catalyzed ATP Polymerization

Incorporation of a nucleotide during DNA or RNA polymerization reactions coincides with the release of a single proton and a pyrophosphate molecule (**Fig. 2.3A**). The release of

these “byproducts” is the basis for label-free sequencing technologies by Ion Torrent and Pyrosequencing (454 sequencing) [104–106]. Similarly, byproducts of the polymerase have also been repurposed for visual readouts of PCR, LAMP, RPA, and type III CRISPR diagnostics [15,107,108]. Visual methods of detecting polymerization byproducts enable sensitive detection of molecular targets using LAMP (100 copies/μL) [16]. However, since type III CRISPR diagnostics lack the exponential amplification of target nucleic acid that occurs in PCR, LAMP, or RPA, the current sensitivities for visual detection of ATP polymerized by the Cas10 polymerase is relatively low (10^{10} copies/μL). Therefore, more sensitive detection methods must be applied to detect Csm or Cmr-generated polymerization byproducts at clinically relevant sensitivities.

Engineered CRISPR Signaling Cascades

The first example of an engineered CRISPR-based signaling cascade leveraged the collateral nuclease activity of target-bound Cas13. In this assay, the collateral cleavage activity of Cas13 generates RNA fragments that activate CRISPR effector nucleases, like Csm6 (**Fig. 2.3B**) [11]. Cas13 cleavage results in a 2’,3’-cyclic phosphate at the 3’ end of the RNA, and certain Cas13 variants preferentially cleave at uridine [11]. Therefore, Zhang and colleagues designed an RNA (i.e., 5’-AAAAUUUUU-3’) that is converted into a Csm6 activator (i.e., 5’-AAAA that ends with a 2’-,3’-cyclic phosphate) following Cas13 processing. Activated Csm6 cleaves a fluorescent reporter RNA, resulting in a 3.5-fold increase in sensitivity over Cas13 alone (**Fig. 3B**) [11]. Recently, this approach has been further enhanced by replacing specific hydroxyl groups in the RNA activator with Fluorine. This substitution prevents Csm6-mediated degradation of the activating ligand and improves sensitivity by 100-fold [74,102]. These

engineered CRISPR signaling cascades may be further improved by replacing Csm6 with a compatible nuclease that does not have ring-nuclease activity (i.e., NucC, Can1 or Can2). This change would eliminate the costs associated with the synthesis of fluorinated RNAs, and could improve the sensitivity by eliminating the negative impact that the 2'-fluoro modification has on nuclease activation [74,109].

An alternative signaling cascade relies on Cas13 collateral cleavage to snip a chimeric RNA-DNA reporter. The reporter is designed such that snipping of the RNA liberates a binding site for Cas12 in the DNA, and this activates the collateral cleavage activity of Cas12 (**Fig. 3C**) [75]. This results in a 1000-fold improvement over direct Cas13-based detection and the two amplification steps of this circuit result in a limit of detection of 800 copies/μL in 100 minutes [75]. Intriguingly, Cas13 may cooperate with natural downstream effector proteins to facilitate antiviral defense. For example, Cas13 has recently been shown to activate a CRISPR-associated pore-forming toxin (Csx28) that disrupts the integrity of the inner membrane, possibly stimulating an abortive infection. This observation suggests that future Cas13-based RNA target detection applications may be repurposed for the depolarization of a biological membrane that results in electrochemical readouts [110].

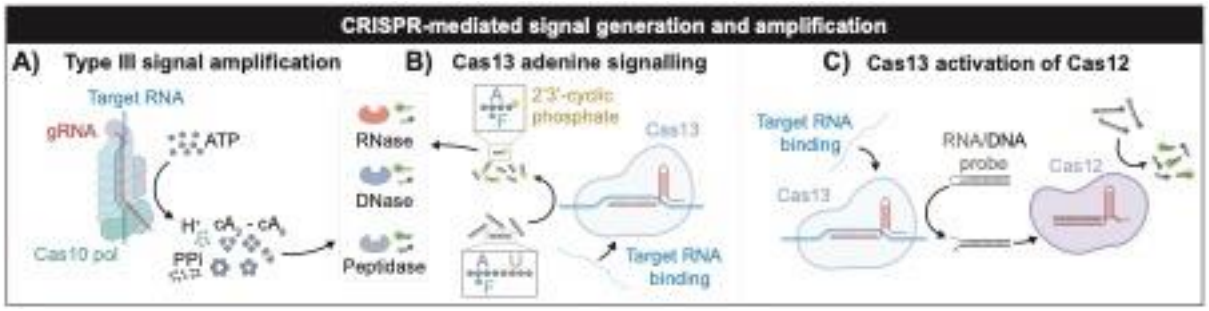


Figure 2.3. CRISPR-mediated signalling inspires new CRISPR-dx applications. (A) Type III (i.e., Cmr and Csm) CRISPR complexes are used for direct detection of RNA. Target RNA binding allosterically activates a Cas10 polymerase subunit (Cas10 pol) in Csm or Cmr that produces a mixture of cyclic oligoadenylate molecules (cA2 – cA6). Cyclic oligoadenylate molecules bind and activate CRISPR associated effectors including RNases, DNases, and peptidases, which can be repurposed to cleave RNA, DNA, peptides, or proenzyme reporters. (B) The collateral nuclease activity of Cas13 produces 2'3'-cyclic phosphates at the 3' ends of cleaved RNAs. Cas13-mediated cleavage of a specially designed reporter RNA thereby produces an artificial linear ligand that activates CRISPR-associated effectors, resulting in >100-fold increase in sensitivity. (C) Cas13-mediated nicking of an inhibiting RNA loop makes an RNA-DNA hybrid duplex accessible for Cas12 RNA-guided DNA binding. One Cas13 thereby activates many Cas12 complexes, which in turn cleave many fluorescent reporter DNAs, increasing the sensitivity 1000-fold over direct Cas13-based detection.

Target-Guided Nucleases

While most of the attention around Cas9 has focused on applications for RNA-guided genetic engineering, Cas9 has also been used for diagnostics [14,20,52,71,111,112]. In 2016, Collins and colleagues used Cas9 to detect RNA from the Zika virus by first converting the viral RNA to DNA using reverse transcriptase and isothermally amplifying the DNA using NASBA [14]. However, Cas9 is a single turnover nuclease, which limits the sensitivity of detection methods that rely on dsDNA cleavage. A new innovation has transformed Cas9 into a versatile diagnostic that can be multiplexed for the identification of many different RNA targets [17]. In this application, the tracrRNA (trans-activating CRISPR RNA), which is required for CRISPR RNA processing [113], has been reprogrammed (reprogrammed tracrRNAs, Rptrs), to be complementary to a sequence on the target RNA of interest. The Rptr thereby "coerces" the target RNA to act as a crRNA for Cas9, which then cleaves a complementary reporter (**Fig. 2.4A**) [17]. The multiplexing potential of this approach is a major advantage, however current applications rely on a pre-amplification step to make up for the single-turnover nature of Cas9 cleavage. Further, these Cas9-based diagnostic reactions must be analyzed using electrophoresis. However, coupling this Cas9-based diagnostic to a microarray-based detection methodology could improve this platform to detect millions of different RNA targets in the same reaction [17,114]. Further, single-molecule imaging of fluorescently labeled Cas9 may circumvent the current requirement for a pre-amplification step. Similar applications may pertain to the evolutionary precursors of Cas9 (i.e., TnpB and IscB) that have recently been reported [47,48,115].

Like CRISPR-Cas systems, prokaryotic Argonaute (pAgo) proteins can be repurposed to act as target-guided nucleases that detect DNA. In 2017, the Argonaute protein from *Thermus thermophilus* was shown to "chop" dsDNA in a guide-independent manner and subsequently load cleaved DNA products as gDNAs (guide DNAs). As a diagnostic, short gDNAs are used to program pAgo for cleavage of a DNA target, generating two target-derived fragments. One of these fragments has a 5' monophosphate, which is required for loading into an unguided (i.e., apo) pAgo protein. This target-derived guide, then directs the pAgo to cleave a complementary fluorescent reporter DNA [18,19]. DNA cleavage by pAgo is a multi-turnover process (**Fig. 2.4B**) [116,117], which suggests that the implementation of signaling cascades or sensitive nuclease detection methods could improve pAgo-dx sensitivities. Finally, pAgos are often associated with, or directly fused to diverse enzymatic effectors, some of which have been shown to deplete cellular NAD+ (Nicotinamide adenine dinucleotide) (i.e., Sir2, or TIR) or to depolarize the bacterial inner membrane (i.e., Aga2) [28,34,35]. Therefore, pAgo-dx may eventually be purposed for detection methodologies that do not rely on nuclease activity [34]. In fact, a pAgo-dx that relies on $NAD(P)^+$-ase activity was recently reported [36]. The pAgo-TIR/APAZ heterodimer from *Crenotalea thermophila* forms a tetramer of homodimers upon binding an RNA guide [36]. RNA-guided binding of pAgo-TIR/APAZ to a ssDNA target activates the $NAD(P)^+$-ase activity of the TIR domain to degrade $NAD(P)^+$ into ADPR(P) (adenosine diphosphate ribose) and NAM (nicotinamide). Therefore, ssDNA can be detected with high specificity by pAgo-TIR/APAZ-mediated degradation of ε-NAD^+ to the fluorescent molecule ε-ADPR [36]. While the sensitivity of this approach is currently modest (~10^{10}

copies/μL) [36], identification of enzyme variants that are naturally more active may improve sensitivity.

Target Binding

Graphene Field Effect Transistor Sensing of DNA

A FET (Field-Effect Transistor) is a transistor that uses an electric field to control the flow of current in a semiconductor and is composed of three electrodes (source, drain, and gate). A voltage applied to the gate alters the conductivity between the drain and source. In graphene FETs (gFETs), the channel between the source and drain electrodes is graphene, which is an excellent electrical conductor with an extraordinarily high surface-to-volume ratio (**Fig. 4C**). This material is highly sensitive to changes in electric fields caused by any molecule attached to the surface [118]. These properties have been repurposed for the sensitive detection of Cas9 binding to target DNAs in a method that has been termed "CRISPR-CHIP" [71,72]. In this application, a graphene layer is functionalized with catalytically dead Cas9 (dCas9) and sequence-specific binding of dCas9 to a complementary dsDNA target causes large changes in the electric field across the gFET channel (**Fig. 2.4C**) [71,72]. These assays are capable of detecting 1000 copies/μL in 15 minutes [71,72]. Further, extending the assay time to 1 hour allows for highly specific detection that discriminates between SNPs [71,72]. gFET-based sensors enable rapid, sensitive, specific, and low-cost detection via an electronic readout [118]. In addition, the multiplexing of gFET sensors is expected to facilitate the testing of up to thousands of targets on a single small device. These properties are expected to enable highly multiplexed, amplification-free detection of DNA-containing infectious agents or human alleles at point-of-care.

Reconstitution of Split Reporter Enzyme

Since Richards' 1958 observation that two proteolytic fragments of bovine pancreatic ribonuclease could interact *in trans* to reconstitute a functional enzyme [119], the reconstitution of "split reporter enzymes" have been used to visualize molecular interactions [120]. Current iterations of this approach rely on tethering two non-functional halves of the reporter to two other proteins that are anticipated to interact [120]. Using this strategy, two halves of a split luciferase can be linked to Cas9s that are programmed to bind complementary target sequences that are adjacent to one another (**Fig. 2.4D**) [79]. However, the nucleotide spacing between Cas9s may require optimization since the efficient reconstitution of the split reporter protein is likely to depend on the helical phase of DNA [121,122].

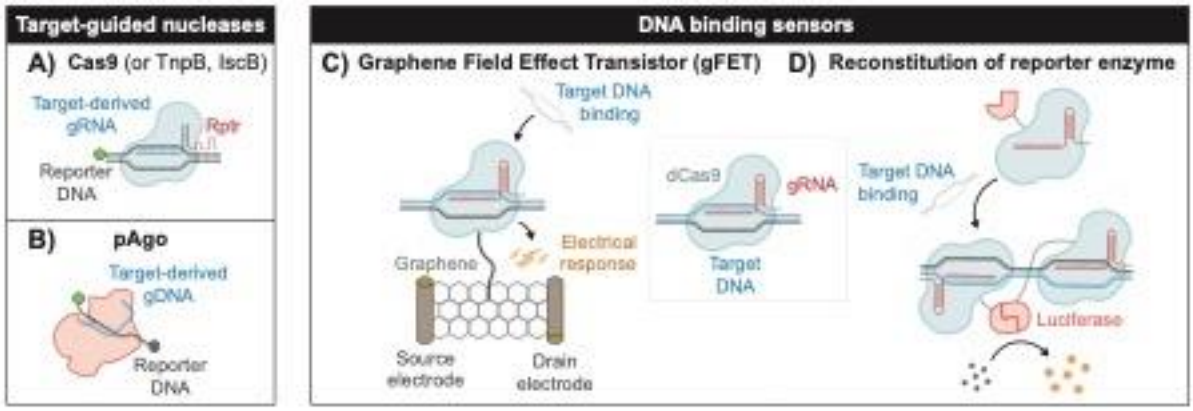


Figure 2.4. Target-guided nucleases and sensitive methods to detect target binding. (A) Cas9 can be repurposed to detect RNA in a multiplexed protocol, by using reprogrammed tracrRNAs (Rptrs) that anneal to the target RNA and coerce the adjacent RNA sequence to behave as a typical RNA guide that results in cleavage of a complementary DNA reporter. The presence of many different RNA targets is determined using gel electrophoresis or a Bioanalyzer. Future microarray-based detection methods may enable multiplexed detection of millions of RNAs. (B) DNA-guided prokaryotic Argonautes cleave dsDNA targets into fragments that are recognized as new gDNAs (guide DNAs), which direct pAgos to cleave complementary DNA reporter. Unlike target-guided Cas9, target-guided Argonautes are multi-turnover enzymes that cleave many reporter DNAs per target-derived guide. (C) RNA-guided Cas9 can be repurposed to detect complementary dsDNA target. Cas9-bound dsDNA induces a large change in the electrical properties of the tethered graphene channel, resulting in rapid and sensitive electrical detection of DNA. (D) Two RNA-guided Cas9s, each fused to different halves of a split reporter enzyme can be repurposed to detect dsDNA. Stable association of each Cas9 to the adjacent complementary sites results in the reconstitution of the reporter enzyme (e.g., Luciferase).

Suggested Reporting Metrics for Molecular Diagnostics

PCR matured into its current status as a gold-standard diagnostic technique through the formulation of guidelines that standardized how to conduct, analyze and report results from quantitative PCR (qPCR) [123,124]. This initiative for standardization was prompted by a technically flawed study that claimed to detect measles viral RNA in children with autism [125], that could not be reproduced [126]. In the above section of this review, we compare the sensitivities and specificities of CRISPR- and pAgo-dx platforms that have been designed to detect a wide variety of RNA and DNA targets. However, an objective comparison between many of these platforms is not possible due to a lack of standardized reporting guidelines. The concentration of target nucleic acid detected by qPCR is commonly related by cycle threshold (Ct) or quantification cycle (Cq) number, which is the cycle number at which the fluorescence reading for a reaction exceeds the background fluorescence threshold. Conversely, CRISPR-dx and pAgo-dx utilize diverse biochemical activities that are measured via diverse methodologies, resulting in diverse signals including fluorescence, visible wavelengths of light, electrical current, and the inelastic scattering of photons. Therefore, no single experiment-specific metric can be used to compare these emerging diagnostic assays, further highlighting the need to standardize the reporting of the limit of detection, repeatability within an experiment, reproducibility between experiments, clinical sensitivity, and diagnostic specificity (Table 2.1, Table 2.S1). These minimal diagnostic metrics are informed by previous guidelines established during the maturation of qPCR as the gold-standard diagnostic, and by FDA's EUA (Food and Drug Administration's Emergency Use Authorization) guidelines for molecular diagnostics during the COVID-19 pandemic [124,127].

Table S1. Recommended Information for publication of CRISPR-dx and pAgo-dx Assays

Item to check
Protein purification (or manufacturer/source if purchased/obtained somewhere else)
Detailed description of protein expression (expressed gene, affinity tag, plasmid backbone, expression strain, inductor concetration, growth time and temperature, etc.)
Lysis conditions
Protocol purification protocol (columns, buffers, chromotography systems)
Gel filtration profiles with collected fractions shown, SDS-PAGE of final protein prep
RNA gels showing guide RNA size co-purified with the proteins
Sample
Description of how samples were acquired
For viral diagnostics: sample type (saliva, nasopharyngeal swab, etc.), composition or manufacturer of viral transport media (if used)
For viral diagnostics: how samples were handled (biosafety measures, was sample inactivated?)
For IVT standards: manufacturer or detailed description of template DNA, buffers and kits used for transcription reactions and product clean-up. How was concentration of standard quantified?
How was target concentration determined in the sample/standard? If qPCR was used, provide details on qPCR conditions, include standard curve (see MIQE guidelines)
Sample storage conditions and duration
Nucleic acid extraction
Volume of sample processed
Elution volume
Procedure and/or instrumentation
Name of kit and details of any modifications
Source of additional reagents used
Pre-amplification (if any)
Complete reaction conditions
Amount of input RNA/DNA and reaction volume
Oligonucleotide primer sequences and concentrations
Enzymes used and buffer composition
Temperature and time
Manufacturer of reagents and catalogue numbers
Target and guide RNA information
Target gene / transcript / sequence with accession number (if available)
Guide sequence and length
Algorithms or source for guide design
In silico specificity screen for guide RNA (BLAST, and so on)
Location of targeted site
CRISPR-dx or pAgo-dx protocol
Complete reaction conditions
Reaction volume and amount of input sample

Outlook

It took nine years for PCR to be developed from an error-prone, insensitive, and time-consuming proof-of-concept into a reliable diagnostic [1,2,45,128], and a total of 16 years to establish guidelines that enable reliable interpretation [124]. In the past six years, CRISPR-dx and Ago-dx have undergone rapid development into an emerging class of molecular diagnostics [18,40]. Coupling CRISPR-dx and Ago-dx to pre-amplification methods have improved their sensitivities by six orders of magnitude to detection limits comparable to RT-qPCR, and fast-tracked their use as molecular diagnostics under the FDA's Emergency Use Authorization (Fig. 1) [129,130]. However, pre-amplification requires primer design, and optimization, increases the cost, and generally increases reaction times or sample handling. Therefore the next generation of CRISPR- and pAgo diagnostics will rely on the discovery and incorporation of new enzyme activities, development of target concentration methods, sensitive detection methodologies, and protein circuits or signaling cascades that all improve sensitivity, specificity and time-to-result (Fig. 2, 3, and 4) [100,131–133]. Several of these innovations have already led to the development of rapid (<60 mins) CRISPR- and pAgo-dx that do not rely on pre-amplification (Fig. 5). Efforts to increase the sensitivity of these methods, and to bypass or obviate the need for RNA extraction, are ongoing (Fig. 5) [22,58,59,67,103]. Machine learning-assisted design of enzymes with improved kinetics, substrate affinity, or selection of more specific and active RNA guides, may facilitate these efforts [134–137]. The continued maturation of these molecular diagnostic tests will assist a transition to a distributed-testing paradigm. The centralized-testing paradigm, which requires shipping samples to a high-complexity lab for RT-qPCR, created a critical bottleneck in SARS-CoV-2 surveillance. Therefore, a distributed-testing model that

enables individuals to self-perform and analyze results will prevent similar bottlenecks in the future. An estimated 6.64 billion people, or ~84% of the world's population, own a smartphone that includes a high-resolution camera [138]. The SARS-CoV-2 pandemic has bolstered efforts to enhance smartphone-based detection and interpretation of test results, machine learning-assisted interpretation of the results, and cloud-based computing are anticipated to empower point-of-care applications [22,60,139–141]. Collectively, we anticipate that continued innovation, driven by demands for faster diagnostics that can be administered at the point-of-care, will elevate CRISPR-dx and Ago-dx to a new standard in molecular diagnostics.

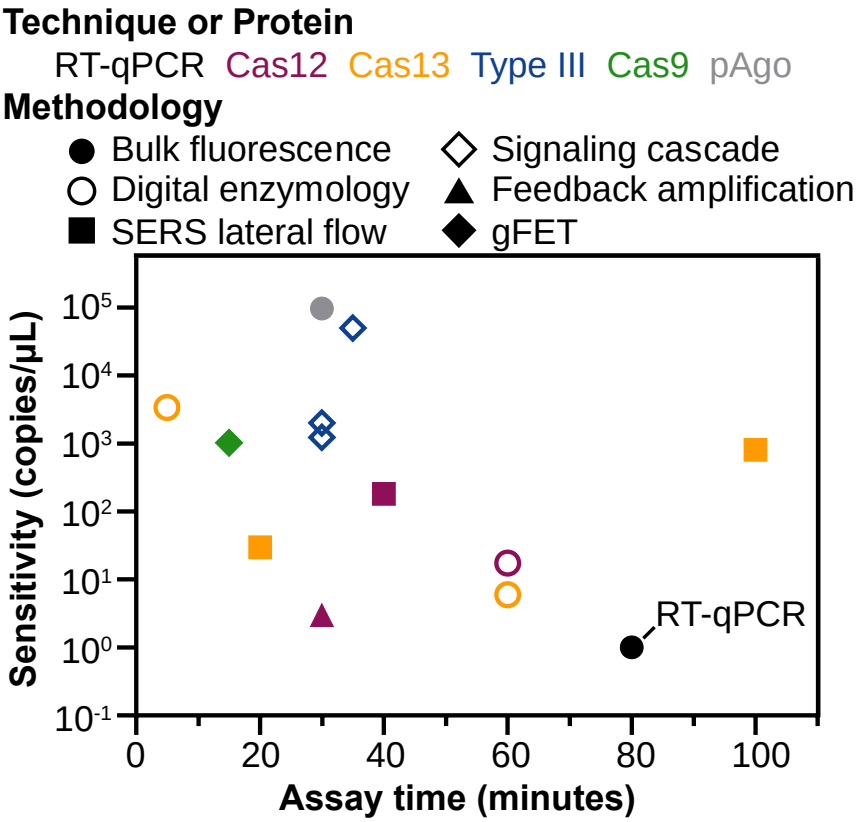


Figure 5. The emerging landscape of pre-amplification-independent CRISPR-dx and pAgo-dx. Unique and sensitive Cas12- , Cas13- , Type III CRISPR- , Cas9- and pAgo-based diagnostic methodologies that do not rely on preamplification compared to the CDC's recommended RT-qPCR protocol for SARS-CoV-2 [18,66,78,142,67–69,71,73–76].

Manuscript References

1. Mullis KB. The Polymerase Chain Reaction (Nobel Lecture). Angew Chemie Int Ed English ;33:1209–13. https://doi.org/10.1002/anie.199412091.

2. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, et al. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985;230:1350–4. https://doi.org/10.1126/science.2999980.

3. Mullis KB, Erlich HA, Arnheim N, Horn GT, Saiki RK, Scharf SJ. PROCESS FOR AMPLIFYING, DETECTING, AND/OR-CLONING NUCLEIC ACID SEQUENCES. 828144, 1987.

4. Gill P, Ghaemi A. Nucleic Acid Isothermal Amplification Technologies—A Review. Nucleosides, Nucleotides and Nucleic Acids 2008;27:224–43. https://doi.org/10.1080/15257770701845204.

5. Deng R, Zhang K, Li J. Isothermal Amplification for MicroRNA Detection: From the Test Tube to the Cell. Acc Chem Res 2017;50:1059–68. https://doi.org/10.1021/acs.accounts.7b00040.

6. Ptasinska A, Whalley C, Bosworth A, Poxon C, Bryer C, Machin N, et al. Diagnostic accuracy of loop-mediated isothermal amplification coupled to nanopore sequencing (LamPORE) for the detection of SARS-CoV-2 infection at scale in symptomatic and asymptomatic populations. Clin Microbiol Infect 2021;27:1348.e1-1348.e7. https://doi.org/10.1016/j.cmi.2021.04.008.

7. Peto L, Rodger G, Carter DP, Osman KL, Yavuz M, Johnson K, et al. Diagnosis of SARS-CoV-2 infection with LamPORE, a high-throughput platform combining loop-mediated isothermal amplification and nanopore sequencing. J Clin Microbiol 2021;59. https://doi.org/10.1128/JCM.03271-20.

8. Stüder F, Petit JL, Engelen S, Mendoza-Parra MA. Real-time SARS-CoV-2 diagnostic and variants tracking over multiple candidates using nanopore DNA sequencing. Sci Rep 2021;11:1–11. https://doi.org/10.1038/s41598-021-95563-w.

9. Ludwig KU, Schmithausen RM, Li D, Jacobs ML, Hollstein R, Blumenstock K, et al. LAMP-Seq enables sensitive, multiplexed COVID-19 diagnostics using molecular barcoding. Nat Biotechnol 2021;39:1556–62. https://doi.org/10.1038/s41587-021-00966-9.

10. Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 2017;356:438–42. https://doi.org/10.1126/science.aam9321.

11. Gootenberg JS, Abudayyeh OO, Kellner MJ, Joung J, Collins JJ, Zhang F. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. Science 2018;360:439–44. https://doi.org/10.1126/science.aaq0179.

12. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 2018;360:436–9. https://doi.org/10.1126/science.aar6245.

13. Broughton JP, Deng X, Yu G, Fasching CL, Servellita V, Singh J, et al. CRISPR–Cas12-based detection of SARS-CoV-2. Nat Biotechnol 2020. https://doi.org/10.1038/s41587-020-0513-4.

14. Pardee K, Green AA, Takahashi MK, Braff D, Lambert G, Lee JW, et al. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. Cell 2016;165:1255–66. https://doi.org/10.1016/j.cell.2016.04.059.

15. Santiago-Frangos A, Hall LN, Nemudraia A, Nemudryi A, Krishna P, Wiegand T, et al. Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. Cell Reports Med 2021;2:100319. https://doi.org/10.1016/j.xcrm.2021.100319.

16. Dao Thi VL, Herbst K, Boerner K, Meurer M, Kremer LP, Kirrmaier D, et al. A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. Sci Transl Med 2020;7075:1–20. https://doi.org/10.1126/scitranslmed.abc7075.

17. Jiao C, Sharma S, Dugar G, Peeck NL, Bischler T, Wimmer F, et al. Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. Science 2021;7106:eabe7106. https://doi.org/10.1126/science.abe7106.

18. He R, Wang L, Wang F, Li W, Liu Y, Li A, et al. Pyrococcus furiosus Argonaute-mediated nucleic acid detection. Chem Commun 2019;55:13219–22. https://doi.org/10.1039/c9cc07339f.

19. Wang F, Yang J, He R, Yu X, Chen S, Liu Y, et al. PfAgo-based detection of SARS-CoV-2. Biosens Bioelectron 2021;177:112932. https://doi.org/10.1016/j.bios.2020.112932.

20. Zhou W, Hu L, Ying L, Zhao Z, Chu PK, Yu XF. A CRISPR–Cas9-triggered strand displacement amplification method for ultrasensitive DNA detection. Nat Commun 2018;9:1–11. https://doi.org/10.1038/s41467-018-07324-5.

21. Li L, Li S, Wu N, Wu J, Wang G, Zhao G, et al. HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. ACS Synth Biol 2019;8:2228–37. https://doi.org/10.1021/acssynbio.9b00209.

22. de Puig H, Lee RA, Najjar D, Tan X, Soenksen LR, Angenent-Mari NM, et al. Minimally

instrumented SHERLOCK (miSHERLOCK) for CRISPR-based point-of-care diagnosis of SARS-CoV-2 and emerging variants. Sci Adv 2021;7:23–6. https://doi.org/10.1126/sciadv.abh2944.

23. Abudayyeh OO, Gootenberg JS. CRISPR diagnostics. Science 2021;372:914–5. https://doi.org/10.1126/science.abi9335.

24. Kaminski MM, Abudayyeh OO, Gootenberg JS, Zhang F, Collins JJ. CRISPR-based diagnostics. Nat Biomed Eng 2021;5:643–56. https://doi.org/10.1038/s41551-021-00760-7.

25. Kim S, Ji S, Koh HR. Crispr as a diagnostic tool. Biomolecules 2021;11. https://doi.org/10.3390/biom11081162.

26. Kostyusheva A, Brezgin S, Babin Y, Vasilyeva I, Glebe D, Kostyushev D, et al. CRISPR-Cas systems for diagnosing infectious diseases. Methods 2021. https://doi.org/10.1016/j.ymeth.2021.04.007.

27. Nguyen PQ, Soenksen LR, Donghia NM, Angenent-Mari NM, de Puig H, Huang A, et al. Wearable materials with embedded synthetic biology sensors for biomolecule detection. Nat Biotechnol 2021. https://doi.org/10.1038/s41587-021-00950-3.

28. Makarova KS, Wolf YI, van der Oost J, Koonin E V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct 2009;4:29. https://doi.org/10.1186/1745-6150-4-29.

29. Hegge JW, Swarts DC, Van Der Oost J. Prokaryotic argonaute proteins: Novel genome-editing tools? Nat Rev Microbiol 2018;16:5–11. https://doi.org/10.1038/nrmicro.2017.73.

30. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. Mol Cell 2013;51:594–605. https://doi.org/10.1016/j.molcel.2013.08.014.

31. Miyoshi T, Ito K, Murakami R, Uchiumi T. Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. Nat Commun 2016;7. https://doi.org/10.1038/ncomms11846.

32. Swarts DC, Szczepaniak M, Sheng G, Chandradoss SD, Zhu Y, Timmers EM, et al. Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. Mol Cell 2017;65:985-998.e6. https://doi.org/10.1016/j.molcel.2017.01.033.

33. Swarts DC, Jore MM, Westra ER, Zhu Y, Janssen JH, Snijders AP, et al. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 2014;507:258–61. https://doi.org/10.1038/nature12971.

34. Zeng Z, Chen Y, Pinilla-Redondo R, Shah SA, Zhao F, Wang C, et al. A short prokaryotic argonaute cooperates with membrane effector to confer antiviral defense. BioRxiv 2021. https://doi.org/https://doi.org/10.1101/2021.12.09.471704.

35. Zaremba M, Dakineviciene D, Golovinas E, Stankunas E, Lopatina A, Sorek R, et al. Sir2-domain associated short prokaryotic Argonautes provide defence against invading mobile genetic elements through NAD+ depletion. BioRxiv 2021. https://doi.org/https://doi.org/10.1101/2021.12.14.472599.

36. Koopal B, Potocnik A, Mutte SK, Aparicio-Maldonado C, Lindhoud S, Vervoort JJM, et al. Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. Cell 2022:1–16. https://doi.org/10.1016/j.cell.2022.03.012.

37. Kuzmenko A, Oguienko A, Esyunina D, Yudin D, Petrova M, Kudinova A, et al. DNA targeting and interference by a bacterial Argonaute nuclease. Nature 2020;587:632–7. https://doi.org/10.1038/s41586-020-2605-1.

38. Jolly SM, Gainetdinov I, Jouravleva K, Zhang H, Strittmatter L, Bailey SM, et al. Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. Cell 2020;182:1545-1559.e18. https://doi.org/10.1016/j.cell.2020.07.036.

39. Lisitskaya L, Aravin AA, Kulbachinskiy A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. Nat Commun 2018;9:1–12. https://doi.org/10.1038/s41467-018-07449-7.

40. East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JHD, Tjian R, et al. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature 2016;538:270–3. https://doi.org/10.1038/nature19802.

41. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 2016;353. https://doi.org/10.1126/science.aaf5573.

42. Lopatina A, Tal N, Sorek R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. Annu Rev Virol 2020;7:371–84. https://doi.org/10.1146/annurev-virology-011620-040628.

43. Rostøl JT, Marraffini LA. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. Nat Microbiol 2019;4:656–62. https://doi.org/10.1038/s41564-018-0353-x.

44. Jain I, Kolesnik M, Minakhin L, Morozova N, Kirillov A, Medvedeva S, et al. tRNA anticodon cleavage by target-activated CRISPR-Cas13a effector 2021. https://doi.org/https://doi.org/10.1101/2021.11.10.468108.

45. Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci 1991;88:7276–80. https://doi.org/10.1073/pnas.88.16.7276.

46. Li S-Y, Cheng Q-X, Wang J-M, Li X-Y, Zhang Z-L, Gao S, et al. CRISPR-Cas12a-assisted nucleic acid detection. Cell Discov 2018;4:20. https://doi.org/10.1038/s41421-018-0028-z.

47. Altae-Tran H, Kannan S, Demircioglu FE, Oshiro R, Nety SP, McKay LJ, et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. Science 2021;374:57–65. https://doi.org/10.1126/science.abj6856.

48. Knott GJ, Lapinaite A. CRISPR Gets Its Origin Story. Cris J 2021;4:631–3. https://doi.org/10.1089/crispr.2021.29137.knl.

49. Ramachandran A, Santiago JG. CRISPR Enzyme Kinetics for Molecular Diagnostics. Anal Chem 2021;93:7456–64. https://doi.org/10.1021/acs.analchem.1c00525.

50. Kellner MJ, Koob JG, Gootenberg JS, Abudayyeh OO, Zhang F. SHERLOCK: nucleic acid detection with CRISPR nucleases. Nat Protoc 2019;14:2986–3012. https://doi.org/10.1038/s41596-019-0210-2.

51. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, et al. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. Biochemistry 2011;50:4402–10. https://doi.org/10.1021/bi2002289.

52. Huang M, Zhou X, Wang H, Xing D. Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 Triggered Isothermal Amplification for Site-Specific Nucleic Acid Detection. Anal Chem 2018;90:2193–200. https://doi.org/10.1021/acs.analchem.7b04542.

53. Wang T, Liu O, Sun H-H, Yin B-C, Ye B-C. Analytical Methods An RNA-Guided Cas9 Nickase-Based Method for Universal Isothermal DNAAmplification n.d. https://doi.org/10.1002/ange.201901292.

54. Shen J, Zhou X, Shan Y, Yue H, Huang R, Hu J, et al. Sensitive detection of a bacterial pathogen using allosteric probe-initiated catalysis and CRISPR-Cas13a amplification reaction. Nat Commun 2020;11:1–10. https://doi.org/10.1038/s41467-019-14135-9.

55. Notomi T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, et al. Loop-mediated isothermal amplification of DNA. Nucleic Acids Res 2000;28:E63. https://doi.org/10.1093/nar/28.12.e63.

56. Piepenburg O, Williams CH, Stemple DL, Armes NA. DNA detection using recombination proteins. PLoS Biol 2006;4:1115–21.

https://doi.org/10.1371/journal.pbio.0040204.

57. Ackerman CM, Myhrvold C, Thakku SG, Freije CA, Metsky HC, Yang DK, et al. Massively multiplexed nucleic acid detection with Cas13. Nature 2020;582:277–82. https://doi.org/10.1038/s41586-020-2279-8.

58. Joung J, Ladha A, Saito M, Kim N-G, Woolley AE, Segel M, et al. Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. N Engl J Med 2020;382:NEJMc2026172. https://doi.org/10.1056/NEJMc2026172.

59. Ramachandran A, Huyke DA, Sharma E, Sahoo MK, Huang C, Banaei N, et al. Electric field-driven microfluidics for rapid CRISPR-based diagnostics and its application to detection of SARS-CoV-2. Proc Natl Acad Sci U S A 2020;117:29518–25. https://doi.org/10.1073/pnas.2010254117.

60. Fozouni P, Son S, Díaz de León Derby M, Knott GJ, Gray CN, D'Ambrosio M V., et al. Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. Cell 2020:1–11. https://doi.org/10.1016/j.cell.2020.12.001.

61. Ferrario A, Scaramuzza M, Pasqualotto E, De Toni A, Paccagnella A. Development of a Disposable Gold Electrodes-Based Sensor for Electrochemical Measurements of cDNA Hybridization. Procedia Chem 2012;6:36–45. https://doi.org/10.1016/j.proche.2012.10.128.

62. Bruch R, Baaske J, Chatelle C, Meirich M, Madlener S, Weber W, et al. CRISPR/Cas13a-Powered Electrochemical Microfluidic Biosensor for Nucleic Acid Amplification-Free miRNA Diagnostics. Adv Mater 2019;31. https://doi.org/10.1002/adma.201905311.

63. Bruch R, Johnston M, Kling A, Mattmüller T, Baaske J, Partel S, et al. CRISPR-powered electrochemical microfluidic multiplexed biosensor for target amplification-free miRNA diagnostics. Biosens Bioelectron 2021;177. https://doi.org/10.1016/j.bios.2020.112887.

64. Zhang D, Yan Y, Que H, Yang T, Cheng X, Ding S, et al. CRISPR/Cas12a-Mediated Interfacial Cleaving of Hairpin DNA Reporter for Electrochemical Nucleic Acid Sensing. ACS Sensors 2020;5:557–62. https://doi.org/10.1021/acssensors.9b02461.

65. Dai Y, Somoza RA, Wang L, Welter JF, Li Y, Caplan AI, et al. Exploring the Trans-Cleavage Activity of CRISPR-Cas12a (cpf1) for the Development of a Universal Electrochemical Biosensor. Angew Chemie - Int Ed 2019;58:17399–405. https://doi.org/10.1002/anie.201910772.

66. Tian T, Shu B, Jiang Y, Ye M, Liu L, Guo Z, et al. An Ultralocalized Cas13a Assay Enables Universal and Nucleic Acid Amplification-Free Single-Molecule RNA Diagnostics. ACS Nano 2021;15:1167–78. https://doi.org/10.1021/acsnano.0c08165.

67. Shinoda H, Taguchi Y, Nakagawa R, Makino A, Okazaki S, Nakano M, et al. Amplification-free RNA detection with CRISPR–Cas13. Commun Biol 2021;4:476. https://doi.org/10.1038/s42003-021-02001-8.

68. Yue H, Shu B, Tian T, Xiong E, Huang M, Zhu D, et al. Droplet Cas12a Assay Enables DNA Quantification from Unamplified Samples at the Single-Molecule Level. Nano Lett 2021;21:4643–53. https://doi.org/10.1021/acs.nanolett.1c00715.

69. Pang Y, Li Q, Wang C, zhen S, Sun Z, Xiao R. CRISPR-cas12a mediated SERS lateral flow assay for amplification-free detection of double-stranded DNA and single-base mutation. Chem Eng J 2022;429:132109. https://doi.org/10.1016/j.cej.2021.132109.

70. Xiao R, Lu L, Rong Z, Wang C, Peng Y, Wang F, et al. Portable and multiplexed lateral flow immunoassay reader based on SERS for highly sensitive point-of-care testing. Biosens Bioelectron 2020;168:112524. https://doi.org/10.1016/j.bios.2020.112524.

71. Hajian R, Balderston S, Tran T, deBoer T, Etienne J, Sandhu M, et al. Detection of unamplified target genes via CRISPR–Cas9 immobilized on a graphene field-effect transistor. Nat Biomed Eng 2019;3:427–37. https://doi.org/10.1038/s41551-019-0371-x.

72. Balderston S, Taulbee JJ, Celaya E, Fung K, Jiao A, Smith K, et al. Discrimination of single-point mutations in unamplified genomic DNA via Cas9 immobilized on a graphene field-effect transistor. Nat Biomed Eng 2021;5:713–25. https://doi.org/10.1038/s41551-021-00706-z.

73. Shi K, Xie S, Tian R, Wang S, Lu Q, Gao D, et al. A CRISPR-Cas autocatalysis-driven feedback amplification network for supersensitive DNA diagnostics. Sci Adv 2021;7:1–10. https://doi.org/10.1126/sciadv.abc7802.

74. Liu TY, Knott GJ, Smock DCJ, Desmarais JJ, Son S, Bhuiya A, et al. Accelerated RNA detection using tandem CRISPR nucleases. Nat Chem Biol 2021;17:982–8. https://doi.org/10.1038/s41589-021-00842-2.

75. Sha Y, Huang R, Huang M, Yue H, Shan Y, Hu J, et al. Cascade CRISPR/cas enables amplification-free microRNA sensing with fM-sensitivity and single-base-specificity. Chem Commun 2021;57:247–50. https://doi.org/10.1039/d0cc06412b.

76. Sridhara S, Goswami HN, Whyms C, Dennis JH, Li H. Virus detection via programmable Type III-A CRISPR-Cas systems. Nat Commun 2021;12:1–10. https://doi.org/10.1038/s41467-021-25977-7.

77. Steens JA, Zhu Y, Taylor DW, Bravo JPK, Prinsen SHP, Schoen CD, et al. SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. Nat Commun 2021;12:1–12. https://doi.org/10.1038/s41467-021-25337-5.

78. Grüschow S, Adamson CS, White MF. Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector. Nucleic Acids Res 2021;49:13122–34. https://doi.org/10.1093/nar/gkab1190.

79. Zhang Y, Qian L, Wei W, Wang Y, Wang B, Lin P, et al. Paired Design of dCas9 as a Systematic Platform for the Detection of Featured Nucleic Acid Sequences in Pathogenic Strains. ACS Synth Biol 2017;6:211–6. https://doi.org/10.1021/acssynbio.6b00215.

80. Basu AS. Digital Assays Part I: Partitioning Statistics and Digital PCR. SLAS Technol 2017;22:369–86. https://doi.org/10.1177/2472630317705680.

81. Rissin DM, Walt DR. Digital concentration readout of single enzyme molecules using femtoliter arrays and poisson statistics. Nano Lett 2006;6:520–3. https://doi.org/10.1021/nl060227d.

82. Rogacs A, Marshall LA, Santiago JG. Purification of nucleic acids using isotachophoresis. J Chromatogr A 2014;1335:105–20. https://doi.org/10.1016/j.chroma.2013.12.027.

83. Grieshaber D, MacKenzie R, Vörös J, Reimhult E. Electrochemical biosensors - Sensor principles and architectures. Sensors 2008;8:1400–58. https://doi.org/10.3390/s8031400.

84. Hayat A, Catanante G, Marty JL. Current trends in nanomaterial-based amperometric biosensors. Sensors (Switzerland) 2014;14:23439–61. https://doi.org/10.3390/s141223439.

85. Le Ru EC, Meyer M, Etchegoin PG. Proof of single-molecule sensitivity in Surface Enhanced Raman Scattering (SERS) by means of a two-analyte technique. J Phys Chem B 2006;110:1944–8. https://doi.org/10.1021/jp054732v.

86. Nie S, Emory SR. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. Science 1997;275:1102–6. https://doi.org/10.1126/science.275.5303.1102.

87. Koczula KM, Gallotta A. Lateral flow assays. Essays Biochem 2016;60:111–20. https://doi.org/10.1042/EBC20150012.

88. Wang K, Wang Y, Wang C, Jia X, Li J, Xiao R, et al. Facile synthesis of high-performance SiO2@Au core-shell nanoparticles with high SERS activity. RSC Adv 2018;8:30825–31. https://doi.org/10.1039/c8ra05213a.

89. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, et al. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. Cell 2009;139:945–56. https://doi.org/10.1016/j.cell.2009.07.040.

90. Marraffini LA, Sontheimer EJ. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 2008;322:1843–5. https://doi.org/10.1126/science.1165771.

91. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, et al. Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. Mol Cell 2012;45:292–302. https://doi.org/10.1016/j.molcel.2011.10.023.

92. Staals RHJ, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, et al. RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. Mol Cell 2014;56:518–30. https://doi.org/10.1016/j.molcel.2014.10.005.

93. Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas Č, Nwokeoji AO, Dickman MJ, et al. Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. Mol Cell 2014;56:506–17. https://doi.org/10.1016/j.molcel.2014.09.027.

94. Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA. Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. Cell 2015;161:1164–74. https://doi.org/10.1016/j.cell.2015.04.027.

95. Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science 2017;357:605–9. https://doi.org/10.1126/science.aao0100.

96. Niewoehner O, Garcia-Doval C, Rostøl JT, Berk C, Schwede F, Bigler L, et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature 2017;548:543–8. https://doi.org/10.1038/nature23467.

97. McMahon SA, Zhu W, Graham S, Rambo R, White MF, Gloster TM. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. Nat Commun 2020;11. https://doi.org/10.1038/s41467-019-14222-x.

98. Lau RK, Ye Q, Birkholz EA, Berg KR, Patel L, Mathews IT, et al. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Mol Cell 2020;77:723-733.e6. https://doi.org/10.1016/j.molcel.2019.12.010.

99. Rouillon C, Schneberger N, Chi H, Peter MF, Geyer M, Boenigk W, et al. SAVED by a toxin: Structure and function of the CRISPR Lon protease. BioRxiv 2021. https://doi.org/doi.org/10.1101/2021.12.06.471393.

100. Makarova KS, Timinskas A, Wolf YI, Gussow AB, Siksnys V, Venclovas Č, et al. Evolutionary and functional classification of the CARF domain superfamily, key sensors in

prokaryotic antivirus defense. Nucleic Acids Res 2020:1–20. https://doi.org/10.1093/nar/gkaa635.

101. Athukoralage JS, White MF. Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. Rna 2021;27:855–67. https://doi.org/10.1261/rna.078739.121.

102. Jia N, Jones R, Yang G, Ouerfelli O, Patel DJ. CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. Mol Cell 2019;75:944-956.e6. https://doi.org/10.1016/j.molcel.2019.06.014.

103. Nemudraia A, Nemudryi A, Buyukyoruk M, Scherffius A, Zahl T, Wiegand T, et al. Sequence-specific capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. Res Sq 2022. https://doi.org/10.21203/rs.3.rs-1466718/v1.

104. Daum LT, Fischer GW. Ion torrent genomic sequencing. US9365904B2, 2013.

105. Nyren P. METHOD OF SEQUENCING DNA BASED ON THE DETECTION OF THE RELEASE OF PYROPHOSPHATE AND ENZYMATIC NUCLEOTIDE DEGRADATION. US6258568B1, 2001.

106. Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. Science 1998;281:363–5. https://doi.org/10.1126/SCIENCE.281.5375.363/ASSET/5A4EFC11-FC20-44E2-B7F1-F9248787B850/ASSETS/GRAPHIC/363-2.GIF.

107. Tanner NA, Zhang Y, Evans TC. Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes. Biotechniques 2015;58:59–68. https://doi.org/10.2144/000114253.

108. Tomita N, Mori Y, Kanda H, Notomi T. Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. Nat Protoc 2008;3:877–82. https://doi.org/10.1038/nprot.2008.57.

109. Garcia-Doval C, Schwede F, Berk C, Rostøl JT, Niewoehner O, Tejero O, et al. Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. Nat Commun 2020;11:1–9. https://doi.org/10.1038/s41467-020-15334-5.

110. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc Natl Acad Sci U S A 2009;106:7702–7. https://doi.org/10.1073/pnas.0901054106.

111. Quan J, Langelier C, Kuchta A, Batson J, Teyssier N, Lyden A, et al. FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. Nucleic Acids Res 2019;47:e83. https://doi.org/10.1093/nar/gkz418.

112. Wang T, Liu Y, Sun HH, Yin BC, Ye BC. An RNA-Guided Cas9 Nickase-Based Method for Universal Isothermal DNA Amplification. Angew Chemie - Int Ed 2019;58:5382–6. https://doi.org/10.1002/anie.201901292.

113. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011;471:602–7. https://doi.org/10.1038/nature09886.

114. Jung C, Hawkins JA, Jones SK, Ke A, Press WH, Finkelstein IJ, et al. Massively Parallel Biophysical Analysis of CRISPR- Cas Complexes on Next Generation Article Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. Cell 2017;170:35-47.e13. https://doi.org/10.1016/j.cell.2017.05.044.

115. Karvelis T, Druteika G, Bigelyte G, Budre K, Zedaveinyte R, Silanskas A, et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. Nature 2021;599:692–6. https://doi.org/10.1038/s41586-021-04058-1.

116. Kuzmenko A, Yudin D, Ryazansky S, Kulbachinskiy A, Aravin AA. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. Nucleic Acids Res 2019;47:5822–36. https://doi.org/10.1093/nar/gkz379.

117. Hegge JW, Swarts DC, Chandradoss SD, Cui TJ, Kneppers J, Jinek M, et al. DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute. Nucleic Acids Res 2019;47:5809–21. https://doi.org/10.1093/nar/gkz306.

118. Béraud A, Sauvage M, Bazán CM, Tie M, Bencherif A, Bouilly D. Graphene field-effect transistors as bioanalytical sensors: Design, operation and performance. Analyst 2021;146:403–28. https://doi.org/10.1039/d0an01661f.

119. Richards FM. ON THE ENZYMIC ACTIVITY OF SUBTILISIN-MODIFIED RIBONUCLEASE. Proc Natl Acad Sci 1958;44:162–6. https://doi.org/10.1073/pnas.44.2.162.

120. Kerppola TK. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. Annu Rev Biophys 2008;37:465–87. https://doi.org/10.1146/annurev.biophys.37.032807.125842.

121. Lee DH, Schleif RF. In vivo DNA loops in araCBAD: size limits and helical repeat. Proc Natl Acad Sci 1989;86:476–80. https://doi.org/10.1073/pnas.86.2.476.

122. Santiago-Frangos A, Buyukyoruk M, Wiegand T, Krishna P, Wiedenheft B. Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. Curr Biol 2021:1–10. https://doi.org/10.1016/j.cub.2021.05.068.

123. Lefever S, Hellemans J, Pattyn F, Przybylski DR, Taylor C, Geurts R, et al. RDML:

Structured language and reporting guidelines for real-time quantitative PCR data. Nucleic Acids Res 2009;37:2065–9. https://doi.org/10.1093/nar/gkp056.

124. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem 2009;55:611–22. https://doi.org/10.1373/clinchem.2008.112797.

125. O'Leary JJ. Potential viral pathogenic mechanism for new variant inflammatory bowel disease [1]. J Clin Pathol - Mol Pathol 2003;56:248. https://doi.org/10.1136/mp.56.4.248.

126. Bustin SA. The reproducibility of biomedical research: Sleepers awake! Biomol Detect Quantif 2014;2:35–42. https://doi.org/10.1016/j.bdq.2015.01.002.

127. FDA. In Vitro Diagnostics EUAs | FDA n.d. https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/in-vitro-diagnostics-euas (accessed January 14, 2022).

128. Higuchi R, Dollinger G, Walsh PS, Griffith R. Simultaneous Amplification and Detection of Specific DNA Sequences. Biotechnology 1992;10:413–7. https://doi.org/10.1038/nbt0492-413.

129. Sherlock Biosciences I. Sherlock CRISPR SARS-CoV-2 kit. FDA 2021. https://www.fda.gov/media/137746/download.

130. Mammoth Biosciences I. SARS-CoV-2 RNA DETECTR Reagent Kit. FDA 2020. https://www.fda.gov/media/141765/download.

131. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V, et al. The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol 2014;21:743–53. https://doi.org/10.1038/nsmb.2879.

132. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJJ, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 2020;18:67–83. https://doi.org/10.1038/s41579-019-0299-x.

133. Makarova KS, Wolf YI, van der Oost J, Koonin E V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct 2009;4:29. https://doi.org/10.1186/1745-6150-4-29.

134. Kim HK, Kim Y, Lee S, Min S, Bae JY, Choi JW, et al. SpCas9 activity prediction by DeepSpCas9, a deep learning–based model with high generalization performance. Sci Adv 2019;5:1–10. https://doi.org/10.1126/sciadv.aax9249.

135. Chuai G, Ma H, Yan J, Chen M, Hong N, Xue D, et al. DeepCRISPR: Optimized

CRISPR guide RNA design by deep learning. Genome Biol 2018;19:1–18. https://doi.org/10.1186/s13059-018-1459-4.

136. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. Nat Methods 2019;16:687–94. https://doi.org/10.1038/s41592-019-0496-6.

137. Huang PS, Boyken SE, Baker D. The coming of age of de novo protein design. Nature 2016;537:320–7. https://doi.org/10.1038/nature19946.

138. Turner A. How Many People Have Smartphones Worldwide (Apr 2022) n.d. https://www.bankmycell.com/blog/how-many-phones-are-in-the-world#1579705085743-b3697bdb-9a8f (accessed April 21, 2022).

139. Arizti-Sanz J, Freije CA, Stanton AC, Petros BA, Boehm CK, Siddiqui S, et al. Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2. Nat Commun 2020;11. https://doi.org/10.1038/s41467-020-19097-x.

140. Kumar M, Gulati S, Ansari AH, Phutela R, Acharya S, Azhar M, et al. FnCas9 based CRISPR diagnostic for rapid and accurate detection of major SARS-CoV2 variants on a paper strip. Elife 2021;10:1–25. https://doi.org/10.7554/eLife.67130.

141. Azhar M, Phutela R, Kumar M, Ansari AH, Rauthan R, Gulati S, et al. Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. Biosens Bioelectron 2021;183. https://doi.org/10.1016/j.bios.2021.113207.

142. CDC. Real-Time RT-PCR diagnostic panel for emergency use only 2020;3:1–42.

CHAPTER THREE

SEQUENCE-SPECIFIC CAPTURE AND CONCENTRATION OF VIRAL RNA BY TYPE III CRISPR SYSTEM ENHANCES DIAGNOSTIC

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Anna Nemudraia

Contributions:

Co-Author: Artem Nemudryi

Contributions:

Co-Author: Murat Buyukyoruk

Contributions:

Co-Author: Andrew M. Scherffius

Contributions:

Co-Author: Trevor Zahl

Contributions:

Co-Author: Tanner Wiegand

Contributions:

Co-Author: Shishir Pandey

Contributions:

Co-Author: Joseph E. Nichols

Contributions:

Co-Author: Laina Hall

Contributions:

Co-Author: Aidan McVey

Contributions: Helen H Lee

Co-Author: Royce A. Wilkinson

Contributions:

Co-Author: Laura R. Snyder

Contributions:

Co-Author: Joshua D. Jones

Contributions:

Co-Author: Kristin S. Koutmou

Contributions:

Co-Author: Andrew Santigo-Frangos

Contributions:

Co-Author: Blake Wiedenheft

Contributions: Conceptualization, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Manuscript Information

Anna Nemudraia[1,2], Artem Nemudryi[1,2], Murat Buyukyoruk[1,3], Andrew M. Scherffius[1,3], Trevor Zahl[1], Tanner Wiegand[1], Shishir Pandey[1], Joseph E. Nichols[1], Laina Hall[1], Aidan McVey[1], Helen H Lee[1], Royce A. Wilkinson[1], Laura R. Snyder[4], Joshua D. Jones[4], Kristin S. Koutmou[4], Andrew Santiago-Frangos[1*], and Blake Wiedenheft[1*]

[1]Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT 59717, USA

[4]Department of Chemistry, University of Michigan, Ann Arbor, MI 48105, USA

[2]These authors contributed equally, [3]These authors contributed equally

[4]Department of Chemistry, University of Michigan, Ann Arbor, MI 48105, USA

*Correspondence: andrew.santiagofrangos@gmail.com and bwiedenheft@gmail.com.

Status of Manuscript:

___ Prepared for submission to a peer-reviewed journal

___ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

_X_ Published in a peer-reviewed journal

Nature Communications

Volume 13

15 December 2022

DOI: 10.1038/s41467-022-35445-5

Abstract

Type-III CRISPR-Cas systems have recently been adopted for sequence-specific detection of SARS-CoV-2. Here, we repurpose the type III-A CRISPR complex from *Thermus thermophilus* (TtCsm) for programmable capture and concentration of specific RNAs from complex mixtures. The target bound TtCsm complex generates two cyclic oligoadenylates (i.e., cA_3 and cA_4) that allosterically activate ancillary nucleases. We show that both Can1 and Can2 nucleases cleave single-stranded RNA, single-stranded DNA, and double-stranded DNA in the presence of cA_4. We integrate the Can2 nuclease with type III-A RNA capture and concentration for direct detection of SARS-CoV-2 RNA in nasopharyngeal swabs with 15 fM sensitivity. Collectively, this work demonstrates how type-III CRISPR-based RNA capture and concentration simultaneously increases sensitivity, limits time to result, lowers cost of the assay, eliminates solvents used for RNA extraction, and reduces sample handling.

Introduction

Although qPCR (quantitative polymerase chain reaction) remains the "gold standard" for nucleic acid detection, it requires sophisticated equipment, trained personnel, efficient specimen transport to high-complexity labs, and reliable reporting systems[1]. While the complexity and turnaround times necessary for qPCR are acceptable for many diagnostic applications, the SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) pandemic reveals an urgent need for diagnostics that are easy to distribute, simple to perform, and fast enough to stop transmission of a contagious disease[1]. Although rapid antigen tests and isothermal amplification methods have helped address this need, these and other emerging methods have limitations related to sensitivity, versatility, or specificity[2,3].

CRISPR RNA-guided diagnostics (CRISPR-dx) are a diverse group of nascent technologies that aim to address current limitations by providing a versatile and programmable platform that is sufficiently sensitive for clinical applications and stable enough for distribution[4,5]. The first CRISPR-based viral diagnostic came from Collins and colleagues in 2016, when they demonstrated that Cas9 could be used to discriminate between different variants of the Zika virus[6]. This approach relies on converting viral RNA to DNA using reverse transcriptase, followed by isothermal DNA amplification prior to sequence-based discrimination by Cas9. The exclusive recognition of double-stranded DNA (dsDNA) by Cas9 seemed to be an intrinsic limitation for diagnostic applications that require RNA detection. However, Beisel and colleagues recently developed a creative method that uses the trans-acting CRISPR-RNA (tracrRNA) to capture complementary RNA guides derived from RNA viruses[7]. In this system, the engineered tracrRNA-crRNA hybrid guides Cas9 to a complementary dsDNA reporter. While

this approach enables RNA detection, Cas9 is a single turn-over enzyme, which may limit its sensitivity. In contrast to Cas9, target recognition by type V (Cas12-DETECTR) and type VI (Cas13-SHERLOCK) CRISPR-systems activates a multi-turnover non-sequence-specific "collateral nuclease" activity that amplifies the signal by cleaving thousands of reporter molecules for every target bound[8,9]. The recently reported rates of collateral cleavage by target bound Cas12 and Cas13 nucleases indicate that the theoretical limit of detection, without prior amplification of the target is 1 pM (~10^6 copies/μL)[10,11].

Like type VI, type III CRISPR systems also recognize RNA. However, unlike any other system, target recognition by type III complexes simultaneously activates polymerase and HD-nuclease domains in the Cas10 subunit[12–14]. The polymerase domain has been estimated to generate ~1000 cyclic oligoadenylates per bound RNA[15], which trans-activate multi-turnover ancillary nucleases that provide defense from invading genetic parasites[16,17]. Thus, this biochemical cascade provides two signal amplifications steps, the second of which activates thousands of multi-turnover ancillary nucleases, rather than relying on a single collateral nuclease (i.e., Cas13) bound to a target. However, initial efforts to implement this approach failed to be sufficiently sensitive for clinical applications without prior amplification of the target RNA[18–20]. The sensitivity of this first-generation diagnostic was in part limited by the use of an ancillary nucleases (e.g., Csm6) that also degrade the cyclic nucleotide activator[21–25]. Recently, Malcolm White's lab demonstrated that alternative ancillary nucleases, which efficiently cleave reporters but do not cleave the signaling molecule, can be used to enhance the sensitivity of type III-based diagnostics[26].

Despite innovations leading to new and improved CRISPR-based diagnostics, point-of-care testing requires new strategies that simplify the workflow and increase the sensitivity without prior RNA purification or amplification (e.g., PCR, LAMP, NASB, RPA, etc.). Here, we bring CRISPR-dx closer to a deployable diagnostic by developing a type III CRISPR-based method for sequence-specific capture and concentration of a specific RNA from a complex mixture. To improve the sensitivity, we purify several different ancillary nucleases (i.e., Can1, Can2, and NucC), systemically test nuclease activation using a series of purified cyclic oligoadenylate standards, and test for ring nuclease activity. We show that Can1 and Can2 nucleases cleave single-stranded RNA (ssRNA) and DNA (ssDNA), as well as double-stranded DNA (dsDNA) in the presence of a Cas10-generated cyclic oligomer composed of four adenosines (i.e., cA_4). We demonstrate how the type III complex can be used to bypass RNA extraction methods, and that coupling type III-based RNA capture with the AaCan2 nuclease further increases the sensitivity of SARS-CoV-2 RNA detection in patient swabs to 15 fM ($8.3x10^3$ copies/µL).

Results

Type III-mediated sequence-specific enrichment of RNA: Type III CRISPR RNA-guided complexes (i.e., Csm and Cmr) bind and cleave complementary single-stranded RNA (ssRNA) targets[27]. Complementary RNA is cleaved in six-nucleotide increments by metal-dependent nucleases (Csm3 or Cmr4) that form the oligomeric "backbone" of the complex[28]. Type III complexes release fragments of the cleaved target, which inactivates ATP polymerization by the Cas10 subunit[28]. Previously, we mutated residues in the Csm3 subunit responsible for target RNA cleavage (D34A), purified the RNase-dead complex (TtCsm$^{Csm3\text{-}D34A}$), and showed that the mutant complex provides more sensitive detection of viral RNA than the wild-type complex[18]. To further increase the sensitivity, we set out to determine if TtCsm$^{Csm3\text{-}D34A}$ could be used to concentrate sequence-specific RNAs. To test this approach, we mixed ^{32}P-labeled target or non-target RNAs with TtCsm$^{Csm3\text{-}D34A}$, incubated for 20 minutes, and concentrated the His-tagged complex using nickel-derivatized magnetic beads (**Fig. 1a, Supplementary Fig. 3.1a**). The beads were concentrated using a magnet, and RNAs were extracted from the bound and unbound fractions. The type III complex captured most of the radiolabeled target RNA (76±5.8%), while non-target RNA primarily remains in the supernatant (**Fig. 3.1b, Supplementary Fig. 3.1b, c**). To determine if type III CRISPR-based RNA capture and concentration results in the synthesis of more cyclic nucleotides, we added Csm-beads to 120 µL of a sample containing a mixture of SARS-CoV-2 RNA and total RNA extracted from HEK 293T cells (**Fig. 3.1c,** see Methods). After concentrating the beads with a magnet, we resuspended the pellet in a buffer containing α-^{32}P-ATP, allowed the cyclic polymerization to proceed, and analyzed the reactions using thin-layer chromatography (TLC). Type III CRISPR-based concentration increases the amount of cA_3

and cA_4, as compared to reaction performed without RNA concentration (**Fig. 3.1c, d, Supplementary Fig. 3.1d**).

Previously, we repurposed TtCsm6, a cA_4-activated ribonuclease, to generate a real-time fluorescent readout for Csm-based RNA detection[18] (**Fig. 3.1e, top**). We reasoned that increased cA_4 levels after RNA enrichment will boost the nuclease activity of TtCsm6 and therefore increase the sensitivity of the RNA detection. To test this hypothesis, we titrated 10^8 to 10^5 copies/μL of SARS-CoV-2 N-gene RNA into total RNA extracted from HEK 293T cells, concentrated the target RNA using $TtCsm^{Csm3\text{-}D34A}$, resuspended the beads in a buffer containing ATP, and then transferred the polymerization products to a reaction containing TtCsm6 and a fluorescent RNA reporter (i.e., FAM-RNA-Iowa Black FQ). Csm-based RNA enrichment increased the sensitivity of the assay 100-fold compared to the assay without the pull-down (**Fig. 1e**). Taken together, these results demonstrate how type III-A CRISPR-complexes can be used to capture sequence-specify RNAs, resulting in a higher concentration of cyclic nucleotides, which improves the sensitivity of sequence-specific RNA detection.

CARF-nucleases Can1 and Can2 exhibit cA_4-specific nuclease activities: Csm6 proteins contain an amino-terminal CARF (CRISPR-associated Rossman Fold) and a carboxy-terminal HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domain[12,14]. Csm6 family proteins form homodimers, and the two CARF-domains bind cA_4[25,29] or cA_6[24], which activate the C-terminal HEPN nuclease domain. However, the CARF domain of some Csm6 proteins also degrades the cyclic nucleotide, which inactivates the nuclease and may limit the sensitivity of Csm6-based assays[30]. To improve the sensitivity, we sought to identify and incorporate a CARF-nuclease that is activated by, but does not degrade, cA_4.

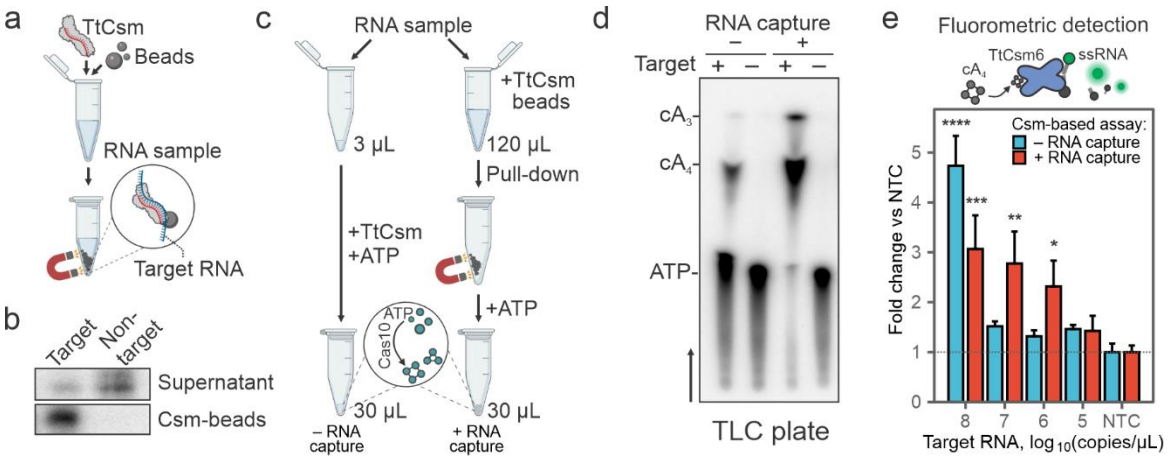


Fig 3.1: Type III CRISPR-based RNA concentration enhances sequence-specific detection of RNA**. a** Schematic of type III CRISPR-based RNA concentration. RNase-dead type III CRISPR complex from *Thermus thermophilus* (e.g., TtCsm$^{Csm3\text{-}D34A}$) is added to a sample to bind complementary "Target RNA". The His-tagged complex is concentrated using nickel-derivatized magnetic beads and a magnet. **b** Sequence-specific RNA enrichment with TtCsm$^{Csm3\text{-}D34A}$ was tested using 25 nM ^{32}P 5`-end labeled RNA. Target or non-target RNA fragments were mixed with His-tagged TtCsm$^{Csm3\text{-}D34A}$ complex (125 nM) and concentrated with nickel-derivatized magnetic beads. Phenol-chloroform extracted RNAs from the supernatants and Csm-beads were resolved using UREA-PAGE. **c** Csm-based direct RNA detection using 3 µL of sample is compared to an assay where TtCsm$^{Csm3\text{-}D34}$ is used to capture RNA at the same concentration from a 120 µL volume. Magnetic beads decorated with TtCsm$^{Csm3\text{-}D34A}$ are added to the sample. After concentrating beads with a magnet, the supernatant is decanted, and he pellet is resuspended in a reaction buffer with ATP to activate polymerase activity of Cas10. Polymerization products (e.g., cA_3 and cA_4) are used for the downstream assays. **d** TtCsm$^{Csm3\text{-}D34A}$ polymerization reactions were performed with α-^{32}P-ATP as shown in **(c)** and products were resolved using thin-layer chromatography (TLC). Black arrow shows migration of solvent in the TLC plate. Bands were annotated using chemically synthesized standards (Supplementary Fig. 1d). 3 µL (- RNA capture) or 120 µL (+ RNA capture) of SARS-CoV-2 N-gene RNA (10^{10} copies/µL) diluted in total RNA of 293T cells were used for reactions. **e** TtCsm6-based fluorescent readout (top panel) is used for detection of cA_4 generated by TtCsm$^{Csm3\text{-}D34A}$ with (red bars) or without RNA capture step (blue bars) as shown in panel **(c)**. SARS-CoV-2 N-gene RNA diluted in total RNA of 293T cells was used as a target. Fluorescence was measured with qPCR instrument and normalized to the no target control (NTC, 293T RNA only, dashed line). In each assay, the mean (n=3) fluorescent signal was compared with one-way ANOVA. Pairwise comparisons with NTC were performed using post hoc Dunnett's test. Data are shown as mean ± SD. $*p < 0.05$; $**p < 0.01$; $***p<0.001$; $****p < 10^{-4}$.

CRISPR ancillary nucleases (Can) are another family of recently identified proteins that are activated by cyclic oligoadenylates and lack ring nuclease activity[31–33]. Like Csm6 proteins, Can proteins also contain amino-terminal CARF domains, but the carboxy-terminal nucleases are distinct. The Can1 protein from *Thermus thermophilus* (TtCan1) has a unique monomeric architecture with two non-identical CARF domains, one nuclease-like domain (NLD) and one restriction endonuclease domain (PD-(D/E)XK)[33], while Can2 nucleases contain a single CARF domain and form symmetrical homodimers[31,32] (**Fig 3.2a**).

To identify Can1 and Can2 orthologs, we generated Hidden Markov models (HMMs) to query publicly available microbial genomes and metagenomes from NCBI and JGI. This analysis identified 204 Can1 and 3,121 Can2 proteins. Based on this analysis, we selected one Can1 protein and three Can2 proteins from thermophilic organisms. We hypothesized that thermostable nucleases will be compatible with RNA detection and target RNA activated polymerization of ATP by the Csm complex from *Thermus thermophilus* (i.e., TtCsm). Elevated temperatures are anticipated to improve accessibility to RNA targets that might otherwise be obscured by secondary structures and the use of thermostable ancillary nucleases may enable detection of a specific RNA in a single tube at a single temperature. In addition, the stability of these proteins[34], may have downstream benefits when it comes to packaging and distribution.

Consistent with the previous research[33], TtCan1 exhibits Mn^{2+}-dependent linearization of supercoiled plasmid DNA in the presence of cA_4 (**Fig. 3.2c**). Given the unique asymmetric interaction between the two non-identical CARF-domains of TtCan1 and cA_4, we hypothesized that other cyclic oligonucleotides without two-fold symmetry (e.g., cA_3 or cAAG) might allosterically activate the nuclease activity. We tested a library of twelve chemically synthesized

cyclic oligonucleotides in a reaction with TtCan1 and three different substrates (i.e., dsDNA, ssRNA, and ssDNA) (**Supplementary Fig. 3.2**). Although none of the other cyclic oligonucleotides activate the TtCan1 nuclease, cA_4-activated TtCan1 robustly cleaves ssRNA and long but not short ssDNA substrates (**Fig. 3.2d, e, Supplementary Fig. 3.2c**). This pattern suggests that the nuclease activity of TtCan1 is either sequence-specific and the preferred cut site is not present in the short oligonucleotide or this nuclease requires a longer substrate for recognition and/or cleavage.

Can2 genes from *Clostridium thermobutyricum* (CthCan2), *Thermus thermophilus* (TtCan2), and *Archaeoglobi* archaeon JdFR-42 (AaCan2) were cloned and expressed in *E. coli* (**Fig. 3.2b**). However, only AaCan2 purified in quantities sufficient for biochemical assays (**Supplementary Fig. 3.3**). We systematically tested the activities of AaCan2 on different substrates with a range of cyclic oligoadenylates (**Supplementary Fig. 4**). Like TtCan1, AaCan2 also linearizes supercoiled plasmid DNA, and cleaves both ssRNA and ssDNA when activated with cA_4 (**Fig. 2f-h, Supplementary Fig. 3.4)**. Cleavage of a ssDNA short oligos (71 nt) produces a discrete band suggesting that the enzyme processes ssDNA to a minimal cleavage product or that the activity is sequence-specific (**Supplementary Fig. 3.4c**). Collectively, our results demonstrate that TtCan1 and AaCan2 function as cA_4-activated nucleases on plasmid DNA, ssRNA, and ssDNA.

Both TtCan1 and AaCan2 linearize supercoiled DNA when activated with cA_4, which is followed by slow DNA degradation (**Supplementary Fig. 3.5a**). To determine if the initial cleavage event is sequence-specific, we sequenced plasmid DNA linearized with TtCan1 or AaCan2. To identify the cleavage site(s), we mapped the start and end positions of each read to

the reference DNA sequence. This analysis identified that while both nucleases preferentially linearize the supercoiled DNA in two regions (1,700-1,750 and 5,800-5,850 bp), the most frequent TtCan1 cut sites map to poly-G sequences, while AaCan2 cuts at poly-A and poly-T sites, which indicates different cleavage preference of the two nucleases (**Supplementary Fig. 3.5b, c**).

While TtCan1 and AaCan2 rapidly digest supercoiled DNA with distinct sequence preferences, we do not detect cleavage at the same sites if DNA is linearized beforehand (**Supplementary Fig** 3.**5d, e**). Further, both nucleases digest single-stranded form of ΦX174 bacteriophage DNA to small fragments but cut at several locations in the supercoiled double-stranded replicative form (RF1) that has identical nucleotide sequence (**Fig. 3.3 e, h, Supplementary Fig. 3.5f**). Taken together, this data suggests that TtCan1 and AaCan2 nucleases specifically act on single-stranded nucleic acids, but not on relaxed dsDNA. Collectively, this suggests that plasmid linearization is driven by cleavage in the underwound regions of the supercoiled DNA.

Can2 ancillary nuclease provides sensitive Csm-based RNA detection: To determine if incorporating TtCan1 or AaCan2 improves sensitivity of the Csm-based RNA detection assay, we screened a library of short synthetic RNA and DNA reporters designed to identify sequences that might be preferred by these nucleases (see Methods, **Supplementary Table 3.1**, and **Supplementary Fig. 3.6**). Consistent with our gel-based assays, cA_4-activated AaCan2 cleaves DNA reporters in the presence of Mn^{2+} and RNA reporters in the presence of either Mg^{2+} or Mn^{2+}, but reactions with RNA reporter and Mn^{2+} consistently result in higher fluorescent signal (**Supplementary Fig. 3.6 a-c**).

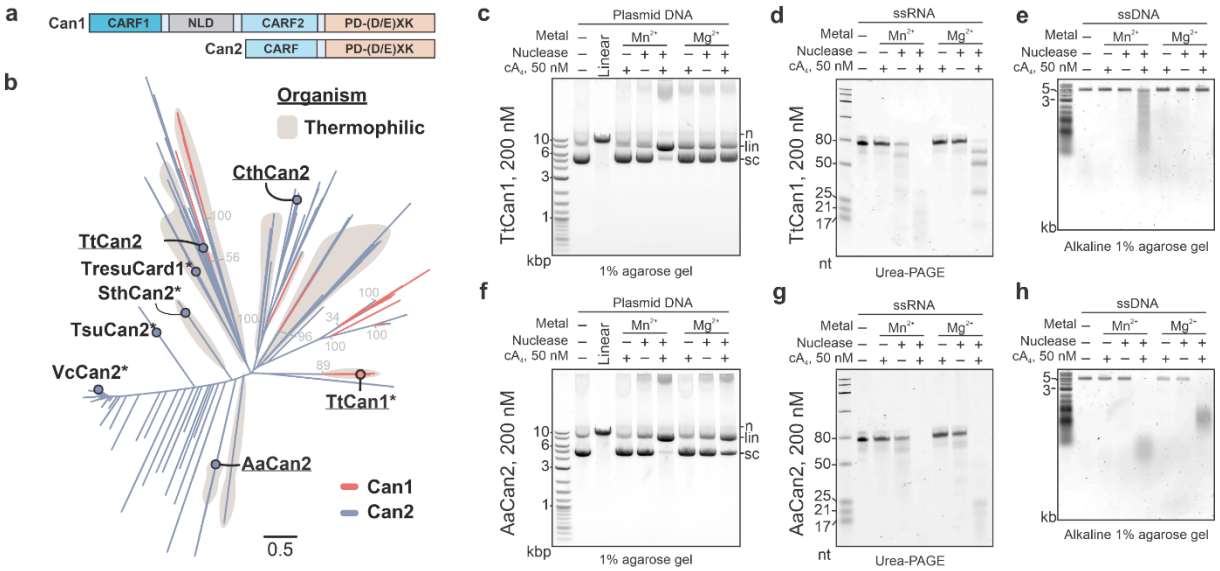


Fig. 3.2: Can1 and Can2 ancillary nucleases function as cA4-activated nucleases on plasmid DNA, ssRNA and ssDNA**. a** Domain organization of Can1 and Can2 proteins. Can2 proteins have two domains – CARF and PD-(D/E)XK superfamily nuclease domain. Can1 is predicted to be derived from Can2 by gene duplication[32]. NLD – nuclease-like domain. **b** Maximum-likelihood phylogeny of 204 Can1 (CARF2 and PD-(D/E)XK nuclease domain) and 3,121 Can2 proteins. The light gray numbers indicate bootstrap values. Previously studied effectors are marked with asterisks (*). Effectors chosen for purification and *in vitro* experiments are underlined. **c - e** Plasmid DNA (15 nM), ssRNA (425 nM), and ssDNA (15 nM) cleavage assay with TtCan1 (200 nM) in the presence of cA_4 (50 nM). The reactions were incubated 15 min at 60°C. n – nicked, lin – linear, sc – supercoiled plasmid. **f - h** Cleavage assays with AaCan2 (200 nM) in the presence of cA_4 (50 nM). Assays were performed with 15 nM plasmid DNA, 425 nM ssRNA or 15 nM ssDNA for 15 min at 60°C. n – nicked, lin – linear, sc – supercoiled plasmid.

Having established that AaCan2 is more active than TtCan1, we set out to compare AaCan2 to the sensitivity of TtCsm6, which we used previously[18]. This comparison was performed by measuring cA_4 concentration-dependent activity for AaCan2 and TtCsm6 using the preferred RNA reporter and conditions that support the highest activity among tested variables (**Supplementary Fig. 3.6, 3.7**). AaCan2 produces a similar fluorescent signal to TtCsm6 when activated with 20-fold less cA_4 (0.5 nM versus 10 nM) (**Fig. 3.3a, b**). Moreover, AaCan2 exhibits an incremental decrease in cleavage rates with decreasing cA_4, while TtCsm6 exhibits a dramatic (non-linear) drop in the activity. This distinction in activity between the enzymes is consistent with the ring-nuclease activity of TtCsm6 rapidly degrading its activator, while AaCan2 binds and preserves the cyclic nucleotide (**Supplementary Fig. 3.8**).

Incorporating AaCan2 into the type III-based detection assay produces high background that is only evident in the presence of the TtCsm-complex, whereas AaCan2 alone demonstrates very little non-specific cleavage (**Supplementary Fig. 3.9a**, **Fig. 3.3b**). This disparity suggests that non-sequence specific activation of the Cas10 polymerase may generate low levels of cA_4, which stably activates AaCan2, whereas the ring-nuclease of TtCsm6 rapidly degrades cA_4 limiting the background signal. To test this hypothesis, we tritiated concentration of TtCsm$^{Csm3\text{-}D34A}$-complex and show that 50-fold less of the complex (i.e., 25 nM to 0.5 nM) reduces the background without compromising target-specific activity, which significantly improves signal-to-noise ($p < 0.001$, **Supplementary Fig. 3.9b, c**).

Finally, we benchmarked AaCan2 and TtCsm$^{Csm3\text{-}D34A}$-complex combination against TtCsm6-based detection. The TtCsm6-based assay reliably detects 10^6 copies/μL of target RNA (**Fig. 3.3c**), while AaCan2-based reactions are more sensitive (10^5 copies/μL) (**Fig. 3.3d**).

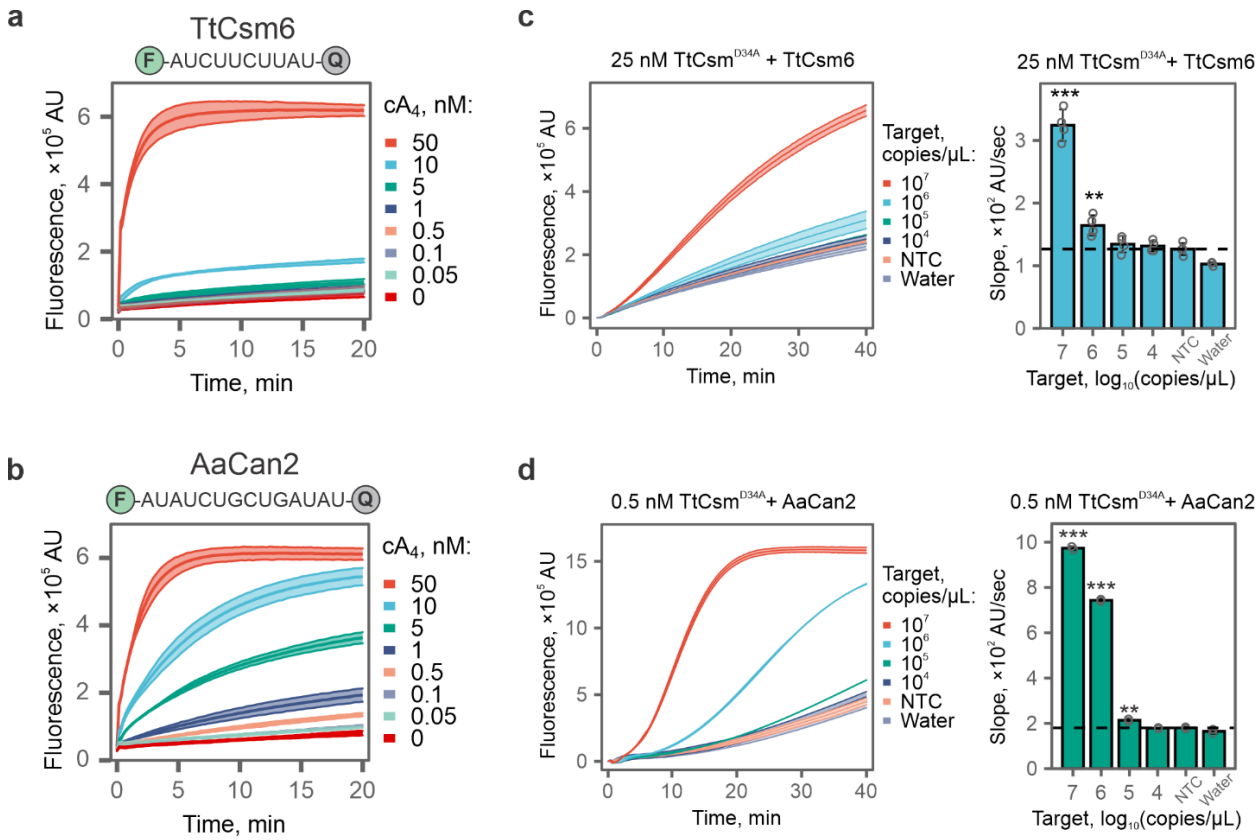


Fig.3.3: Incorporation of Can2 nuclease improves sensitivity of Csm-based RNA detection. **a** TtCsm6 (300 nM) and **b** AaCan2 (300 nM) cleavage assays with fluorescent ssRNA reporter (top) in the presence of varying cA_4 activator concentrations (shown with colors). Data are shown as the mean (center line) of three replicates ± S.D. (ribbon). The optimal fluorescent reporter (top) was determined using RNA library screen in Supplementary Fig. 6. TtCsm RNA detection coupled with TtCsm6- **(c)** and AaCan2-based **(d)** assays were performed using samples with target RNA concentrations ranging from 10^7 to 10^4 copies/µL. Samples were prepared by spiking IVT fragments of SARS-CoV-2 N gene into total RNA extracted from nasopharyngeal swab patient sample negative for SARS-CoV-2. Cleavage of fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument (left). Data were plotted as mean of four replicates. Simple linear regression was used to calculate slopes for linear regions of the curves. Bars show mean values (n = 4) ± S.D. (right). Data was analyzed with one-way ANOVA followed by multiple comparisons to NTC sample using one-tailed post-hoc Dunnett's test. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Incorporating cA_3-dependent nuclease activity does not provide additional sensitivity of RNA detection:While our assay uses cA_4-activated collateral cleavage of ssRNA reporters, the TtCsm$^{Csm3\text{-}D34A}$ complex also produces cA_3 (**Fig. 3.1d, Supplementary Fig. 3.1d**). We hypothesized that combining cA_3- and cA_4-sensing nucleases might enhance the sensitivity of TtCsm-based detection (**Fig. 3.4a**). NucC (Nuclease, CD-NTase associated) is an endonuclease activated by cA_3 [26,35,36]. We purified three thermophilic NucC orthologs and tested cA_3-dependent dsDNA cleavage (**Supplementary Fig. 3.10**). The NucC from *Clostridium tepidum* (CtNucC) has the highest dsDNase activity and digests plasmid DNA into 300-400 bp fragments in the presence of cA_3 (**Fig. 3.4b, left; Supplementary Fig. 3.11a**). Deep sequencing of cleavage fragments determined that all purified NucC nucleases have a preference for 5'-ANNT-3' sequence motif, which is consistent with previously published work[36] (**Fig. 3.4b**, **right; Supplementary Fig. 3.11b-e**).

Next, we set out to determine if CtNucC and AaCan2 could be combined into a single reaction to improve the sensitivity of RNA detection with TtCsm$^{Csm3\text{-}D34A}$. To perform fluorescent assays with CtNucC, we designed a 31-bp dsDNA reporter comprising six repeats of the optimal cleavage site (**Supplementary Table 1**). The lowest concentration of cA_3 detected by CtNucC is 0.05 nM (**Fig. 3.4c, Supplementary Fig. 3.12, 3.13**). However, TtCsm$^{Csm3\text{-}D34A}$ coupled with CtNucC and dsDNA reporter only detects high concentrations of target RNA (i.e., 10^7 copies/µL; **Fig. 3.4d**). Further, combining CtNucC with AaCan2 and matching fluorescent probes (i.e., dsDNA and ssRNA, respectively) (**Fig. 3.4a**) into a single reaction does not improve the sensitivity compared to detection with AaCan2 alone (**Fig. 3.4d, Supplementary Fig. 3.13**).

While CtNucC is sensitive to cA_3 activation, the TtCsm-complex may not produce sufficient concentrations of this cyclic nucleotide to increase sensitivity over AaCan2 detection alone.

Type III CRISPR-based RNA capture and detection directly from clinical samples: Most CRISPR-based diagnostics assays reported to date use purified RNA as the starting material for viral detection assays[37]. The cost of RNA extraction, time to extract the RNA, and specialized equipment necessary to extract the RNA is often not discussed. We hypothesized that thermostable type III CRISPR complexes (e.g., TtCsm) could be used to capture target RNA directly from patient samples, without prior RNA extraction. Sequence-specific capture and concentration directly from patient samples could drastically reduce time, reagents, and equipment costs.

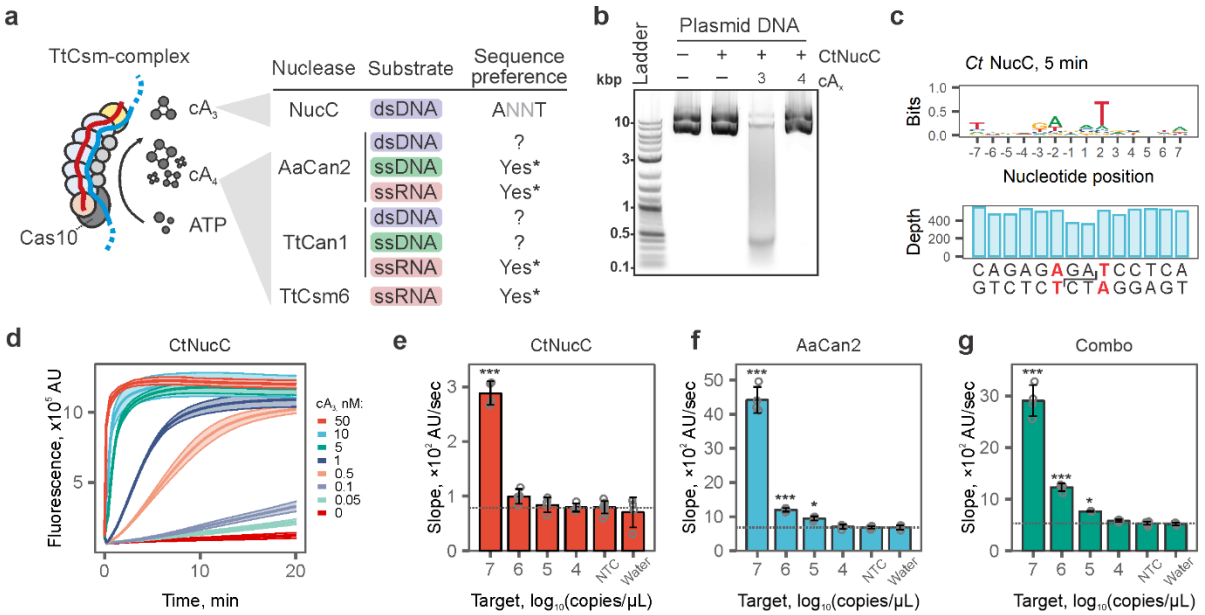


Fig.3.4: Incorporation of cA3-activated nucleases into Csm-based RNA detection assay. **a** The target bound TtCsm complex primarily generates cA_4 and cA_3. Schematic summarizes cA_4- and cA_3-dependent activities of nucleases biochemically tested. N/D – not detected; Asterisks (*) indicate nucleases that have sequences preferences (Supplementary Fig. 3.6). **b** CtNucC (15 nM) is activated by cA_3 (20 nM) and cleaves plasmid DNA into short fragments in 15 min. **c** Deep sequencing of DNA fragments generated after 5 min of incubation with CtNucC revealed the preferential cleavage sites (ANNT). The reduced sequencing depth at the cut site is consistent with a cleavage mechanism producing 3'-overhangs that are removed by T4 DNA polymerase when sequencing library is prepared. **d** CtNucC (300 nM) cleavage assay with fluorescent dsDNA reporter across eight concentrations of cA_3 (shown with colors). Data is shown as mean (center line) of four replicates ± S.D. (ribbon). **e-g** TtCsm RNA detection assays coupled with CtNucC (dsDNA reporter that include preferred cleavage sequence), AaCan2 (ssRNA reporter), and combination of AaCan2 and CtNucC (both reporters). Reactions were performed using samples with target RNA concentrations ranging from 10^7 to 10^4 copies/µL. Samples were prepared by spiking IVT fragment of SARS-CoV-2 N gene in total RNA of SARS-CoV-2 negative nasal swab. Cleavage of the fluorescent reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument. Simple linear regression was used to determine slopes for four replicates. See Supplementary Fig. 3.13 for fluorescent curves used in the analysis. Data were plotted as mean (n = 4) ± S.D. and analyzed with one-way ANOVA. All samples were compared to the non-target RNA control (NTC) using one-tailed post-hoc Dunnett's test. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

To address this challenge, we set out to incorporate “capture and concentration” into the fluorescent based detection method. However, the magnetic beads used to pull-down target RNA unexpectedly obscure the fluorescent signal released when AaCan2 cleaves the reporter (**Supplementary Fig. 3.14d**). Initially, we sidestepped the problem by performing a second magnetic pull-down after the polymerization and transferring the products (without the beads) to a cleavage reaction with the ancillary nuclease (**Figure 3.1e** and **Supplementary Figure 3.14a, b**). However, this introduces an additional liquid handling step (three steps total), complicating the assay. In addition, reactions with AaCan2 following RNA pull-down produce high background (**Supplementary Fig. 3.14b**), which results in a low signal-to-noise ratio. We hypothesized that both limitations could be resolved by reducing the concentration of the beads used for RNA capture. To test this hypothesis, we titrated Csm-beads in the pull-down and then combined the polymerization reaction and AaCan2 nuclease into a single step (**Supplementary Fig. 3.14 c, d)**. Reducing the concentration of Csm-beads 100-fold rescues fluorescent signal, improves signal-to-noise and streamlines the assay by eliminating the need for additional sample handling.

Next, we tested six lysis buffers supplemented with detergents (i.e., Triton X-100 or NP-40), and chelating agents (i.e., EDTA or EGTA), to identify lysis conditions compatible with the two-step detection assay that relies on Csm-based RNA capture, followed by detection with AaCan2. The lysis buffers tested do not inhibit the assays with purified RNA and increase the signal-to-noise ratio (**Supplementary Fig. 3.15**). This improvement may be a consequence of the detergents, which according to the manufacturer, reduce aggregation of the magnetic beads. Further, when we titrate the viral RNA, the optimized two-step assay generates a fluorescent

signal that is significantly different from the negative control (293T RNA) at target concentrations ≥8.3 fM (5×10^3copies/μL, **Figure 3.5a - c**).

To determine the sensitivity of direct SARS-CoV-2 RNA detection in patient samples, we used two unprocessed SARS-CoV-2 positive patient samples that were serially diluted in a negative patient swab sample (**Fig. 5d**). Each sample was divided in two parts. One part (120 μL) was tested directly with the two-step type III RNA detection assay and the other part (125 μL) was used for RNA extraction. Purified RNA was tested with the two-step type III detection protocol (120 μL) and RT-qPCR (5 μL) (**Fig. 5d**). Csm-based RNA capture assay detects SARS-CoV-2 RNA in both unprocessed patient swabs and purified RNAs with Ct <24, which corresponds to ~$8.7x10^3$ copies/μL (~15 fM) (**Fig.** 3.**5e**).

Finally, to test cross-reactivity of TtCsm-based RNA capture coupled with AaCan2-based fluorescent detection, we use a panel of seven respiratory viruses, including SARS-CoV-1, Middle East respiratory syndrome coronavirus (MERS-CoV), seasonal coronaviruses 229E (HCoV-229E), NL63 (HCoV-NL63), and HKU1 (HCoV-HKU1). To perform the assay, we used the highest concentration of genomic RNAs available at ATCC and recommended by FDA (~ $5x10^5$ copies/μL)[38]. No cross-reactivity is detectable with RNAs from Influenza B, Human respiratory syncytial virus (RSV), and HCoV-229E (**Fig. 3.5f**). However, we found that reactions with SARS-CoV-1, MERS-CoV, HCoV-HKU1, and HCoV-NL63 RNAs generate a weak signal that is slightly higher than the threshold in negative control samples. Importantly, this signal is significantly lower than the signal obtained for SARS-CoV-2 RNA used at the same concentration ($p < 0.001$, one-way ANOVA with post-hoc Tukey HSD test).

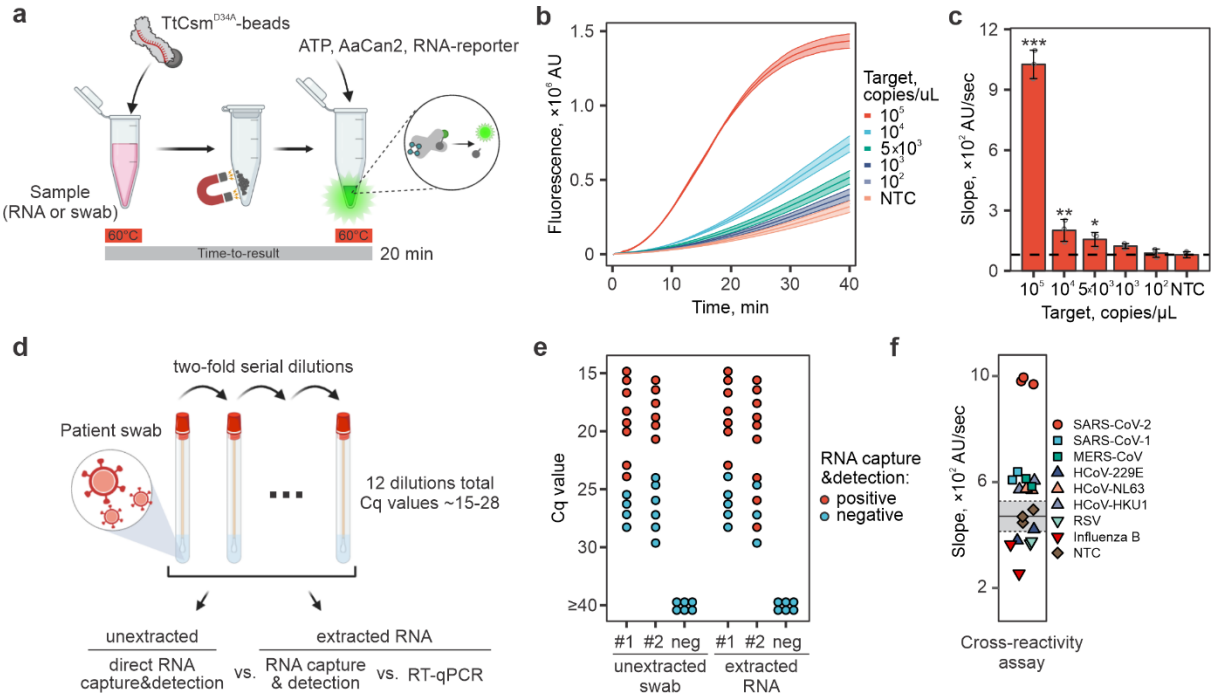


Fig.3.5: TtCsm-based RNA capture directly detects SARS-CoV-2 in clinical samples. **a** Schematic of TtCsm-based RNA capture assay from patient sample coupled with AaCan2-based fluorescent detection. Two-step protocol includes incubation of TtCsm-beads with RNA sample at 60°C followed by concentration of the magnetic beads, and addition of a buffer containing ATP, AaCan2 and a fluorescent RNA reporter (i.e., FAM-RNA-Iowa Black FQ). b RNA detection assay as shown in (a) was performed using samples with target RNA concentrations ranging from 107 to 102 copies/µL. Samples were prepared by spiking IVT fragments of SARS-CoV-2 N gene into in total human RNA (HEK 293T cells). Cleavage of fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument (left). Data were plotted as mean of three replicates ± S.D. (ribbon). c Simple linear regression was used to calculate slopes for linear regions of the curves. Bars show mean values (n = 3) ± S.D. (right). Data was analyzed with one-way ANOVA followed by multiple comparisons to NTC sample using one-tailed post-hoc Dunnett's test. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. d Patient samples (e.g., nasopharyngeal swabs) tested positive for SARS-CoV-2 RNA with RT-qPCR were serially diluted in a negative patient swab sample and dilutions were tested with three different methods. Non-processed swabs were tested directly with TtCsm-based RNA capture & detection assay, while purified RNAs were tested with RT-qPCR and TtCsm-based RNA capture & detection. e Twenty-four nasopharyngeal mock samples were prepared from two patient swabs (#1, Ct~14.8 and #2, Ct~15.6) and tested as shown in (d). Red dots show samples that tested positive in type III detection, blue dots show samples that tested negative. f RNAs of common respiratory viruses and coronaviruses related to SARS-CoV-2 (5x105 copies/µL) were tested with TtCsm-AaCan2 RNA capture and detection assay (three-step protocol). Reaction kinetics were measured using a real-time PCR instrument and slopes were quantified in the linear range of the fluorescence curves. Dotted line shows mean of three negative samples (NTC). Mean ± 2.33 S.D. for negative samples is shown with gray rectangle.

Discussion

Development of CRISPR-based diagnostics has primarily focused on type V (Cas12) and type VI (Cas13) CRISPR systems, and the sensitivity of these techniques has improved from picomolar[39] to attomolar concentrations[30] over the last six years. However, most CRISPR-based viral diagnostics described to date still require nucleic acid extraction and pre-amplification to reach clinically relevant sensitivities[4].

While sensitivity continues to improve, less attention and little progress has been made to develop methods that bypass RNA extraction. Purifying nucleic acids from patient samples prior to testing requires specialized equipment that increases the cost, labor, and time-to-result. Thus, testing protocols that require RNA extraction represent a major limitation for point-of-care implementation, which is critical for limiting the spread of a contagious agent. Here, we demonstrate how type III CRISPR systems can be used for sequence-specific capture and detection of viral RNA directly from unprocessed patient samples without pre-amplification. In 2021, the first attempts to repurpose type III CRISPR systems for SARS-CoV-2 diagnostics achieved 0.1 – 1 nM sensitivity of RNA detection without pre-amplification[18,19]. More recent improvements using different type III complexes or different ancillary nucleases have been used to detect spiked SARS-CoV-2 RNA with ~2-4 fM sensitivity in 30 min[20,26] (**Supplementary Fig. 3.17**). However, the reported time-to-result for most CRISPR diagnostics does not account for the time (i.e., 20-40 min) or cost (i.e., $5-7) of extracting RNA, which varies depending on the technique, availability of reagents, accessibility of high complexity laboratories, reliability of shipping, and the number of samples. Here, we contribute to the ongoing development of type III-based diagnostics by developing a method for sequence-specific capture and concentration of

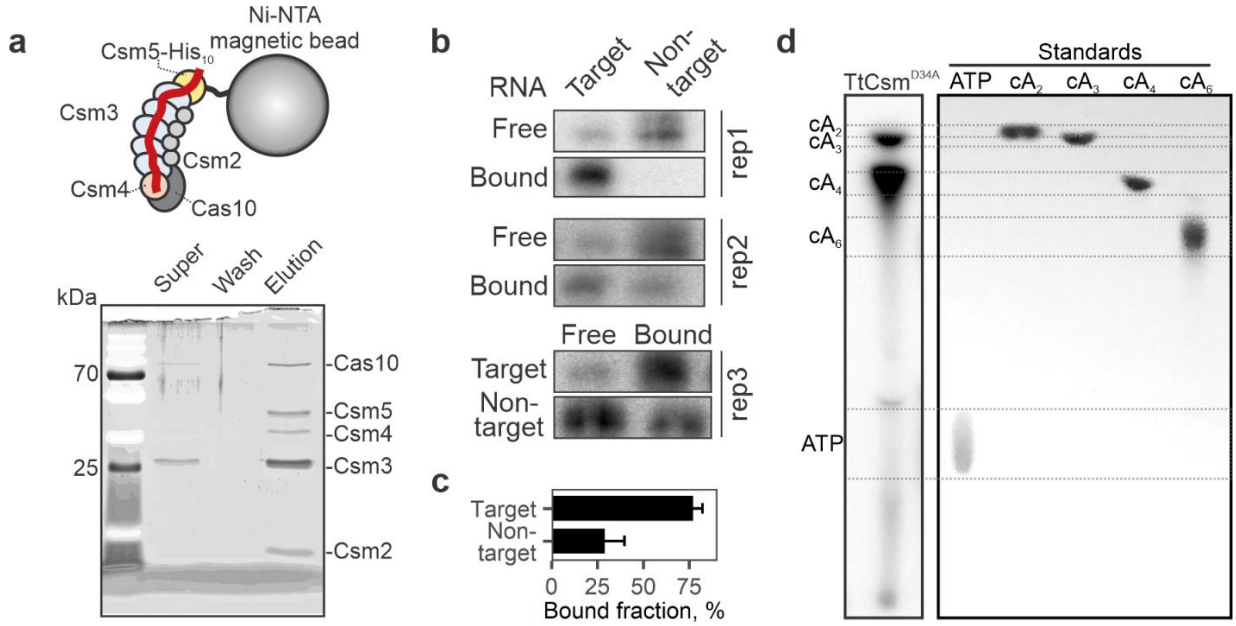
target RNAs directly from unprocessed patient samples. This approach enables direct detection of $8.7x10^3$ copies of SARS-CoV-2 RNA per µL (~15 fM) in clinical samples without laboratory-based RNA extraction or pre-amplification in less than 20 min total. While the sensitivity of the approach is still inferior to RT-qPCR, it is sufficient to identify infected individuals capable of spreading SARS-CoV-2[40] and is comparable to rapid antigen tests[2].

Like Cas13, type III systems also recognize RNA, and the most sensitive detection methods developed to date for either approach rely on collateral nuclease activity to release a fluorescent signal[4]. While methods that incorporate tandem CRISPR nucleases (i.e., Cas13 and Csm6) are currently more sensitive (~50 aM), the intrinsic amplification of RNA recognition by type III systems may ultimately improve sensitivity. Type III systems uniquely amplify RNA recognition in two sequential steps: first, through Cas10-mediated polymerization of cOAs, and second, through cOA-mediated activation of multi-turnover effectors (e.g., Can). In addition to the advantages that might come from consecutive stages of signal amplification, the separation of target recognition by the type III surveillance complex (i.e., Csm or Cmr) from collateral cleavage by ancillary effectors also enables programmable RNA capture. Unlike Cas13, which relies on the same active site for target and non-target collateral cleavage[41], the RNase-dead TtCsm complex ($TtCsm^{Csm3\text{-}D34A}$) can capture and maintain target RNA from a larger volume and concentrate these RNAs for various downstream applications. Incorporating RNA capture increases the sensitivity of type III CRISPR-based diagnostic and allows direct detection in clinical samples without RNA extraction, a prerequisite for most current platforms. Further, the separation of on-target (cis-) and collateral (trans-) activities in type III CRISPR systems provides a flexible platform for creating combinatorial assemblies of detection (i.e., Csm/Cmr

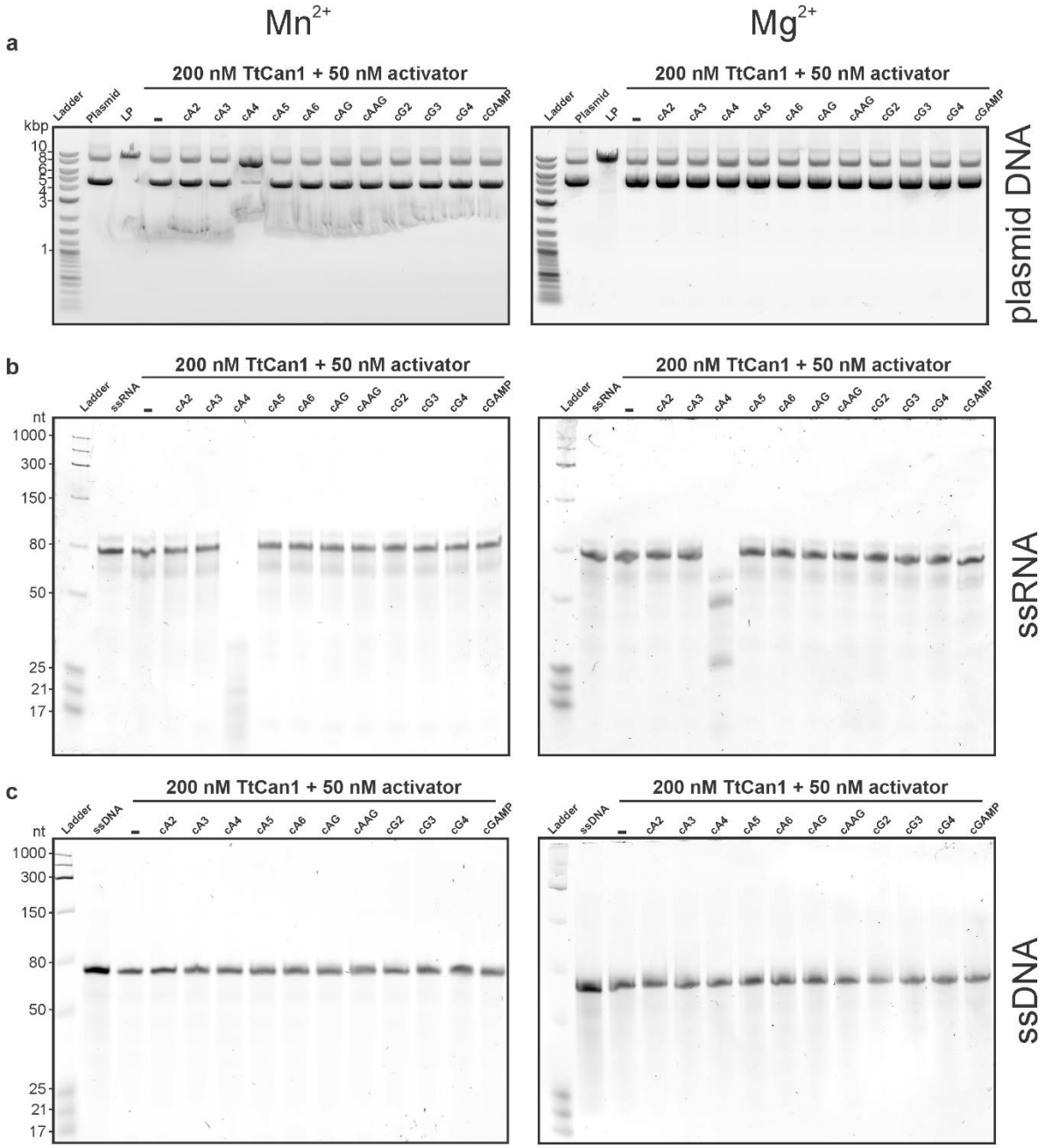
complexes) and readout (i.e., ancillary effectors) modules based on diverse CARF-, SAVED-(SMODS-associated and fused to various effector domains)[16] or CBASS (Cyclic oligonucleotide-based antiphage signaling system) proteins for applications in diagnostics and beyond.

Direct detection of RNA in clinical samples without RNA extraction or pre-amplification will advance direct-to-consumer diagnostics. However, eliminating pre-amplification removes additional selectivity imparted by the traditional use of primers. To evaluate cross-reactivity of TtCsm-based RNA capture coupled with AaCan2-based fluorescent detection, we tested a panel of seven respiratory viruses. At high RNA concentrations ($5x10^5$ copies/μL) we detect a weak, but reproducible signal to some of the coronaviruses (i.e., SARS-CoV-1, MERS-CoV, HCoV-HKU1, and HCoV-NL63) but not to HCoV-229E or other respiratory viruses (i.e., RSV, Influenza B). Each of the coronaviruses have 4-7 mismatches in first two segments (S1 - S2) of crRNA:target duplex (**Supplementary Fig. 18**), which permit target RNA binding but significantly decreases polymerization activity of the Cas10 subunit[19]. Coronaviruses that generate a weak signal have two mismatches in the first segment (S1; nucleotides +2 and +5), while HCoV-229E, which generates no signal, has an additional mismatch at the 1st position. The first two nucleotides in segment S1 have the greatest effect on cOA production[19,26,42], therefore the additional mismatch in HCoV-229E at the 1st position might explain the complete loss of signal, as compared to the weak signal generated by the other coronaviruses which contain two mismatches in segment 1 (S1). Our original bioinformatic pipeline for crRNA design filtered out guides with potential cross-reactivity, however we prioritized the total number of mismatches in segments S1 and S2 rather than the position of these mismatches[18]. The cross-

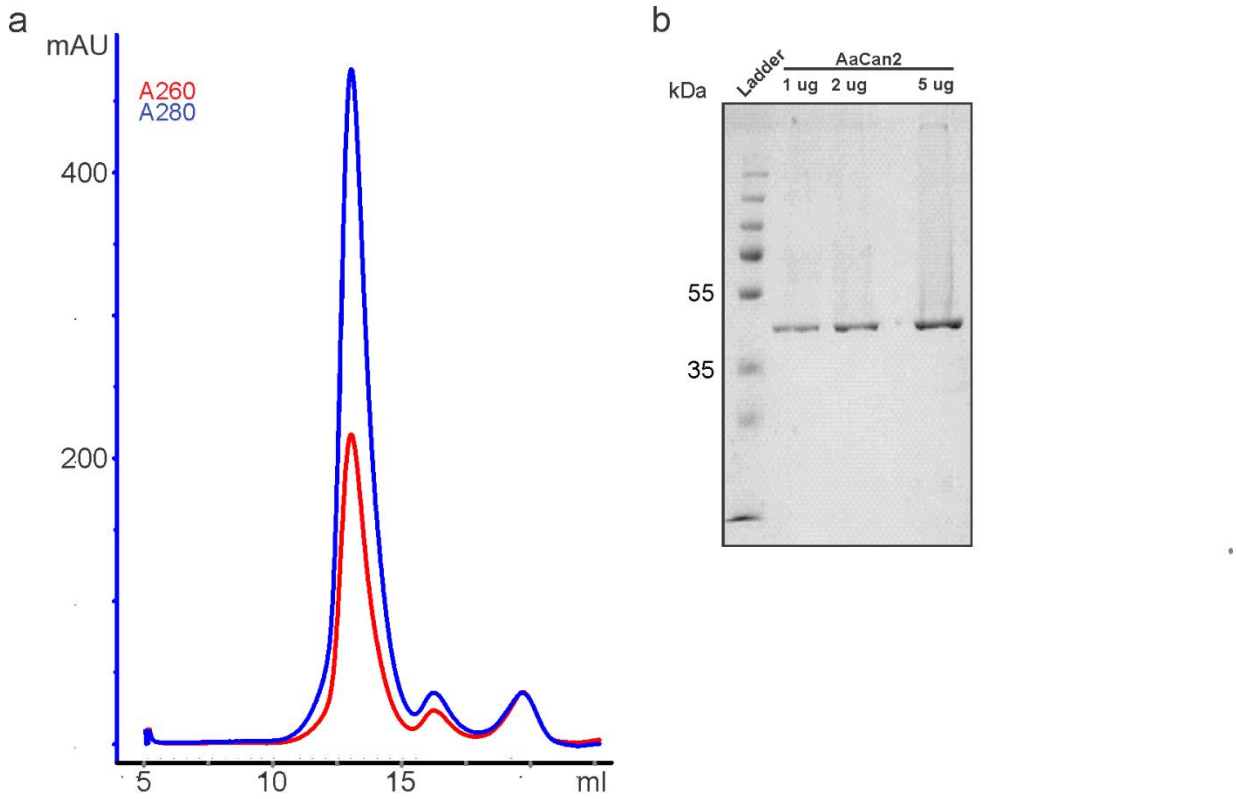
reactivity assay in this manuscript and published biochemical data[19,26,42] demonstrate the importance of mismatch positions +1 and +2, which is expected to improve the specificity of the next generation of guides. Overall, we anticipate that type III-based RNA pull-down techniques that bypass RNA extraction, combined with further optimization of lysis conditions, more efficient guide design, and the integration of next generation of signal detection methods (e.g., real-time sequencing, digital enzymology, amperometry, etc.) will help bringing type III CRISPR diagnostics closer to current standards of rapid molecular testing.



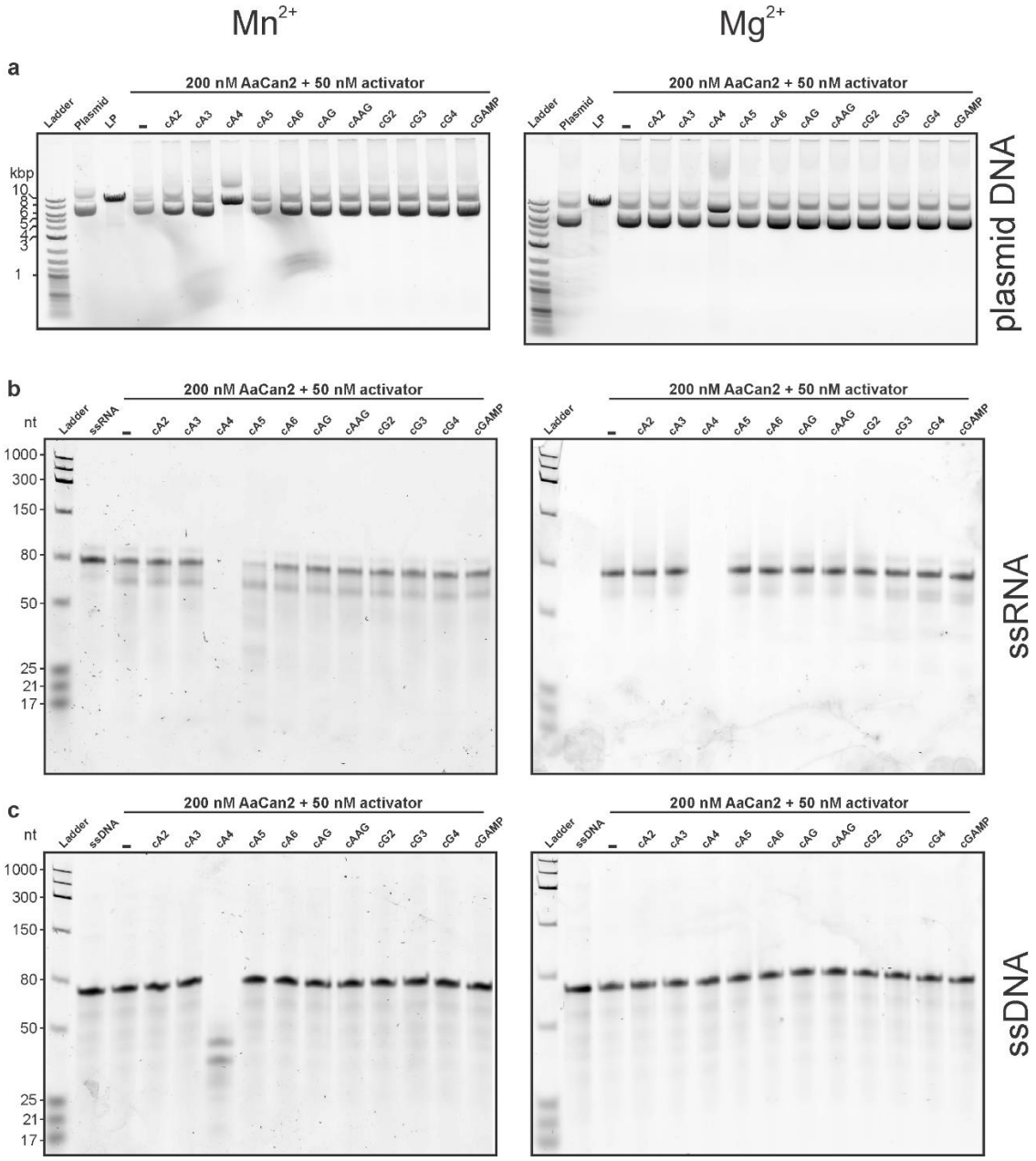
Supplementary Figure 3.1: Type III CRISPR-based RNA concentration enhances detection. **a** *Top panel:* schematic of TtCsm-complex bound to a Ni-NTA magnetic bead. *Bottom panel:* SDS-PAGE of TtCsm-complex after pull-down with Ni-NTA magnetic beads. After binding, beads were separated with a magnet, and supernatant ("Super") was decanted. The beads were washed once ("Wash"), and the protein was eluted using 250 mM imidazole ("Elution"). **b** Sequence-specific RNA enrichment with TtCsm$^{Csm3\text{-}D34A}$ complex was tested using ^{32}P 5`-end labeled IVT RNA fragments. Target and non-target RNA fragments were mixed with TtCsm$^{Csm3\text{-}D34A)}$ complex, incubated at 65°C for 20 min and pulled-down using a Ni-NTA magnetic beads. After the magnetic pull-down, RNA was phenol-chloroform extracted from supernatants ("Free") and bead pellets ("Bound"). Extracted RNA was resolved using UREA-PAGE, exposed to a phosphor screen, and imaged on a Typhoon 5 imager (Amersham). Three replicates are shown. **c** RNA bands shown in **b** were quantified using ImageJ and the bound fraction was calculated for each replicate as follows, [bound/(bound + free)*100%]. Bars show mean of 3 replicates ± S.E.M.; **d** TtCsm$^{Csm3\text{-}D34A}$ complex produces cA_3 and cA_4. TtCsm$^{Csm3\text{-}D34A}$ complex (25 nM) was mixed with target RNA (10^{11} copies/ul) to activate Cas10 subunit. Polymerization reaction was performed with α-^{32}P-ATP. Reaction products were phenol-chloroform extracted and resolved on the same thin-layer chromatography (TLC) plate with chemically synthesized standards. Radiolabeled products (left) were imaged with Typhoon 5 phosphor imager. Chemically synthesized standards (cA_2-cA_6) were visualized using UV shadowing (right).



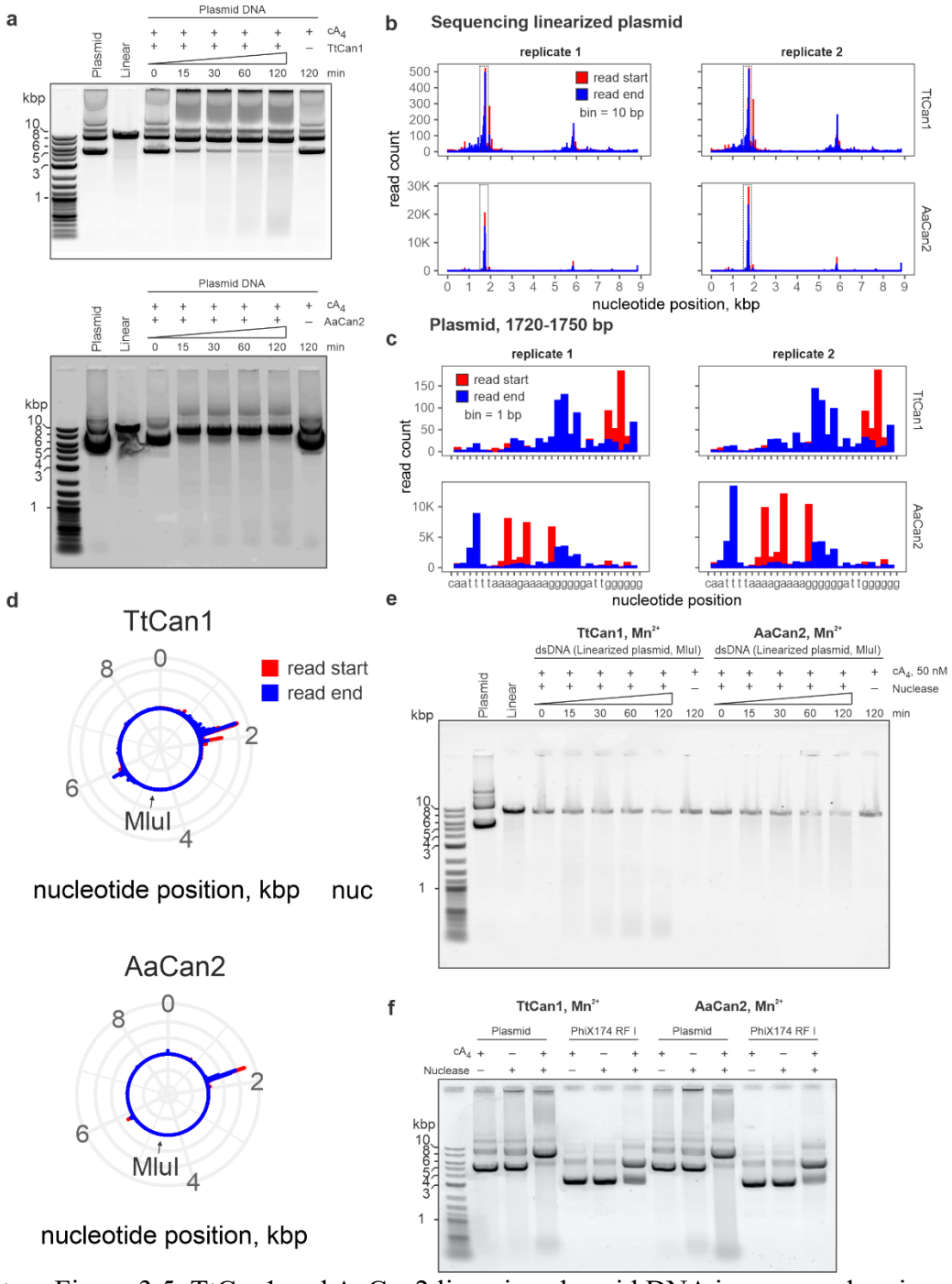
Supplementary Figure 3.2: Biochemical characterization of Thermus thermophilus Can1 (TtCan1) nuclease activities. Cleavage assays with TtCan1 (200 nM) in the presence of various cyclic oligonucleotides (50 nM) in Mn2+ or Mg2+ containing buffers. Reactions were performed with 15 nM plasmid DNA (**a**), 425 nM ssRNA (**b**) or 15 nM ssDNA (**c**), for 15 min at 60°C. Reactions with plasmid DNA were stopped by adding Gel Loading Dye, Purple (6X) (NEB) and visualized in 1% agarose gel. For ssDNA and ssRNA cleavage assays, 2X RNA Loading Dye (NEB) was added, and 10 µL was used for 12% UREA PAGE.



Supplementary Figure 3.3: Purification of Can2 from Archaeoglobi archaeon JdFR-42. **a** Size-exclusion chromatography (SEC) of *Archaeoglobi* archaeon Can2 protein (AaCan2). SEC was performed using a Superdex 200 10/300 GL size-exclusion column (Cytiva). **b** SDS-PAGE of purified AaCan2 in three different concentrations. Expected molecular weight is 42.8 kDa.

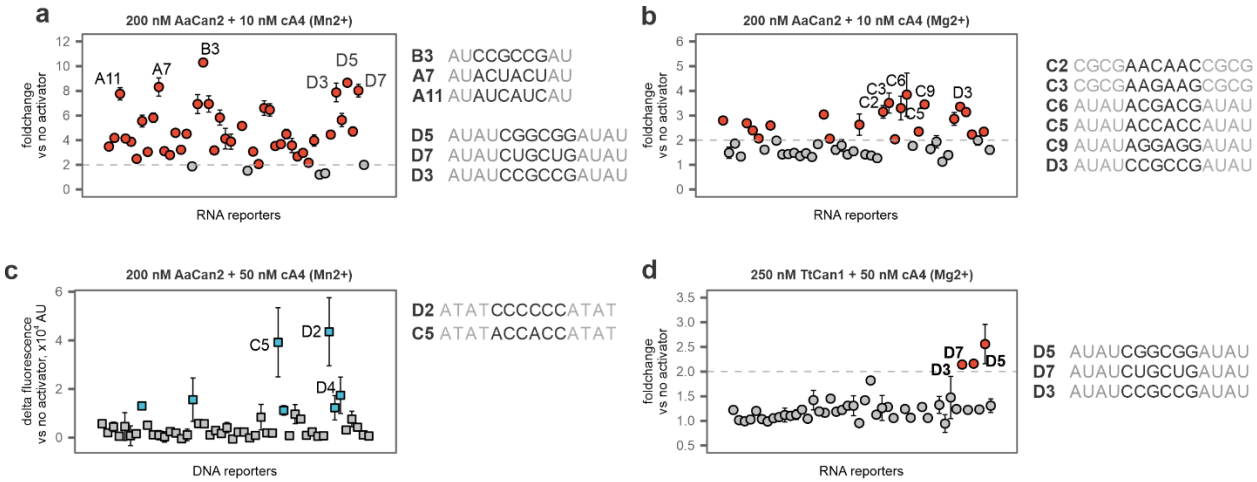


Supplementary Figure 3.4: Biochemical characterization of Can2 from Archaeoglobi archaeon JdFR-42. Cleavage assays with AaCan2 (200 nM) in the presence of various cyclic oligonucleotides (50 nM) in Mn2+ or Mg2+ containing buffers. Assays were performed with 15 nM plasmid DNA (**a**), 425 nM ssRNA (**b**) or 15 nM ssDNA (**c**), for 15 min at 60°C. Reactions with plasmid DNA were stopped by adding Gel Loading Dye, Purple (6X) (NEB) and were loaded on 1% agarose gel. For ssDNA and ssRNA cleavage assays, 2X RNA Loading Dye (NEB) was added and 10 µL was loaded on 12% UREA PAGE.

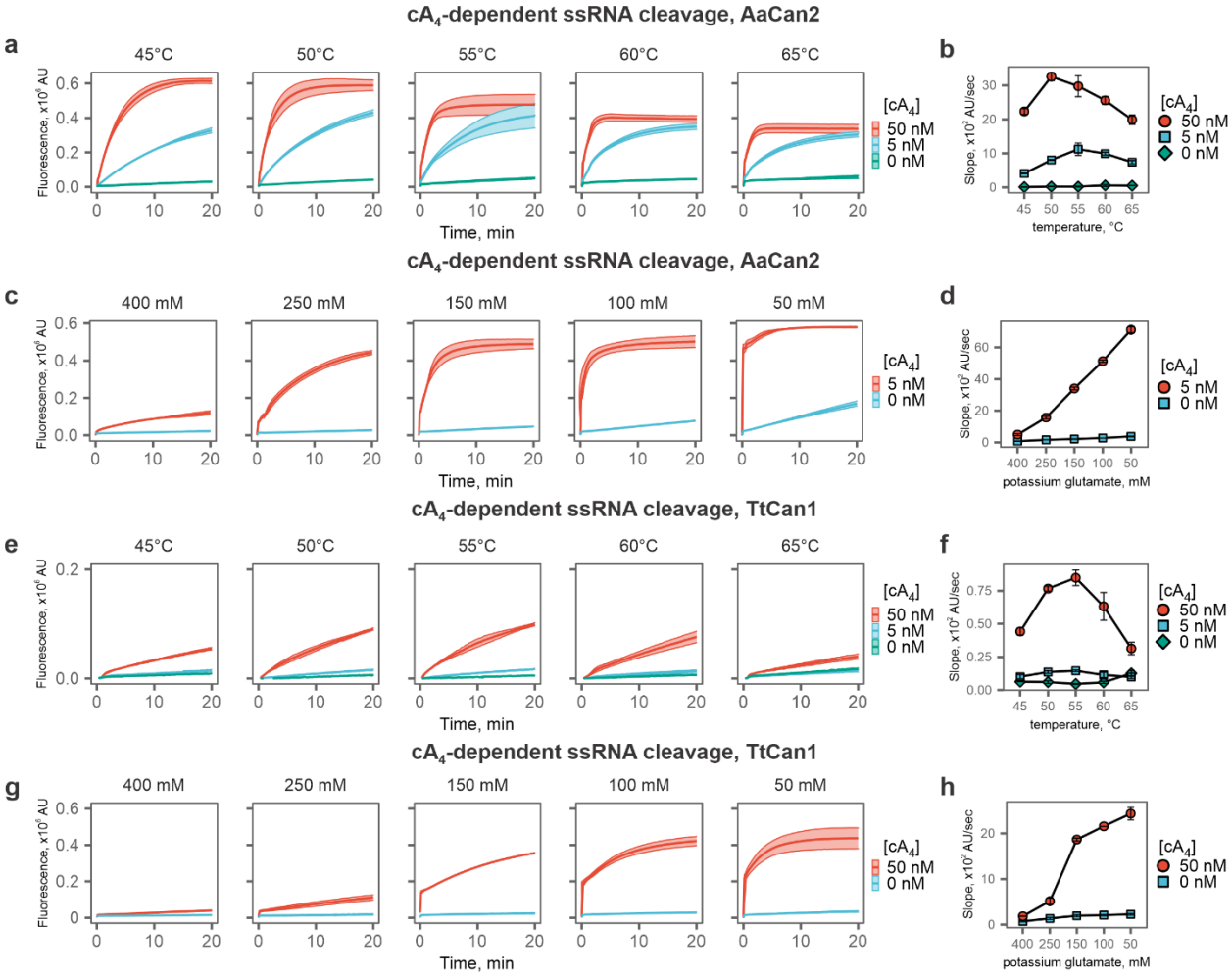


Supplementary Figure 3.5: TtCan1 and AaCan2 linearize plasmid DNA in unwound regions. a Time-course of plasmid DNA (Addgene #105621) incubated with TtCan1 (top, 200 nM) and AaCan2 (bottom, 200 nM) in the presence of cA4 (50 nM). b Plasmid DNA (Addgene #105621, 8,833 bp) was incubated with TtCan1 or AaCan2 in the presence of cA4 for 15 min. DNA was purified using Clean-up and Concentration kit (Zymo Research), end-repaired and sequenced using Oxford Nanopore platform. Sequencing reads (≥8,800) were mapped to the reference sequence. Start and end positions of each read were extracted with bedtools and plotted with

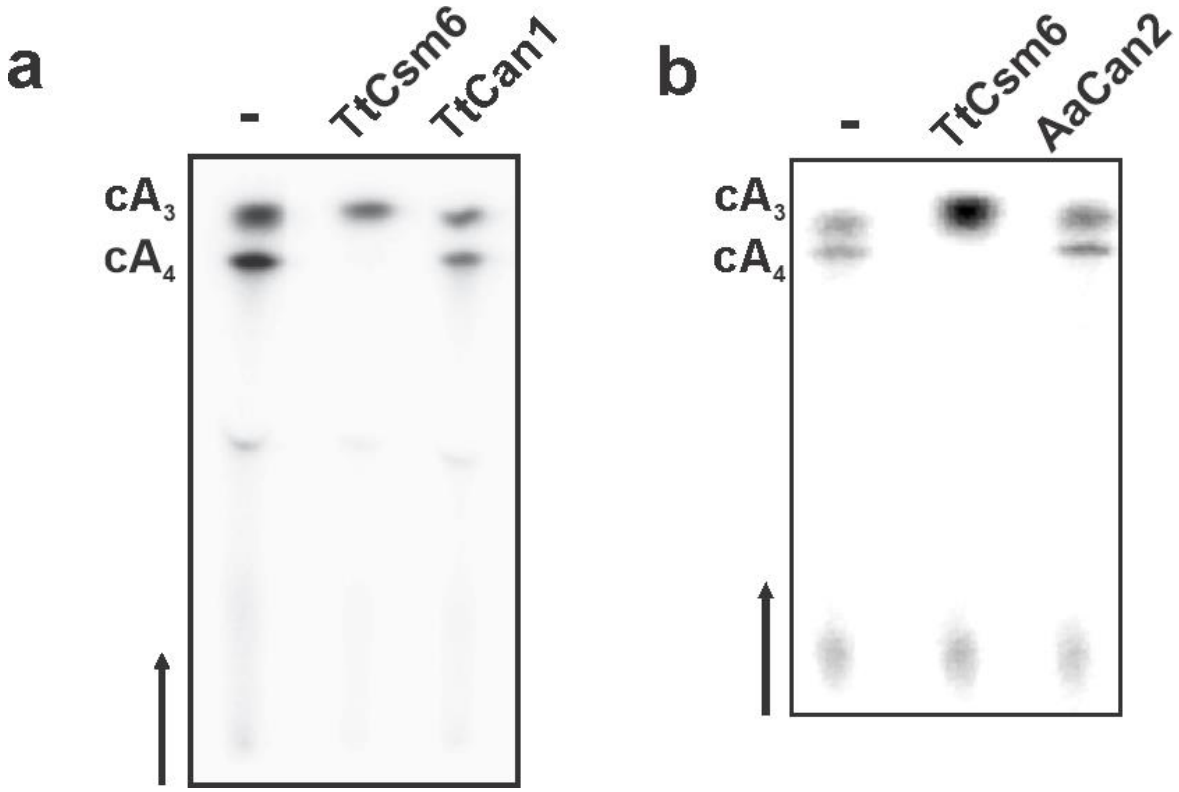
ggplot. Bars show number of reads that start (red) or end (blue) at the specific nucleotide position (bin width = 10 bp). c Region with the highest cleavage frequency (1,720-1,750 bp; dotted box in b) is shown. d - e Plasmid DNA (Addgene #105621) was linearized using MluI restriction nuclease, purified with DNA Clean-up and Concentration kit (Zymo Research), incubated with TtCan1 (200 nM) or AaCan2 (200 nM) in the presence of cA4 for 0, 15, 30 60 or 120 min, and visualized using agarose gel electrophoresis. f Plasmid DNA (Addgene, #105621) or the replicative form (RF I, NEB N3021) of ΦX174 bacteriophage DNA (which is supercoiled double-stranded DNA) were incubated with TtCan1 (200 nM, left) or AaCan2 (200 nM, right) in the presence of cA4 (50 nM) for 15 min.



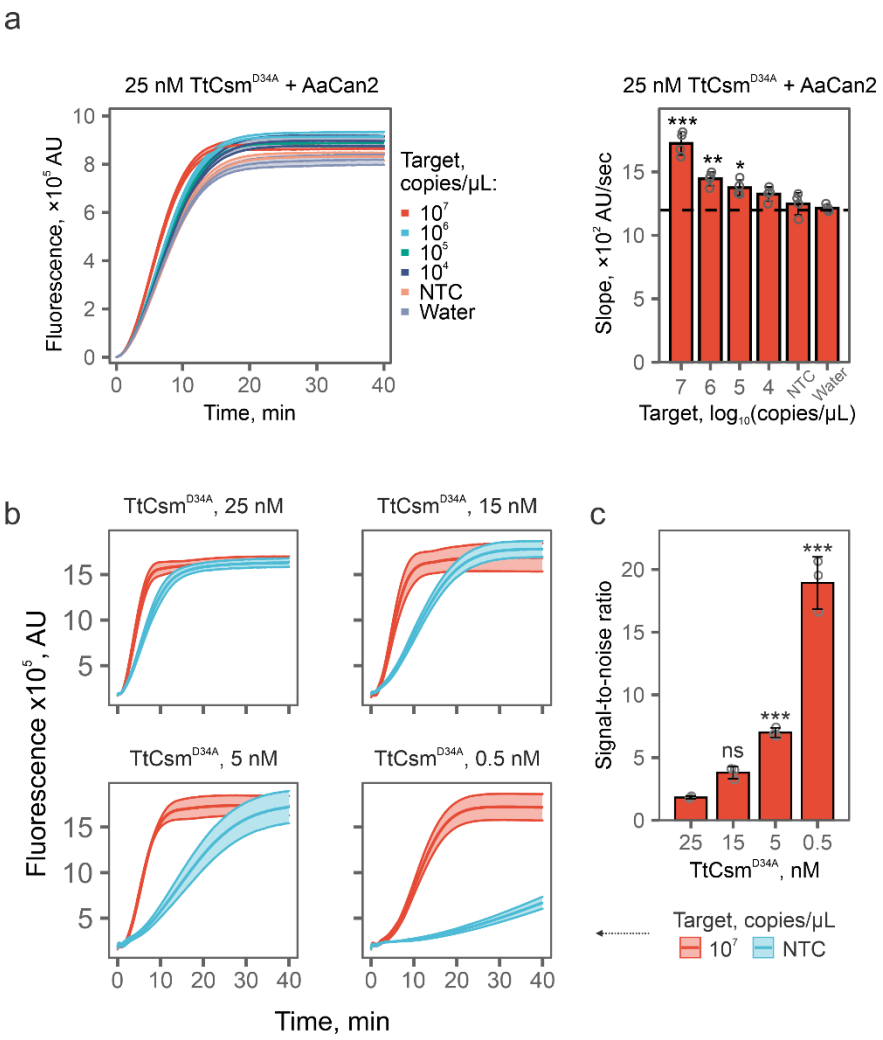
Supplementary Figure 3.6: Sequence cleavage preference for AaCan2 and TtCan1. A library of 46 synthetic RNAs or DNAs tethered to a 5'fluorophore and 3'quencher (see Methods) was screened with AaCan2 (a, b, c) and TtCan1 (d, only RNA reporters) in the presence of cA4. Reactions were incubated at 55°C in a qPCR instrument with fluorescent readings taken every 10 sec. Fluorescence at 10 min incubation time was normalized to the no activator control. Data are shown as mean ± S.D (n=3).



Supplementary Figure 3.7: TtCan1 and AaCan2 nuclease activities at different temperatures and salt concentrations. (a – d) AaCan2 (300 nM) and (e – f) TtCan1 (300 nM) cleavage assays with ssRNA fluorescent reporter (300 nM, D7 see Supplementary Table 1) in the presence cA4 activator (0, 5 or 50 nM) at different temperatures (45-65°C, 250 mM monopotassium glutamate) or salt concentrations (50-400 mM monopotassium glutamate, 55°C). Data is shown as mean (center line) of two replicates ± S.D. (ribbon) (a, c). Simple linear regression was used to calculate slopes for linear regions of the curves (b, d). Dots show mean values (n = 2) ± S.D.

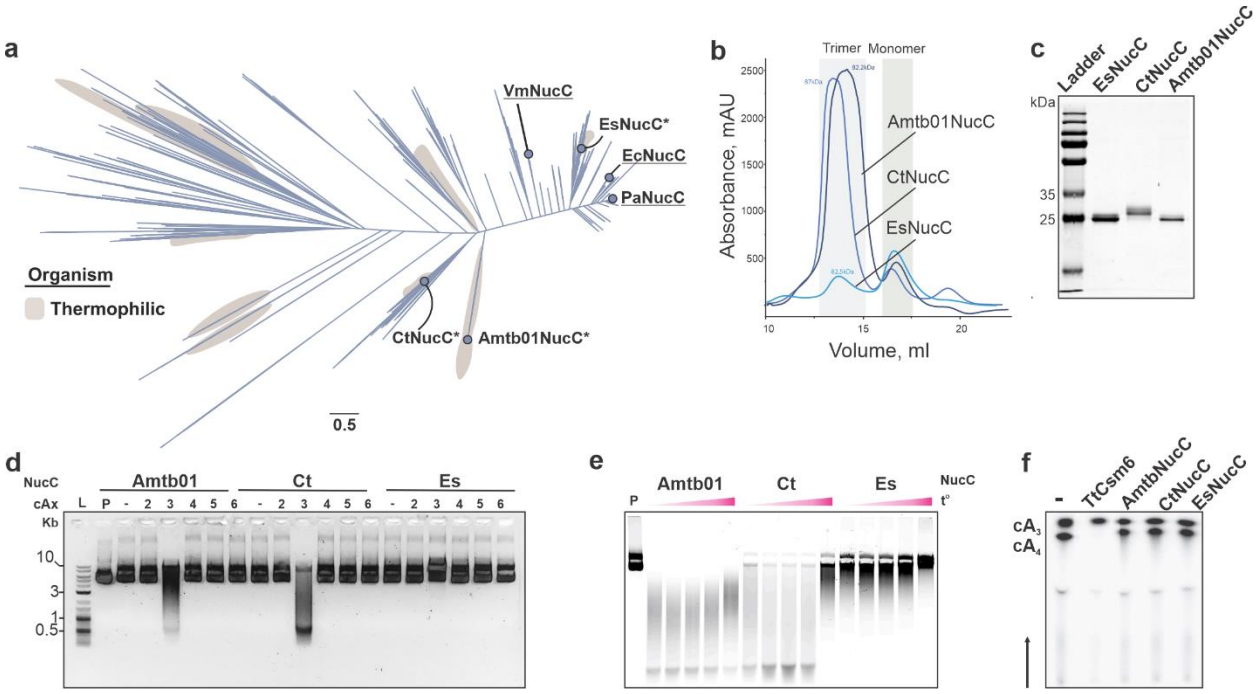


Supplementary Figure 3.8: TtCan1 and AaCan2 do not degrade cA3 or cA4. α-32P-labeled cA3 and cA4 were produced using TtCsmCsm3-D34A complex in the presence of target RNA. Radiolabeled cA3 and cA4 were extracted using phenol-chloroform and mixed with TtCsm6, TtCan1 (500 nM) (a) or AaCan2 (500 nM) (b) and incubated for 1 hour at 55°C. Reaction products were phenol-chloroform extracted and resolved using thin-layer chromatography (TLC). TtCsm6 cleaves cA4. TtCan1 and AaCan2 degrade neither cA3 nor cA4.

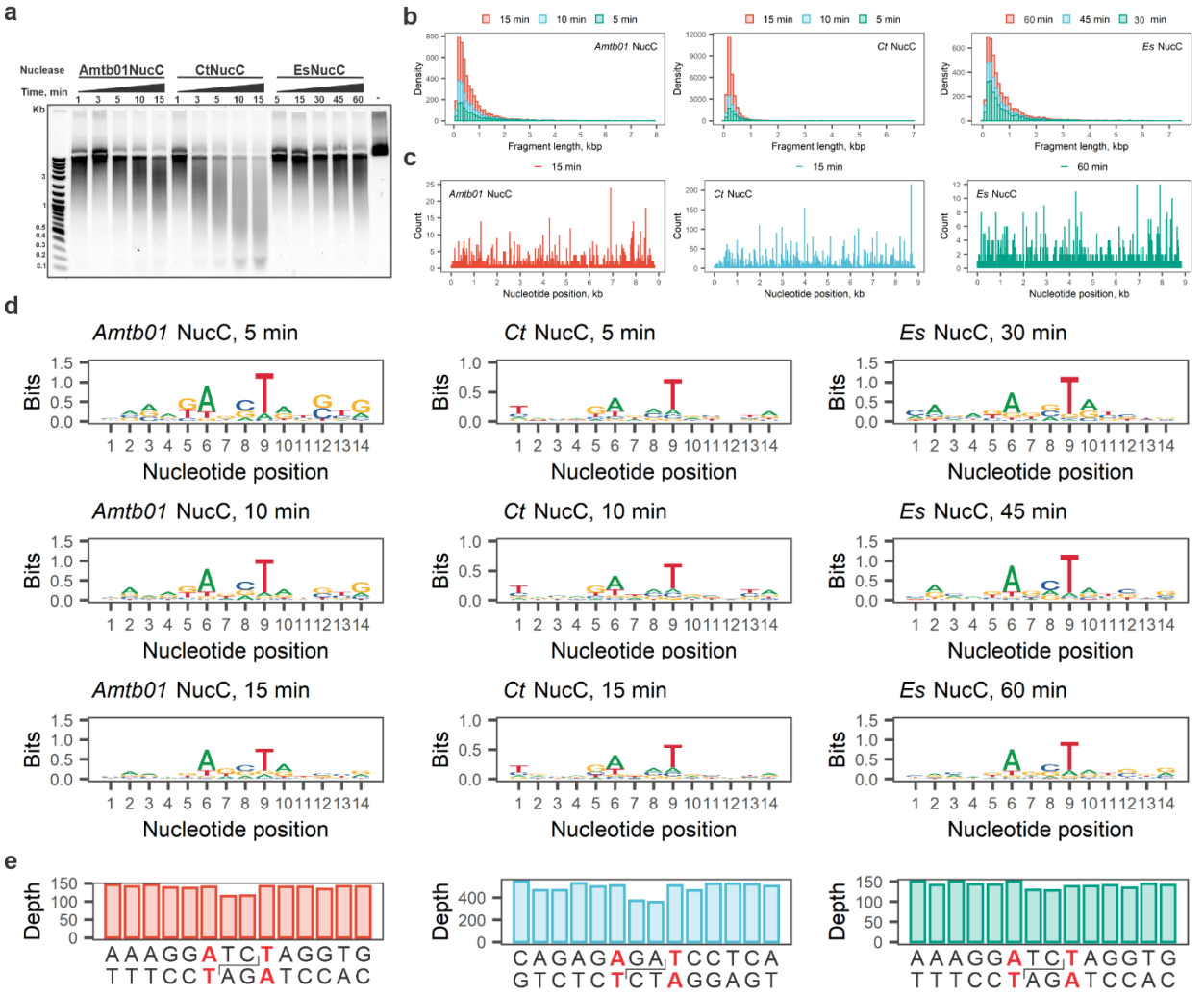


Supplementary Figure 3.9: Decreasing TtCsm-complex concentration in RNA detection assay reduces the background signal. a TtCsm RNA detection assays coupled with AaCan2-based fluorometric detection were performed using samples with target RNA concentrations ranging from 107 to 102 copies/µL. Samples were prepared by spiking IVT fragments of the SARS-CoV-2 N gene into total RNA extracted from a clinical sample that was negative for SARS-CoV-2 by qRT-PCR. Cleavage of the fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument. Left: Data were plotted as mean (centerline) of 4 replicates ± S.D. (ribbon). Right: Simple linear regression was used to calculate slopes for linear regions of the curves (at 10 min). Bars show mean values (n = 4) ± S.D. (right). Data was analyzed with one-way ANOVA followed by multiple comparisons to the NTC sample using one-tailed post-hoc Dunnett's test. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. b Reactions were performed as described in a with varying TtCsm-complex concentrations (25 nM – 0.5 nM). Cleavage of fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument. Data is shown as mean (center line) of three replicates ± S.D. (ribbon). c Signal-to-noise ratios (y-axis) were quantified by dividing slope of fluorescence curves in the assays with target RNA (n= 3, shown in b) by mean (n = 3) slope of no target control (NTC). Data is plotted as mean (n = 3) ± S.D. and analyzed with one-way ANOVA. All

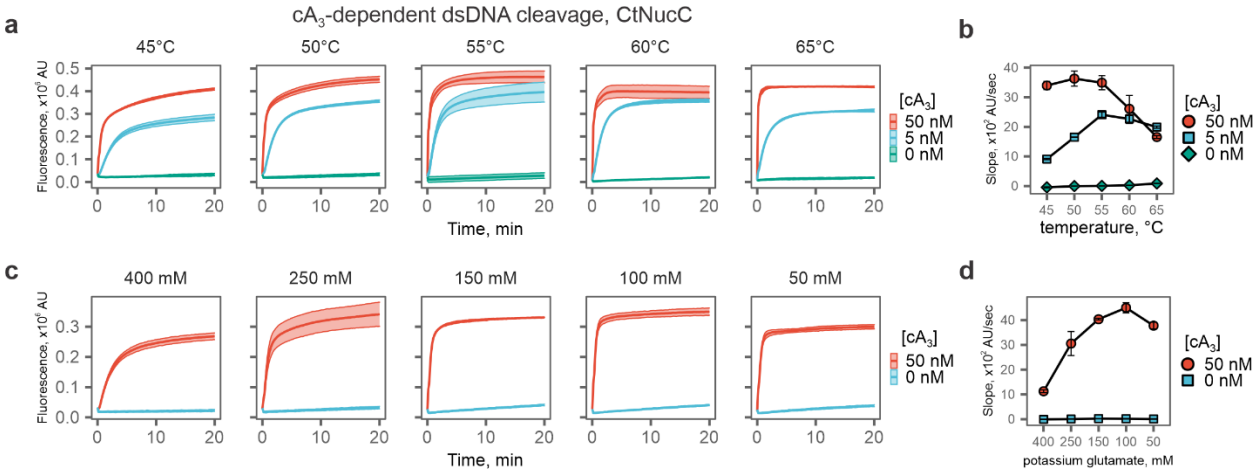
samples were compared to the reaction with 25 nM TtCsmD34A using two-sided post-hoc Dunnett's test. *** $p < 0.001$; ns – non-significant, $p > 0.05$.



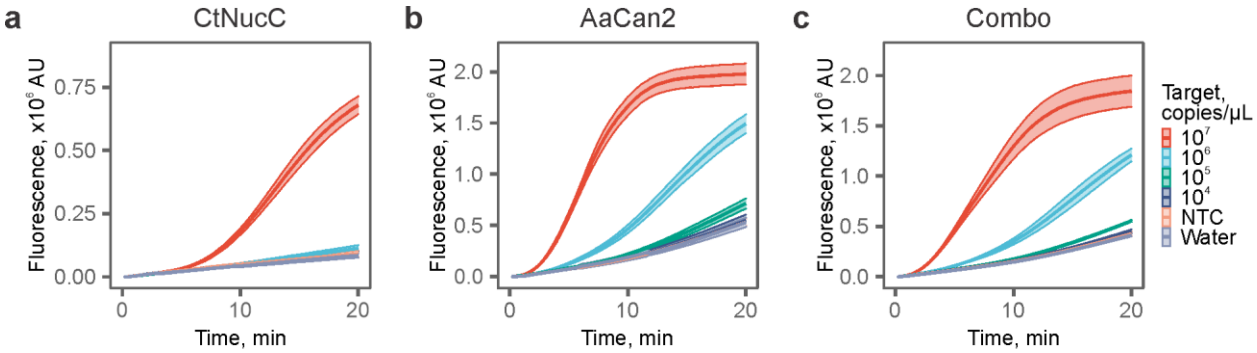
Supplementary Figure 3.10: Purification and biochemical characterization of thermostable NucC orthologs. **a** Maximum-likelihood phylogeny of 1,510 NucC orthologs. Previously studied effectors are underlined on the tree. Asterisks (*) identify effectors chosen for purification and *in vitro* experiments. **b** Size-exclusion chromatography (SEC) profiles of purified NucC orthologs. SEC was performed using a Sepharose 200 Increase 10/300 GL size-exclusion column (Cytiva). **c** SDS-PAGE of purified NucC orthologs. Expected molecular weight of EsNucC, CtNucC and Amtb01NucC are ~27.5kDA, ~29kDa and ~27.4kDa, respectively. **d** Plasmid DNA (Addgene #105621, 1 µg) digested with thermostable NucC orthologs (45 nM) in the presence of cyclic oligoadenylates with 2–6 bases (cA_2-cA_6, 45nM). L is the DNA ladder, P is plasmid DNA. **e** DNase activity of NucC orthologs was tested across a temperature range (45-60°C). The plasmid (1 µg) was mixed with 45 nM EsNucC + 45 nM cA_3, 15 nM CtNucC + 15 nM cA_3 or 45nM Amtb01NucC + 45 nM cA_3 for 15 min. **f** cA_3 and cA_4 hydrolysis assay for EsNucC, CtNucC and Amtb01NucC. The assay was performed as described in Supplementary Figure 8.



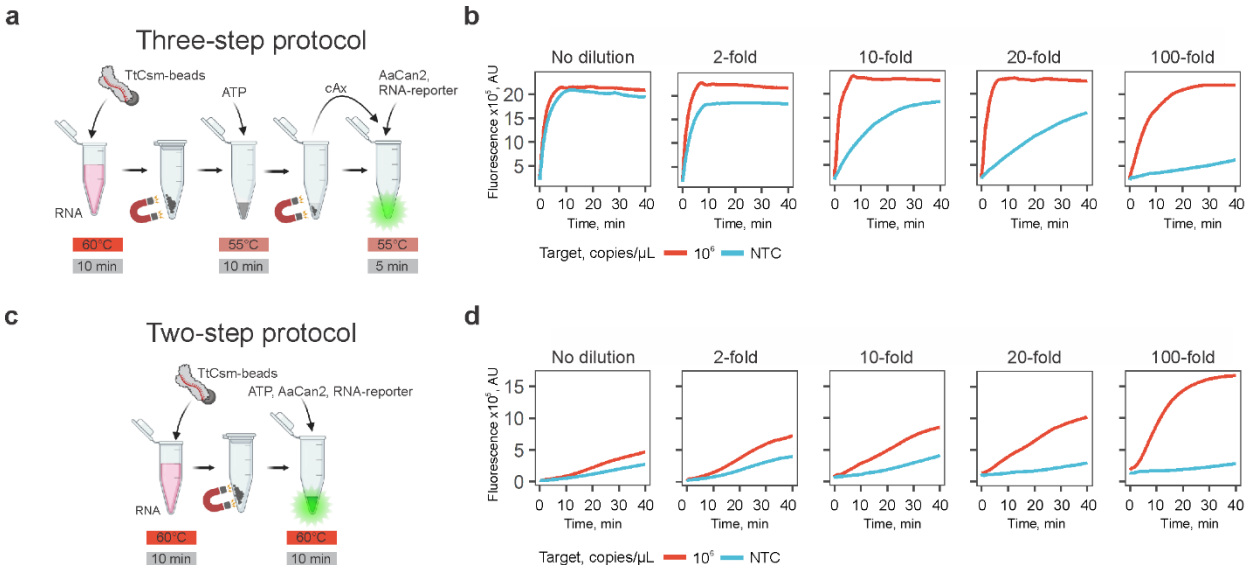
Supplementary Figure 3.11: Thermostable NucC homologs Amtb01NucC, EsNucC, and CtNucC demonstrate sequence cleavage preference of plasmid DNA. **a** Time-course plasmid DNA (Addgene #105621) digestion with thermostable NucC homologs in the presence of cA_3. The plasmid (1 µg) was mixed with EsNucC (45 nM) + cA_3 (45 nM), CtNucC (15 nM) + cA_3 (15 nM) or Amtb01NucC (45 nM) + cA_3 (45 nM) for different periods of time. **b** Plasmid DNA cleavage fragments shown in **a** were end-prepped and deep-sequenced using Oxford Nanopore. Plots show distribution of the read length generated from fragments after different digestion times. **c** Position of read ends (i.e., cut sites) were identified and counted to generate cleavage maps for each nuclease. Resulting maps show cut sites at the longest incubation time tested:15 min of digestion with Amtb01NucC and CtNucC, and after one hour – with EsNucC. **d** Seven nucleotide sequences from both ends of the 5% most frequently cut sites were extracted and used to generate sequence logos. Resulting logos indicate enrichment for A at -1 position (nucleotide 6) and T at +1 position (nucleotide 9). **e** Sequences of top cleaved site are shown for each nuclease. Sequencing depth shows a pattern (i.e., reduced sequencing depth at cut site) which is consistent with 3'-end overhang removal during end-prep step with T4 DNA polymerase when sequencing libraries were prepared.



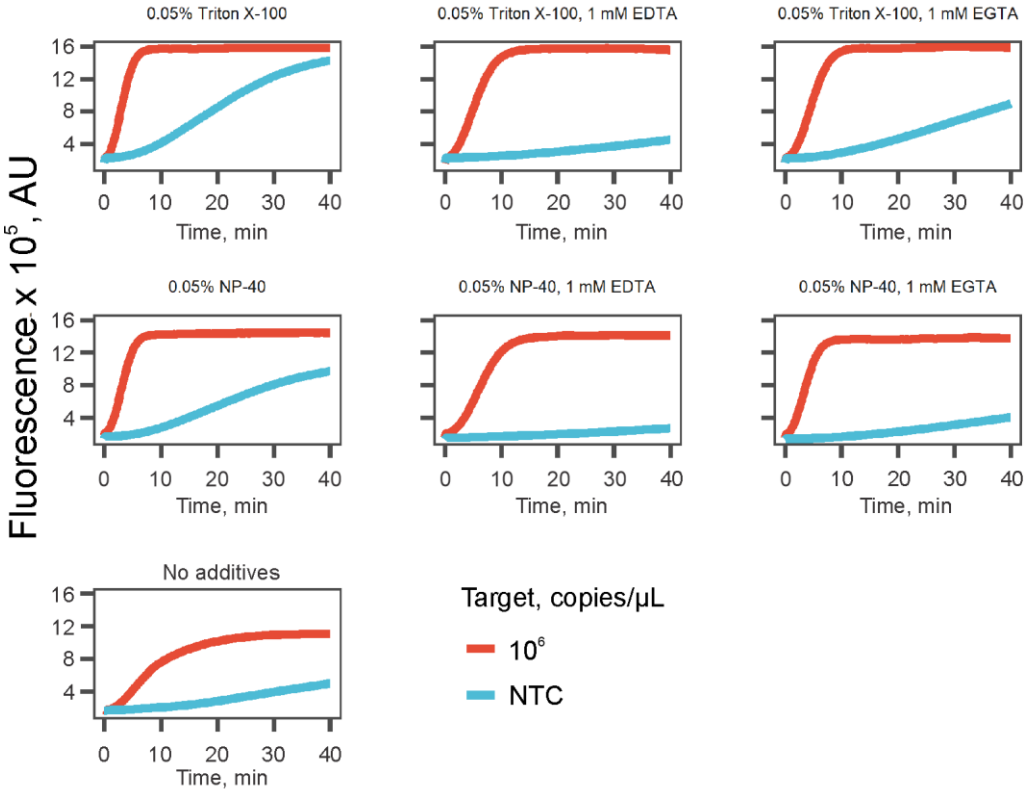
Supplementary Figure 3.12: CtNucC nuclease activity at different temperatures and salt concentrations. CtNucC (300 nM) cleavage assays with fluorescent dsDNA reporter in the presence cA_3 activator (5 nM or 50 nM) at different temperatures (45-65°C, 250 mM monopotassium glutamate) (**a, b**) and salt concentrations (50-400 mM monopotassium glutamate, 55°C) (**c, d**). Data is shown as mean (center line) of two replicates ± S.D. (ribbon) (**a, c**). Simple linear regression was used to calculate slopes for linear regions of the curves (**b, d**). Dots show mean values (n = 2) ± S.E.M.



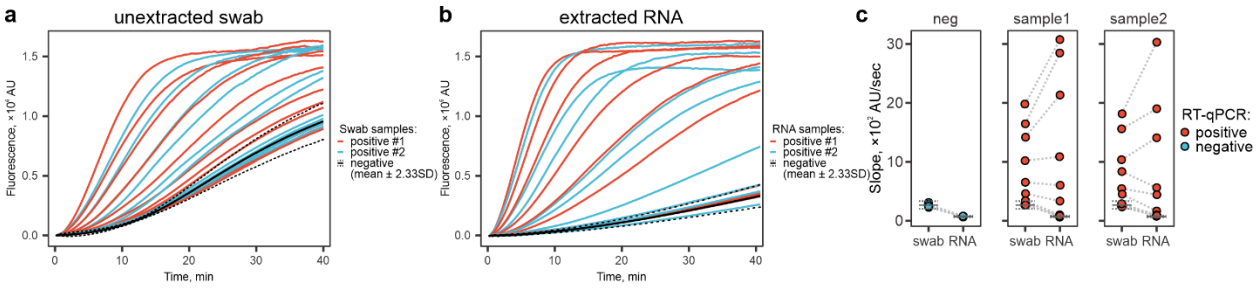
Supplementary Figure 3.13: Integration of cA3-activated nucleases into Csm-based RNA detection assay does not improve sensitivity. **a** TtCsm RNA detection assays coupled with AaCan2 (ssRNA reporter), CtNucC (dsDNA reporter) and combination of AaCan2 and CtNucC (both reporters). Reactions were performed using samples with target RNA concentration ranging from 10^7 to 10^4 copies/µL. Samples were prepared by spiking IVT fragment of SARS-CoV-2 N gene in total RNA of SARS-CoV-2 negative nasal swab. Cleavage of the fluorescent reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument. Data is shown as mean (center line) of four replicates ± S.D. (ribbon).



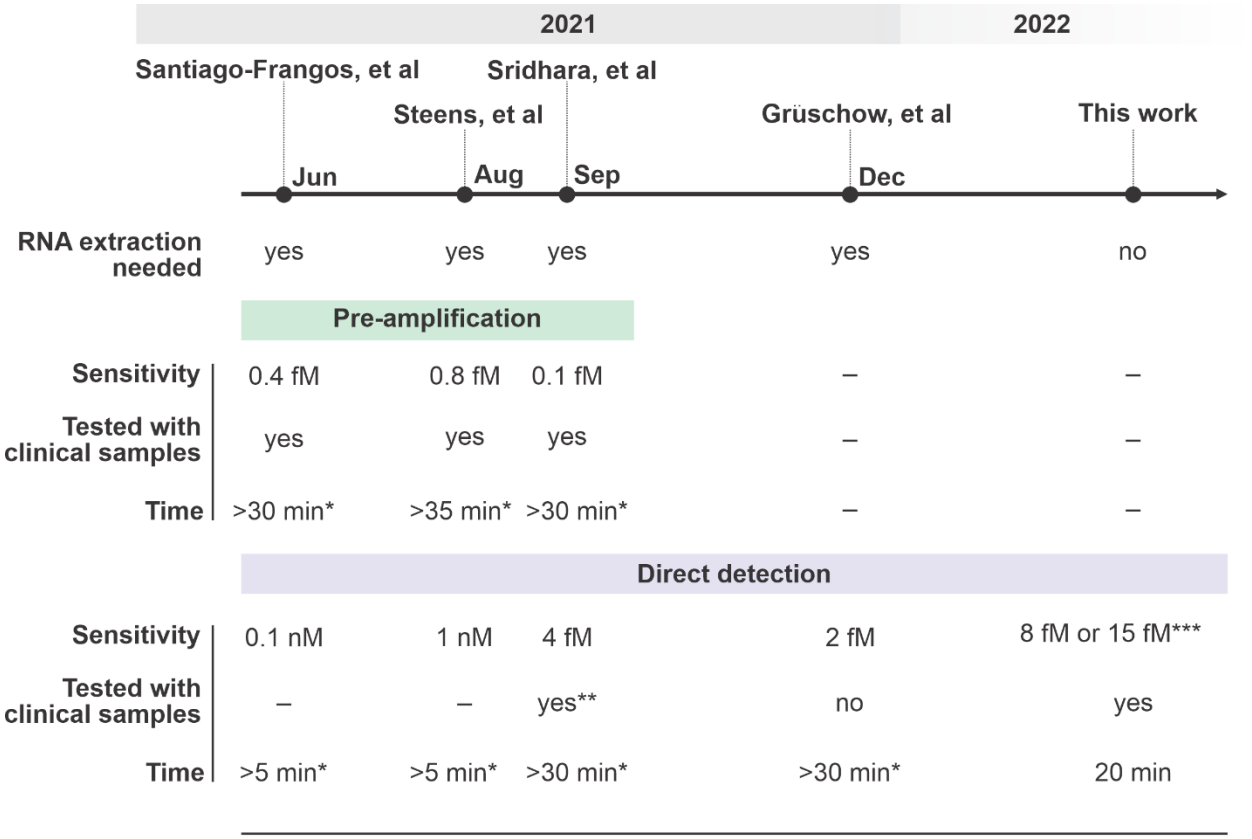
Supplementary Figure 3.14: Optimization of TtCsm-based RNA capture and detection assay. **a** Three-step protocol includes incubating TtCsm-beads with RNA sample at 60°C for 10 minutes (Step 1), concentrating the beads with a magnet, resuspending the beads in a buffer containing ATP, 10 min incubation (Step 2). Then concentrate the beads again and transferring the polymerization products (supernatant) to a reaction containing AaCan2 and a fluorescent RNA reporter (i.e., FAM-RNA-Iowa Black FQ) (Step 3). **b** Decreasing TtCsm-beads concentration increases signal-to-noise ratio. The reactions were performed as shown in **a** with varying Csm-beads dilutions in the pull-down step (Step 1). SARS-CoV-2 N-gene RNA diluted in total human RNA (HEK 293T cells) was used as the target. Cleavage of fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument. **c** The two-step protocol includes incubating TtCsm-beads with RNA sample at 60°C for 10 minutes (Step 1), concentrating the beads with a magnet, and one-pot polymerization and fluorescent detection (Step 2). In the Step 2, concentrated TtCsm-beads bound to the target RNA are resuspended in a buffer containing ATP, AaCan2 and a fluorescent RNA reporter (i.e., FAM-RNA-Iowa Black FQ). **d** Decreasing TtCsm-beads concentration rescues target-specific fluorescent signal and increases signal-to-noise ratio. The reactions were performed as shown in **c** with different Csm-beads dilutions in the Step 1. SARS-CoV-2 N-gene RNA diluted in total human RNA (HEK 293T cells) was used as the target. Cleavage of fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument.



Supplementary Figure 3.15: Testing effect of detergents and chelating agents on TtCsm-based RNA capture and detection assay. Csm-based target RNA concentration followed by one-pot cA_4 production (2-step protocol, **Supplementary Fig. 14c**) with fluorescent detection was performed from mock sample containing indicated additives. The mock sample was made by spiking SARS-CoV-2 RNA fragment (final concentration 10^6 copies/µL) into total human RNA (HEK 293T cells). Total human RNA was used as the no target control (NTC). Cleavage of fluorescent RNA reporter was detected by measuring fluorescence every 10 sec in a real-time PCR instrument.



Supplementary Figure 3.16: Direct detection of SARS-CoV-2 RNA in clinical samples using TtCsm-based RNA capture. **a, b** Two nasopharyngeal swabs were serially diluted in a negative patient swab sample (#1, Ct~14.8, red lines; #2, Ct~15.6, blue lines) and twenty-four dilutions in total were tested with type III RNA capture & detection assay directly (**a**) or from priorly purified RNA (**b**). Black solid line shows mean fluorescence in negative samples (n = 6), dotted black lines show ± 2.33 S.D. interval from mean. **c** Simple linear regression was used to calculate slopes for linear regions of the curves in **a** and **b**. Horizontal dotted lines show mean slope in negative samples (n = 6) ± 2.33 S.D. Reactions that generated signal higher than upper bound of this interval were considered positive for SARS-CoV-2 RNA in the type III CRISPR diagnostic assay. Dashed gray lines connect same samples that were either tested directly with type III RNA capture & detection assay (swab) or extracted (RNA) before type III RNA detection.

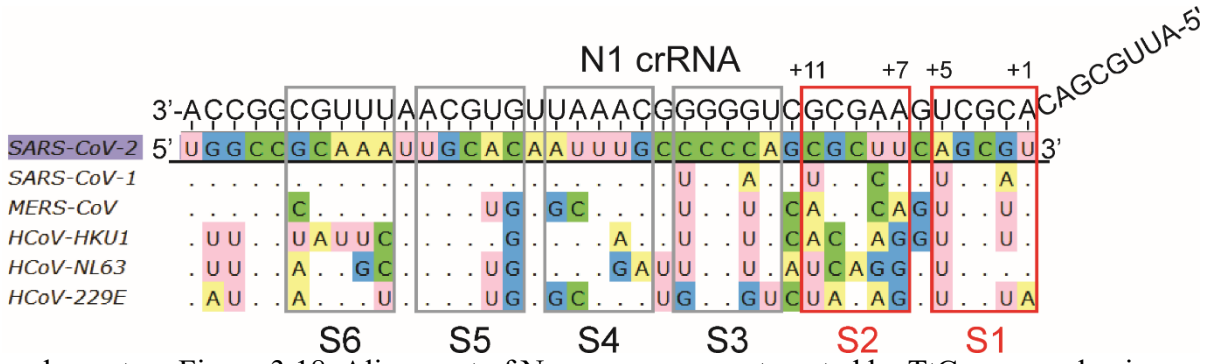


* Does not account for RNA extraction, which takes 20-40 min depending on the technique and the number of samples.

** Three clinical RNA samples with low Ct ≤17.2 tested positive and two samples with high Ct (≥29.2) tested negative. Samples close to the reported LoD were not tested.

*** 15 fM for nasopharyngeal swab samples

Supplementary Figure 3.17: The evolution of type III-based diagnostics. The first iterations of type III-based diagnostics required RNA extraction and pre-amplification to reach clinically relevant sensitivity. More recent improvements of type-III based diagnostics have been used to detect spiked SARS-CoV-2 RNA with ~2-4 fM sensitivity without pre-amplification. This work demonstrates how type-III based diagnostic detects ~ 8 fM of pure SARS-CoV-2 RNA and ~ 15 fM of SARS-CoV-2 RNA directly in unprocessed patient samples (i.e., nasopharyngeal swabs) in 20 min.



Supplementary Figure 3.18: Alignment of N gene sequences targeted by TtCsm-complex in SARS-CoV-2 and other coronaviruses in the cross-reactivity assay. Matching nucleotides are shown with dots, mismatches with nucleotide symbols. Boxes show segments (S1-6) of the crRNA:target RNA duplex bound by backbone subunits of the Csm-complex. Mismatches in S1 and S2 (red) permit RNA binding but significantly reduce ATP polymerization by Cas10 subunits.

Supplementary table 3.1. Fluorescent RNA and DNA reporters.

N	**ssRNA Reporter**
A01	5'-FAM-AUAAAAAAAU-IABkFQ
A02	5'-FAM-AUAACAACAU-IABkFQ
A03	5'-FAM-AUAAGAAGAU-IABkFQ
A04	5'-FAM-AUAAUAAUAU-IABkFQ
A05	5'-FAM-AUACCACCAU-IABkFQ
A06	5'-FAM-AUACGACGAU-IABkFQ
A07	5'-FAM-AUACUACUAU-IABkFQ
A08	5'-FAM-AUAGCAGCAU-IABkFQ
A09	5'-FAM-AUAGGAGGAU-IABkFQ
A10	5'-FAM-AUAGUAGUAU-IABkFQ
A11	5'-FAM-AUAUCAUCAU-IABkFQ
A12	5'-FAM-AUAUGAUGAU-IABkFQ
B01	5'-FAM-AUAUUAUUAU-IABkFQ
B02	5'-FAM-AUCCCCCCAU-IABkFQ
B03	5'-FAM-AUCCGCCGAU-IABkFQ
B04	5'-FAM-AUCCUCCUAU-IABkFQ
B05	5'-FAM-AUCGGCGGAU-IABkFQ
B06	5'-FAM-AUCGUCGUAU-IABkFQ
B07	5'-FAM-AUCUGCUGAU-IABkFQ
B08	5'-FAM-AUCUUCUUAU-IABkFQ
B10	5'-FAM-AUGGUGGUAU-IABkFQ
B11	5'-FAM-AUGUUGUUAU-IABkFQ
B12	5'-FAM-AUUUUUUUAU-IABkFQ
C01	5'-FAM-CGCGAAAAAACGCG-IABkFQ-3'

C02	5'-FAM-CGCGAACAACCGCG-IABkFQ-3'
C03	5'-FAM-CGCGAAGAAGCGCG-IABkFQ-3'
C04	5'-FAM-CGCGAAUAAUCGCG-IABkFQ-3'
C05	5'-FAM-AUAUACCACCAUAU-IABkFQ-3'
C06	5'-FAM-AUAUACGACGAUAU-IABkFQ-3'
C07	5'-FAM-CGCGACUACUCGCG-IABkFQ-3'
C08	5'-FAM-AUAUAGCAGCAUAU-IABkFQ-3'
C09	5'-FAM-AUAUAGGAGGAUAU-IABkFQ-3'
C10	5'-FAM-CGCGAGUAGUCGCG-IABkFQ-3'
C11	5'-FAM-CGCGAUCAUCCGCG-IABkFQ-3'
C12	5'-FAM-CGCGAUGAUGCGCG-IABkFQ-3'
D01	5'-FAM-CGCGAUUAUUCGCG-IABkFQ-3'
D02	5'-FAM-AUAUCCCCCCAUAU-IABkFQ-3'
D03	5'-FAM-AUAUCCGCCGAUAU-IABkFQ-3'
D04	5'-FAM-AUAUCCUCCUAUAU-IABkFQ-3'
D05	5'-FAM-AUAUCGGCGGAUAU-IABkFQ-3'
D06	5'-FAM-AUAUCGUCGUAUAU-IABkFQ-3'
D07	5'-FAM-AUAUCUGCUGAUAU-IABkFQ-3'
D08	5'-FAM-CGCGCUUCUUCGCG-IABkFQ-3'
D10	5'-FAM-AUAUGGUGGUAUAU-IABkFQ-3'
D11	5'-FAM-CGCGGUUGUUCGCG-IABkFQ-3'
D12	5'-FAM-CGCGUUUUUUCGCG-IABkFQ-3'
N	**ssDNA Reporter**
A01	/56-FAM/ATA AAA AAA T/3IABkFQ/
A02	/56-FAM/ATA ACA ACA T/3IABkFQ/
A03	/56-FAM/ATA AGA AGA T/3IABkFQ/

A04	/56-FAM/ATA ATA ATA T/3IABkFQ/
A05	/56-FAM/ATA CCA CCA T/3IABkFQ/
A06	/56-FAM/ATA CGA CGA T/3IABkFQ/
A07	/56-FAM/ATA CTA CTA T/3IABkFQ/
A08	/56-FAM/ATA GCA GCA T/3IABkFQ/
A09	/56-FAM/ATA GGA GGA T/3IABkFQ/
A10	/56-FAM/ATA GTA GTA T/3IABkFQ/
A11	/56-FAM/ATA TCA TCA T/3IABkFQ/
A12	/56-FAM/ATA TGA TGA T/3IABkFQ/
B01	/56-FAM/ATA TTA TTA T/3IABkFQ/
B02	/56-FAM/ATC CCC CCA T/3IABkFQ/
B03	/56-FAM/ATC CGC CGA T/3IABkFQ/
B04	/56-FAM/ATC CTC CTA T/3IABkFQ/
B05	/56-FAM/ATC GGC GGA T/3IABkFQ/
B06	/56-FAM/ATC GTC GTA T/3IABkFQ/
B07	/56-FAM/ATC TGC TGA T/3IABkFQ/
B08	/56-FAM/ATC TTC TTA T/3IABkFQ/
B09	/56-FAM/ATG GGG GGA T/3IABkFQ/
B10	/56-FAM/ATG GTG GTA T/3IABkFQ/
B11	/56-FAM/ATG TTG TTA T/3IABkFQ/
B12	/56-FAM/ATT TTT TTA T/3IABkFQ/
C01	/56-FAM/CGC GAA AAA ACG CG/3IABkFQ/
C02	/56-FAM/CGC GAA CAA CCG CG/3IABkFQ/
C03	/56-FAM/CGC GAA GAA GCG CG/3IABkFQ/
C04	/56-FAM/CGC GAA TAA TCG CG/3IABkFQ/
C05	/56-FAM/ATA TAC CAC CAT AT/3IABkFQ/

C06	/56-FAM/ATA TAC GAC GAT AT/3IABkFQ/
C07	/56-FAM/CGC GAC TAC TCG CG/3IABkFQ/
C08	/56-FAM/ATA TAG CAG CAT AT/3IABkFQ/
C09	/56-FAM/ATA TAG GAG GAT AT/3IABkFQ/
C10	/56-FAM/CGC GAG TAG TCG CG/3IABkFQ/
C11	/56-FAM/CGC GAT CAT CCG CG/3IABkFQ/
C12	/56-FAM/CGC GAT GAT GCG CG/3IABkFQ/
D01	/56-FAM/CGC GAT TAT TCG CG/3IABkFQ/
D02	/56-FAM/ATA TCC CCC CAT AT/3IABkFQ/
D03	/56-FAM/ATA TCC GCC GAT AT/3IABkFQ/
D04	/56-FAM/ATA TCC TCC TAT AT/3IABkFQ/
D05	/56-FAM/ATA TCG GCG GAT AT/3IABkFQ/
D06	/56-FAM/ATA TCG TCG TAT AT/3IABkFQ/
D07	/56-FAM/ATA TCT GCT GAT AT/3IABkFQ/
D08	/56-FAM/CGC GCT TCT TCG CG/3IABkFQ/
D09	/56-FAM/ATA TGG GGG GAT AT/3IABkFQ/
D10	/56-FAM/ATA TGG TGG TAT AT/3IABkFQ/
D11	/56-FAM/CGC GGT TGT TCG CG/3IABkFQ/
D12	/56-FAM/CGC GTT TTT TCG CG/3IABkFQ/

dsDNA Reporter
5`-FAM-TGAACTGAACTGAACTGAACTGAACTGAACT
AGTTCAGTTCAGTTCAGTTCAGTTCAGTTCA-BHQ-3`

Supplementary Table 3.2. RNase P RNA levels in nasopharyngeal swab samples and diluted total RNA from HEK 293T cells used as negative control in the manuscript.

Sample	Ct value (RP primers, 2019-nCoV CDC EUA Kit, IDT#10006606)
RNA patient sample 1	26.85
RNA patient sample 2	24.73
RNA patient sample 3	35.41
RNA patient sample 4	28.29
RNA patient sample 5	28.58
RNA patient sample 6	20.36
RNA patient sample 7	25.24
RNA patient sample 8	25.14
RNA patient sample 9	28.15
RNA patient sample 10	30.37
Total RNA from HEK 293T (diluted)	23.35

Supplementary Table 3.3. Oligonucleotide primers used to generate DNA templates for in vitro transcription with T7 polymerase

Primer name	Sequence (5'-3')
SARS-CoV-2 N1 T7 template Forward	GATAATACGACTCACTATAGGGAACTGATTACAAACATTGGCCGCAAATTGCA CAATT
SARS-CoV-2 N1 T7 template Reverse	GCGCGACATTCCGAAGAACGCTGAAGCGCTGGGGGCAAATTGTGCAATTTGCG GCC
SARS-CoV-1 N1 T7 template Forward	GATAATACGACTCACTATAGGGAACTGATTACAAACATTGGCCGCAAATTGCA CAATT
SARS-CoV-1 N1 T7 template Reverse	GCGTGACATTCCAAAGAATGCAGAGGCACTTGGAGCAAATTGTGCAATTTGCG GCC

Methods

Ethical Statement

The study was reviewed by the Montana State University Institutional Review Board (IRB) For the Protection of Human Subjects (FWA 00000165). De-identified clinical samples were obtained with IRB approval (protocol #DB033020) and informed consent from patients undergoing testing for SARS-CoV-2 at Bozeman Health Deaconess Hospital.

Human Clinical Sample Collection and)reparation

Nasopharyngeal swabs from patients that either tested negative or positive for SARS-CoV-2 were collected in viral transport media. RNA was extracted from all patient samples using the QIAamp Viral RNA Mini Kit (QIAGEN).

Nucleic Acids

Sodium salts of cyclic di-, tri-, tetra-, penta- and hexa-adenosine monophosphates (cA_2-$_6$), sodium salts of cyclic tri- and tetra- guanosine monophosphates (cG_2, cG_3), sodium salts of cyclic (guanosine- (2′ −> 5′)- monophosphate- adenosine- (3′ −> 5′)- monophosphate) (cGAMP), cyclic adenosine monophosphate-guanosine monophosphate (cAG), and cyclic diguanosine-5′-monophosphate (cG_2) were purchased from Biolog Life Science Institute. Fluorescent reporters (RNA and DNA) were purchased from IDT (Supplementary Data 1). The dsDNA reporter was ordered as a duplex from IDT. Target and non-target RNAs of SARS-CoV-2 N-gene were in vitro transcribed with MEGAscript T7 (Thermo Fisher Scientific) from PCR products generated from pairs of synthesized overlapping DNA oligos (Supplementary Data 1) (Eurofins). Transcribed RNAs were purified by denaturing PAGE. Total RNA from HEK 293T cells was extracted using

TRIzol reagent. The genomic RNAs of SARS-CoV-1, Middle East respiratory syndrome coronavirus (MERS-CoV), seasonal coronaviruses 229E (HCoV-229E), NL63 (HCoV-NL63), and HKU1 (HCoV-HKU1) were purchased from ATCC.

Non-Targeting Control (NTC)

Total RNA extracted from SARS-CoV-2 negative nasopharyngeal swabs or total RNA extracted from HEK 293T cells (ATCC, CRL-3216) were used as negative controls. RNA extracted from HEK 293T cells was diluted to match the average Ct level (~27) obtained for RNAseP mRNA in RNA samples extracted from nasopharyngeal swabs (Supplementary Table 1). The RT-qPCR for RNase P mRNA was performed using CDC RP primers and probe (2019-nCoV CDC EUA Kit, IDT#10006606).

Plasmids

Plasmids encoding the type III-A Csm complex from *Thermus thermophilus* (pCDF-5xT7-TtCsm; Addgene #128572 and pACYC-TtCas6-4xcrRNA4.5; Addgene #127764), were a gift from Jennifer Doudna. Vector pCDF-5xT7-TtCsm was used as a template for site-directed mutagenesis to mutate the D34 residue in Csm3 to alanine (D34A) and inactivate Csm3-mediated cleavage of target RNA (pCDF-5xT7-TtCsmCsm3-D34A)42. The CRISPR array in pACYC-TtCas6-4xcrRNA4.5 was replaced with a synthetic CRISPR array (GeneArt) containing five repeats and four identical spacers, designed to target the N-gene of SARS-CoV-2 (i.e., pACYC-TtCas6-4xgCoV2N1)18. TtCas6 was PCR was PCR-amplified from the pACYC-TtCas6-4xcrRNA4.5 plasmid and cloned between the NcoI and XhoI sites in the pRSF-1b backbone (Millipore Sigma) (pRSF-TtCas6). Expression vector encoding TtCsm6 nuclease,

pC0075 TtCsm6 His6-TwinStrep-SUMO-BsaI, was a gift from Feng Zhang (Addgene plasmid #115270)43.

Gene fragments encoding for Can1 from *Thermus thermophilus* (TtCan1; "WP_011229147.1"), Can2 from *Archaeoglobi archaeon* JdFR-42 (AaCan2; "2730024700"), *Clostridium thermobutyricum* (CtCan2; "WP_195972101.1"), and *Thermus thermophilus* (TtCan2; "WP_143585921.1"), were codon optimized for expression in *E. coli*, synthesized by GenScript, and cloned into pC0075 vector (Addgene #115270) in frame with the N-terminal His6-TwinStrep-SUMO tag using NcoI and XhoI restriction sites to replace the TtCsm6 gene. NucC from *Clostridium tepidum* BSD2780120874b_170522_A10 (CtNucC; "WP_195923598.1"), *Elioraea* sp. Yellowstone (EsNucC; "WP_141855040.1") and *Acidimicrobiales bacterium* mtb01 (Amtb01NucC; "TEX45487.1"), were cloned into pC0075 backbone using the same restriction sites as for Can1 and Can2 genes.

Protein Expression and Purification

Expression and purification of the TtCsm$^{Csm3\text{-}D34A}$ complex and TtCsm6 were performed as previously described18. TtCan1, AaCan2, CtCan2, TtCan2, CtNucC, EsNucC, and Amtb01NucC) were purified according to the following protocol. Each expression vector was transformed into *Escherichia coli* BL21(DE3) cells and grown in LB Broth (Lennox) (Thermo Fisher Scientific) at 37 °C to an OD600 of 0.5. Cultures were then incubated on ice for 1 h, and then induced with 0.5 mM IPTG for overnight expression at 16 °C. Cells were lysed with sonication in Lysis buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM TCEP) and lysate was clarified by centrifugation at 10,000 × *g* for 25 min, 4 °C. The lysate was heat-treated at 55 °C for 45 min and clarified by centrifugation at 10,000 × *g* for 25 mins at 4 °C. His_6-TwinStrep-tagged

protein was bound to a StrepTrap HP column (Cytiva) and washed with Lysis buffer. The protein was eluted with Lysis buffer supplemented with 2.5 mM desthiobiotin and concentrated (10k MWCO Corning Spin-X concentrators) at 4 °C. Affinity tags were removed from the protein using His-tagged SUMO protease (100 µL of 2.5 mg/mL protease per 20 mg of protein) during dialysis against SUMO digest buffer (30 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM dithiothreitol (DTT), 0.15% Igepal) at 4 °C overnight. The tag and the protease were applied to HisTrap HP column (Cytiva), and the flow-through was concentrated using Corning Spin-X concentrators at 4 °C. Finally, the protein was purified using a HiLoad Superdex 200 26/600 size-exclusion column (Cytiva) in storage buffer (20 mM Tris-HCl pH 7.5, 1 mM DTT,400 mM monopotassium glutamate, 5% glycerol). Fractions containing the target protein were pooled, concentrated, aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C.

^{32}P-Labeling of RNA Oligos

Target (SARS-CoV-2 N1) and non-target RNAs were transcribed from PCR extended duplex oligos using home-made T7 RNA polymerase (Supplementary Table 2) (Eurofins). The IVT RNAs were gel purified and dephosphorylated with Quick CIP (NEB) for 20 min at 37 °C in 1× CutSmart Buffer (NEB). The phosphatase was inactivated by heating at 80 °C for 5 min before 5′ end-labeling the RNAs with T4 polynucleotide kinase (NEB) and [γ-^{32}P]-ATP (PerkinElmer) for 30 min at 37 °C. The kinase was heat inactivated by heating at 65 °C for 20 min.

Binding and Pull-Down of RNA Oligos with TtCsm

For the experiments shown in Fig. 1b and Supplementary Fig. 1b, c, ^{32}P-labeled RNA (25 nM) was incubated with TtCsm$^{Csm3\text{-}D34A}$ (160 nM) targeting SARS-CoV-2 N-gene in 1×

Binding Buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM TCEP) for 20 min at 65 °C. The reaction mixtures were added to 10 µL of HisPur Ni-NTA Magnetic beads (ThermoFisher) equilibrated in Binding Buffer and incubated on ice 30 min with vortexing every 10 min. The beads were separated from the supernatant using a magnet and washed with 50 µL 1× binding buffer. The RNA was extracted from supernatant (unbound fraction) and beads (bound fraction) using Acid Phenol: chloroform (Ambion). Extracted RNA was resolved using UREA-PAGE, exposed to a phosphor screen, and imaged on a Typhoon 5 imager (Amersham). Bands corresponding to the IVT RNAs were quantified using ImageJ v1.52t and the percent bound calculated [bound/(bound + free)*100%].

Complexing of TtCsm with Magnetic Beads

The HisPur Ni-NTA Magnetic beads (ThermoFisher) were washed two times with a 1× Binding Buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM TCEP). For one reaction, 5 µL of equilibrated beads were mixed with TtCsm^{D34A} complex (25 nM) in 1× Binding Buffer (V = 50 µL) and incubated for 30 min on ice. The beads with the complex (Csm-beads) were concentrated with a magnet and resuspended in 5 µL of 1× Binding Buffer.

Thin-Layer Chromatography (TLC)

For the experiments shown in Fig. 1c, 3 µL of positive sample (target RNA diluted in NTC, 10^{10} copies/µL) or 3 µL of NTC were mixed with TtCsm$^{Csm3\text{-}D34A}$ complex (25 nM) and 250 µM ATP supplemented with [α-^{32}P]-ATP (PerkinElmer) in the reaction buffer (20 mM Tris-HCl pH 7.8, 250 mM monopotassium glutamate, 10 mM ammonium sulfate, 1 mM TCEP (tris(2-carboxyethyl)phosphine)), 5 mM magnesium sulfate). The reaction was incubated at 60 °C for 1 h. For the pull-down reactions, 120 µL of positive or negative samples were mixed with 5 µL of

Csm-beads in Binding Buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM TCEP) for 10 min at 60 °C. The Csm-beads were concentrated with a magnet and the supernatant was discarded. The Csm pellets were resuspended in 30 µL of the reaction buffer and 250 µM ATP supplemented with [α-32P]-ATP (PerkinElmer). Reaction products were phenol-chloroform extracted and resolved on silica TLC plates (Millipore).

Samples (1 µL) were mixed with 100 mM sodium acetate, pH 5.2 (2 µL), and spotted 1.5 cm above the bottom of the TLC plate. The plate was placed inside a 2 L beaker filled to ~0.5 cm with developing solvent (0.2 M ammonium bicarbonate pH 9.3, 70% ethanol, and 30% water) and capped with aluminum foil. The plate was run for 2 h at room temperature and dried. TLC plate was exposed to a phosphor screen and imaged with Typhoon phosphor imager. Chemically synthesized standards (2 µM) were resolved on the same TLC plate and visualized using UV shadowing.

To test cA_3 and cA_4 hydrolysis in the presence of ancillary nuclease, radiolabeled cA_3 and cA_4 produced above were mixed with nuclease (500 nM) in the reaction buffer and incubated for 1 hour at 55 °C. Reaction products were phenol-chloroform extracted and resolved using TLC for 45 min as described above.

Type III-Based RNA Detection

3 µL of RNA sample was mixed with 250 µM ATP, 0.5 nM $TtCsm^{D34A}$ complex, 300 nM of nuclease (TtCsm6, AaCan2, or CtNucC) with corresponding reporter in a reaction buffer (20 mM Tris-HCl pH 7.8, 150 mM or 250 mM monopotassium glutamate, 10 mM ammonium sulfate, 1 mM TCEP (tris(2-carboxyethyl)phosphine)), 5 mM magnesium sulfate (for TtCsm6 and CtNucC) or 5 mM manganese(II) chloride (for AaCan2) in a 30 µL reaction. The reporter B8

(300 nM) was used for the reaction with TtCsm6, D7 (300 nM) – with AaCan2, and dsDNA probe (300 nM) – with CtNucC. Reactions were incubated at 55 °C or 60 °C. Cleavage of fluorescent reporters was detected by measuring fluorescence every 10 sec in a real-time PCR instrument QuantStudio 3 (Applied Biosystems).

Type III-Based RNA Pull-Down and Detection

To bind TtCsm^{D34A} complex with the magnetic beads, the HisPur Ni-NTA Magnetic beads (ThermoFisher) were washed two times with a 1× Binding Buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM TCEP). 5 µL of equilibrated beads were mixed with TtCsm^{D34A} complex (30 nM) in 1× Binding Buffer (V = 50 µL) and incubated for 30 min on ice. The beads with the complex (Csm-beads) were concentrated with a magnet and resuspended in 5–500 µL of 1× Binding Buffer. For one reaction 5 µL of resuspended Csm-beads was used.

Pull-Down and Detection from RNA Samples and Nasopharyngeal Swabs

120 µL of sample was mixed with 5 µL of Csm-beads in 1× Binding Buffer for 10 min at 60 °C (in experiments shown in Fig. 5 of 1× Binding Buffer was supplemented with 0.01% Triton X-1001 mM EDTA). For pull-downs from nasopharyngeal swabs RNase Inhibitor (N2615, Promega) was added to the final concentration 40 U/µl. Csm-beads were concentrated with a magnet and the supernatant was discarded. For three-step protocol the Csm-beads pellet was resuspended in 20 µL of the 1× reaction buffer (20 mM Tris-HCl pH 7.8, 250 mM monopotassium glutamate, 10 mM ammonium sulfate, 1 mM TCEP (tris(2-carboxyethyl)phosphine)), 5 mM magnesium sulfate / manganese(II) chloride) containing ATP (250 µM). The reaction was incubated at least 10 min at 60 °C, the Csm-beads were pelleted, and

the supernatant (10 µL) was transferred to a new reaction with TtCsm6 (300 nM) and B8 RNA Reporter (300 nM) or AaCan2 (300 nM) and D7 RNA Reporter (300 nM) in 1× reaction buffer (V = 30 µL) (Supplementary Data 1). Reactions were incubated at 55 °C. For two-step protocol the Csm-beads pellet was resuspended in 30 µL of the 1× reaction buffer (20 mM Tris-HCl pH 7.8, 250 mM monopotassium glutamate, 10 mM ammonium sulfate, 1 mM TCEP (tris(2-carboxyethyl)phosphine)), 5 mM magnesium sulfate/manganese(II) chloride) containing ATP (250 µM), RNA Reporter (300 nM), and nuclease AaCan2 (300 nM). The reaction was incubated at least 10 min at 60 °C. Cleavage of the fluorescent RNA reporter was detected by measuring fluorescence every 10 s in a real-time PCR instrument QuantStudio 3.

RT-qPCR

RT-qPCR was performed using N1 and RP CDC primers (2019-nCoV CDC EUA Kit, IDT#10006606). RNA was extracted from patient samples with QIAamp Viral RNA Mini Kit (QIAGEN, # 52906) and used for one-step RT-qPCR in ABI 7500 Fast Real-Time PCR System according to CDC protocols (https://www.fda.gov/media/134922/download). In brief, 20 µL reaction included 8.5 µL of Nuclease-free Water, 1.5 µL of Primer and Probe mix (IDT, 10006713), 5 µL of TaqPath 1-Step RT-qPCR Master Mix (ThermoFisher, A15299) and 5 µL of the RNA. Nuclease-free water was used as negative template control (NTC). Amplification was performed as follows: 25 °C for 2 min, 50 °C for 15 min, 95 °C for 2 min followed by 45 cycles of 95 °C for 3 s and 55 °C for 30 s. To quantify viral RNA in the samples, standard curve for N1 primers was generated using a dilution series of a SARS-CoV-2 synthetic RNA fragment (RTGM 10169, NIST) spanning N gene with concentrations ranging from 10 to 10^6 copies per µL. Three

technical replicates were performed at each dilution. The NTC showed no amplification throughout the 45 cycles of qPCR.

Nanopore Sequencing of DNA Cleavage Fragments

DNA cleavage fragments were sequenced using Oxford Nanopore with Ligation Sequencing Kit (SQK-LSK109). After incubation with nucleases, cleavage fragments were column-purified using DNA Clean & Concentrator-5 kit (Zymo Research, D4004) as instructed. Next, for each sample 200 ng of purified DNA was used to prepare sequencing libraries with NEBNext® Ultra™ II DNA Library Prep Kit (NEB, E7645S). Briefly, DNA was end-repaired with NEBNext Ultra II End Prep Enzyme Mix, which fills 5′- and removes 3′- overhangs. Next, end-repaired fragments were barcoded with Native Barcoding Expansion kit (ONT, EXP-NBD104) using Ultra II Ligation Master Mix (NEB). Barcoded DNA fragments were pooled together and purified with magnetic beads (Omega Bio-tek, M1378-01). Freshly mixed 80% ethanol was used to wash magnetic bead pellet. Sequencing adapters (AMII) were ligated to barcoded DNA using NEBNext® Quick Ligation Module (NEB, E6056S). Ligation reactions were purified with magnetic beads. SFB buffer (ONT, EXP-SFB001) was used for washes. Resulting DNA library was eluted from the beads in 20 µL of EB buffer (QIAGEN, #19086). DNA concentration was measured with Qubit dsDNA HS Assay (ThermoFisher, Q32851), and 20 ng was loaded on the Nanopore MinION (MIN-101B, R9.4.1 flow cell). The flow cell was primed, and library was loaded according to Oxford Nanopore protocol (SQK-LSK109 kit). The sequencing run was performed in the fast base calling mode in the MinKNOW software v5.2.4.

Sequenced reads were demultiplexed using guppy-barcoder v6.2.11+e17754edc (ONT) and aligned with minimap2 v2.17-r954-dirty (-ax map-ont mode) to the reference plasmid

sequence that was modified by adding 1000 bp overlaps at the 5′- and 3′- ends. Overlapping regions were introduced to account for circular nature of the plasmid. Resulting alignments (BAM files) were sorted and indexed using samtools v1.13. Next, *bamtobed* function in bedtools package v2.30.0 was used to generate BED files and read coordinates were extracted. Read end coordinates were used to calculate cleavage fragment length distributions and map frequencies of cuts at specific locations. To analyze the sequence preferences of each nuclease, 14 bp windows surrounding read ends were extracted with *getfasta* function from bedtools package v2.30.0. Resulting fasta files were used to calculate position weigh matrices (PWMs) with *getPwmFromFastaFile()* function in DiffLogo v2.16.0R package. Finally, PWMs were plotted as sequence logos using ggseqlogo v0.1R package. Sequencing depth around the most frequent cut site for each nuclease was calculated with samtools v1.13 *depth* function and plotted with ggplot2 package v3.3.5 in RStudio v2022.07.1+554.

RNA and DNA Reporter's Libraries

To determine the optimal RNA or DNA reporter for each cOA-activated nuclease, we constructed a library of variable single-stranded RNA or DNA molecules tethering a FAM fluorophore to an Iowa Black quencher. The Biostrings package in R was used to construct a library of reporter sequences containing each of the 64 unique trinucleotide combinations possible. Since multiple unique trinucleotides could be included in a single reporter (e.g., 5′-FAM-AUAGAAGAAU-IABkFQ-3′ contains AGA, GAA, and AAG), we narrowed our initial library of 64 reporters to remove redundant sequences. This resulted in a library of 24 unique reporter sequences, each of which were integrated into reporters of different length

(Supplementary Data 1). The R-script used to design these reporters is accessible on GitHub (WiedenheftLab/RNA_reporter_design, https://doi.org/10.5281/zenodo.7368892).

In vitro DNA and RNA Cleavage Assays

All reactions were performed in a buffer containing 20 mM Tris-HCl pH 7.8, 50–250 mM monopotassium glutamate, 10 mM ammonium sulfate, 1 mM TCEP, 5 mM magnesium sulfate or 5 mM manganese chloride. Plasmid DNA cleavage assays were performed by incubating 15 nM of Lenti-luciferase-P2A-Neo (Addgene #105621) plasmid with TtCan1, AaCan2 or CtNucC (15–200 nM) in the presence of cOAx (20–50 nM) in 10 µL reaction. After 5–15 min incubation, Gel Loading Dye, Purple (6X) (NEB) was added and 4 µL was loaded on 1% agarose gel. For ssDNA and ssRNA cleavage assays, 0.425 µM of 71 nt DNA oligo (CGTCGTACCGGTTAGAGGATGGTGCAAGCGTAATCTGGAACATCGTATGGGTATGCC CACGGTGTCCACGGCG, Eurofins), 0.425 µM of 74 nt IVT RNA SARS-CoV-2 N-gene (Supplementary Table 2) or 15 nM of ssDNA ΦX174 bacteriophage (NEB, N3023) were incubated with TtCan1 (200 nM) or AaCan2 (200 nM) in the presence of cOAx (20–50 nM) in 10 µL. After 5–15 min incubation, 10 µL was loaded on 12% UREA PAGE or 1% alkaline agarose gel.

Phylogenetic Analysis of Can1 and Can2 Proteins

A DELTA-BLAST was initiated, using previously described Can1 and Can2 proteins as queries[31,32,33] to generate individual lists of closely related proteins with an e-value cutoff of 10^{-4} and 50% query coverage. The resulting sequences were then used as queries to initiate a PSI-BLAST search with an E-value cutoff of 10^{-4} and 50% query coverage. This step was repeated until convergence and redundant sequences were removed with CD-HIT v4.7[44]. In case

of Can1, sequences from a previously published dataset16 that contain two CARF domains and a nuclease domain were used to generate multiple sequence alignment of Can1-related proteins. In total, 29 sequences of Can1-related proteins and 2,531 sequences of Can2-related proteins were used separately to generate multiple sequence alignment with a local version of MAFFT v7.42945 (–localpair —maxiterate 1000). The generated alignments for Can1 and Can2 were curated with MaxAlign v1.146 to remove misaligned or non-homologous sequences. The resulting dataset—comprised of 29 Can1-like and 1,283 Can2-like proteins, respectively—were then individually realigned with MAFFT and HMMbuild47 (HMMER v3.2.1) was used to generate HMM profiles from each alignment. The resulting profiles were used to search a local database of prokaryotic genomes from NCBI (downloaded on June 11, 2021) and list of sequences identified in BLAST search from previous steps. An initial search performed with these HMM profiles identified 1442 Can1 and 5,431 Can2 homologs, which were manually filtered according to the presence of domains that define each protein, as well as the presence of conserved residues found in CARF and nuclease domains. The resulting set of 204 Can1 and 3,121 Can2 proteins were merged into a single file and aligned in MAFFT (LINSI option) for downstream phylogenetic analyses. Next, Trimal v1.448 was used to remove columns in the alignment comprised of ≥70% gaps. Thermostable homologs of Can1 and Can2 were annotated according to organisms that they are originated. ProtTest v3.4.249 was used to select an evolutionary model, and a phylogenetic tree was constructed in IQ-TREE v1.6.150 using the recommended model (i.e., LG+G+F). The phylogenetic tree was plotted using the ggTree package in R51

Phylogenetic Analysis of NucC

A phylogenetic tree of NucC proteins was generated using the same methods as described above for Can1/Can2 proteins. Briefly, DELTA-BLAST and PSI-BLAST searches with previously identified NucC proteins35 generated a list of closely related proteins (e-value cutoff of 10^{-4} and minimum 50% query coverage). The resulting dataset was filtered with CD-HIT v4.7 to remove redundant sequences. The resulting 1230 NucC sequences were aligned with MAFFT (–localpair —maxiterate 1000), and poorly aligned and highly gapped sequences were removed with MaxAlign. The resulting set of 896 NucC sequences were re-aligned with MAFFT as previously described, and the resulting alignment was used to generate a NucC HMM profile which we used to search within prokaryotic genomes from NCBI. This search identified 1774 hits, which were filtered according to the presence of restriction endonuclease-like domain (i.e., ID_{x30}EAK-motif containing), gate-loop and cA_3 binding domains and were aligned with MAFFT. The remaining NucC homologs were curated according to organisms they are originated from to identify thermostable NucC homologs. The resulting alignment of 1510 NucC proteins with 21 thermostable homologs was used to generate a phylogenetic tree with FastTree v2.1.1052 and was plotted using the ggTree package in R.

Statistics & Reproducibility

All statistical analyses were performed in RStudio. Analysis of Variance Models (ANOVA) were calculated with *aov()* function in the stats R package. Multiple comparisons between positive samples and negative controls were performed using Dunnett's test with multcomp R package. Reaction slopes were determined by extracting coefficients from linear models fitted to fluorescence data with *lm()* function in R. The linear regions of the fluorescence

curves were identified using rolling regression with *auto_rate()* function in respR package. Statistical threshold for detecting SARS-CoV-2 in patient samples with Csm-based assay was set as mean of negative control ± 2.33S.D., which captures 98% of variation in negative samples (2% false positive). Statistical significance levels used in the figures are ***$p < 0.001$, **$p < 0.01$, and *$p < 0.05$. No statistical method was used to predetermine the sample size. The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. No data were excluded from the analyses.

Manuscript References

1. Drain, P. K. Rapid diagnostic testing for SARS-CoV-2. N. Engl. J.Med. https://doi.org/10.1056/NEJMcp2117115 (2022).

2. Allan-Blitz, L. T. & Klausner, J. D. A real-world comparison of SARSCoV- 2 rapid antigen testing versus PCR testing in Florida. J. Clin.Microbiol. 59, e0110721 (2021).

3. Jessica, L. et al. Evaluation of Abbott BinaxNOWRapid Antigen Test for SARS-CoV-2 Infection at Two Community-Based Testing Sites—Pima County, Arizona, November 3–17. https://www.cdc.gov/mmwr/volumes/70/wr/pdfs/mm7003e3-H.pdf (2020).

4. Kaminski, M. M., Abudayyeh, O. O., Gootenberg, J. S., Zhang, F. & Collins, J. J. CRISPR-based diagnostics. Nat. Biomed. Eng. 5,643–656 (2021).

5. Abudayyeh, O. O. & Gootenberg, J. S. CRISPR diagnostics. Science 372, 914–915 (2021).

6. Pardee, K. et al. Rapid, low-cost detection of Zika virus using programmable biomolecular components. Cell 165, 1255–1266 (2016).

7. Jiao, C. et al. Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. Science 372, 941–948 (2021).

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

9. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 360, 436 (2018).

10. Santiago, J. G. Inconsistent treatments of the kinetics of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) impair assessment of its diagnostic potential. QRB Discov. 3, e9 (2022).

11. Huyke, D. A. et al. Enzyme kinetics and detector sensitivity determine limits of setection of amplification-free CRISPR-Cas12 and CRISPR-Cas13 diagnostics. Anal. Chem. 94, 9826–9834 (2022).

12. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science 357, 605–609 (2017).

13. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č. & Siksnys, V. Spatiotemporal control of type III-A CRISPR-Cas immunity: coupling DNA degradation with the target RNA recognition. Mol. Cell 62, 295–306 (2016).

14. Niewoehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature 548, 543–548 (2017).

15. Athukoralage, J. S. et al. The dynamic interplay of host and viral enzymes in type iii crispr-mediated cyclic nucleotide signalling. Elife 9, e55852 (2020).

16. Makarova, K. S. et al. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. Nucleic Acids Res. 48, 8828–8847 (2020).

17. Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-ACRISPR–Cas immunity. Nat. Microbiol. 4, 656–662 (2019).

18. Santiago-Frangos, A. et al. Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. Cell Rep. Med. 2, 100319 (2021).

19. Steens, J. A. et al. SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringentCARF ribonuclease activation. Nat. Commun. 12, 1–12 (2021).

20. Sridhara, S., Goswami, H. N., Whyms, C., Dennis, J. H. & Li, H. Virus detection via programmable Type III-A CRISPR-Cas systems. Nat. Commun. https://doi.org/10.1038/s41467-021-25977-7 (2021).

21. Smalakyte, D. et al. Type III-A CRISPR-associated protein Csm6 degrades cyclic hexa-adenylate activator using both CARF and HEPN domains. Nucleic Acids Res. 48, 9204–9217 (2020).

22. Athukoralage, J. S., Graham, S., Grüschow, S., Rouillon, C. &White, M. F. A type III CRISPR ancillary ribonuclease degrades its cyclic oligoadenylate activator. J. Mol. Biol. 431, 2894–2899 (2019).

23. Athukoralage, J. S., Rouillon, C., Graham, S., Grüschow, S. &White, M. F. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. Nature 562, 277–280 (2018).

24. Garcia-Doval, C. et al. Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. Nat. Commun. 11, 1–9 (2020).

25. Jia, N., Jones,R., Yang,G.,Ouerfelli,O. & Patel,D. J.CRISPR-Cas III-A Csm6 CARF domain is a ring nuclease triggering stepwise cA4 cleavage with ApA>p formation terminating RNase activity. Mol. Cell 75, 944–956.e6 (2019).

26. Grüschow, S., Grüschow, G., Adamson, C. S. & White, M. F. Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector. Nucleic Acids Res. https://doi.org/10.1093/NAR/GKAB1190 (2021).

27. Hale, C. R. et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945–956 (2009).

28. Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. &White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. Elife 7, e36734 (2018).

29. Molina, R. et al. Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. Nat. Commun. 10, 1–14 (2019).

30. Liu, T. Y. et al. Accelerated RNA detection using tandem CRISPR nucleases. Nat. Chem. Biol. 17, 982–988 (2021).

31. Rostøl, J. T. et al. The Card1 nuclease provides defence during type III CRISPR immunity. Nature 590, 624–629 (2021).

32. Zhu,W. et al. TheCRISPR ancillary effector Can2 is a dual-specificity nuclease potentiating type III CRISPR defence. Nucleic Acids Res. 49, 2777–2789 (2021).

33. McMahon, S. A. et al. Structure andmechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. Nat. Commun. 11, 1–11 (2020).

CHAPTER FOUR

CRISPR-BASED ENGINEERING OF RNA VIRUSES

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Co-Author: Artem Nemudryi

Contributions: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing -original draft, Writing – review and editing

Author: Anna Nemudraia

Contributions: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing -original draft, Writing – review and editing

Co-Author: Joseph E. Nichols

Contributions: Conceptualization, Investigation, Methodology, Validation

Co-Author: Andrew M. Scherffius

Contributions: Investigation

Co-Author: Trevor Zahl

Contributions: Investigation

Co-Author: Blake Wiedenheft

Contributions: Conceptualization, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing -original draft, Writing – review and editing

Manuscript Information

†Artem Nemudryi, †Anna Nemudraia, Joseph Nichols, Andrew M. Scherffius, Trevor Zahl, and *Blake Wiedenheft

Science Advances

___ Prepared for submission to a peer-reviewed journal

___ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

_X_ Published in a peer-reviewed journal

Science Advances

Volume 9

Issue 37

8 August 2023

DOI: 10.1126/sciadv.adj8277

Abstract

CRISPR RNA–guided endonucleases have enabled precise editing of DNA. However, options for editing RNA remain limited. Here, we combine sequence-specific RNA cleavage by CRISPR ribonucleases with programmable RNA repair to make precise deletions and insertions in RNA. This work establishes a recombinant RNA technology with immediate applications for the facile engineering of RNA viruses.

Introduction

In 1972, Paul Berg combined restriction endonucleases and DNA ligases to construct the first recombinant DNA molecule (1). This work established a new paradigm for DNA manipulation that has altered the way "questions are formulated and the way solutions are sought" (2). Initially, DNA manipulations were performed in vitro, and the recombinant DNA was delivered to cells for propagation. However, two decades after the first demonstration of recombinant DNA, site-specific nucleases were introduced into cells for in vivo editing (3). This approach used homing endonucleases to make site-specific DNA breaks that are repaired by cellular DNA repair pathways. But the repertoire of sequences recognized by homing endonucleases is limited, and changing the target specificity is difficult (4). Demand for cellular engineering solutions resulted in innovations that made nucleases increasingly more programmable (5, 6), culminating in the repurposing of CRISPR RNA (crRNA)–guided nucleases for DNA targeting (7, 8).

Despite the progress in DNA editing, tools to manipulate RNA without a DNA intermediate remain limited. Although Cas13, a type VI crRNA-guided endoribonuclease, has been successfully programmed to target cellular transcripts, target recognition also activates a collateral nuclease activity that degrades nontarget RNAs (9, 10). In prokaryotes, this off-target nuclease activity triggers cell death, providing community-level protection from invading parasites, but the collateral nuclease activity precludes applications that require precision (11). Nuclease-inactivated Cas13 (dCas13) has been engineered to stimulate alternative splicing in eukaryotes and deliver RNA-modifying enzymes to specific RNA sequences for nucleobase conversions (adenosine to inosine or cytosine to uracil) (12–14).

Type III CRISPR systems, like type VI (Cas13), target RNA. However, unlike Cas13, the type III CRISPR complexes lack collateral ribonuclease activity (15, 16). The absence of collateral cleavage has allowed for precise targeting of cellular transcripts, with minimal off-target effects and cytotoxicity (17–20). Here, we combine type III CRISPR-based RNA cleavage with splinted RNA ligation to make programmed deletions and insertions in target RNA sequences. Manipulating RNA is a prerequisite for studying RNA viruses, and we apply crRNA-guided nucleases to delete or substitute sequences in the genome of the Sindbis virus (SINV). This work establishes a concept for manipulating RNA and demonstrates how this recombinant RNA technology enables rapid and facile engineering of RNA viruses.

Results

Coupling CRISPR-Based RNA Cleavage and Splint Ligation for RNA Editing

Type III-A CRISPR complexes comprise five protein subunits (i.e., Csm2 to Csm5 and Cas10) that self-assemble in an unequal stoichiometry around a crRNA (fig. 4.S1A). The Csm3 protein is an endoribonuclease that forms a helical oligomer along the crRNA and cleaves the bound target RNA in 6-nucleotide (nt) intervals (fig. 4.S1B) (16). We hypothesized that the regular cleavage pattern of Csm3 could be leveraged to delete fragments of RNA between each of the cuts (Fig. 4.1A). To test this hypothesis, we cloned, expressed, and purified the type III-A CRISPR complex of Streptococcus thermophilus (SthCsm complex) with a crRNA targeting the nucleocapsid (N) gene of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (fig. 4.S1, C to E). The purified complex cleaves target RNA with ~98% efficiency, producing fragments of the expected size (fig. 4.S1, F and G).

RNA cleavage by SthCsm generates a 2′,3′-cyclic phosphate and a 5′-hydroxyl at each cut site (15). The RtcB ligase of Escherichia coli can directly join ends produced by SthCsm cleavage, while RNA ligation by bacteriophage T4 ligases requires a 3′-hydroxyl and 5′-phosphate (Fig. 4.1B) (21–23). To facilitate ligation of selected RNA fragments, we used complementary DNA splints that bridge two RNAs. These splints can be designed to imitate different RNA substrates (Fig. 4.1C) (24). To identify conditions that result in efficient ligation, we first cleaved in vitro transcribed (IVT) RNA derived from the N-gene from SARS-CoV-2 using a programmed SthCsm complex. Then, we tested three RNA ligases and two DNA splint designs with or without end repair of the cleaved RNA target. As expected, the RtcB directly ligates RNA fragments produced by SthCsm cleavage (0.69 ± 0.01% with splint A), while T4 RNA ligases (T4 Rnl1 and Rnl2) require PNK (polynucleotide kinase) end repair before ligation (Fig. 4.1D). RtcB ligation involves fewer RNA manipulations (i.e., no end repair), but ligation with T4 RNA ligases was ~25-fold more efficient (up to 17.5 ± 3.0% with T4 Rnl1). Therefore, we chose the latter for further experimentation.

To analyze recombinant RNAs in the ligation mixtures, we reverse-transcribed, amplified, and sequenced the targeted regions. Sequencing reads generated with the unedited RNA uniformly cover the length of the amplicon (Fig. 4.1E, top). However, target cleavage followed by splint ligation results in a sharp drop in the sequencing depth at the target site (Fig. 4.1E). Analysis of the sequencing data at the target site identified programmed deletions with high efficiency and accuracy (Fig. 4.1F). Furthermore, we demonstrate that DNA splints can be designed to simultaneously capture multiple editing outcomes (splint A; Fig. 4.1F) or select a

specific deletion size (splint B; Fig. 4.1F). These data confirm that target cleavage with SthCsm, followed by splint ligation, results in efficient RNA editing at the specified location.

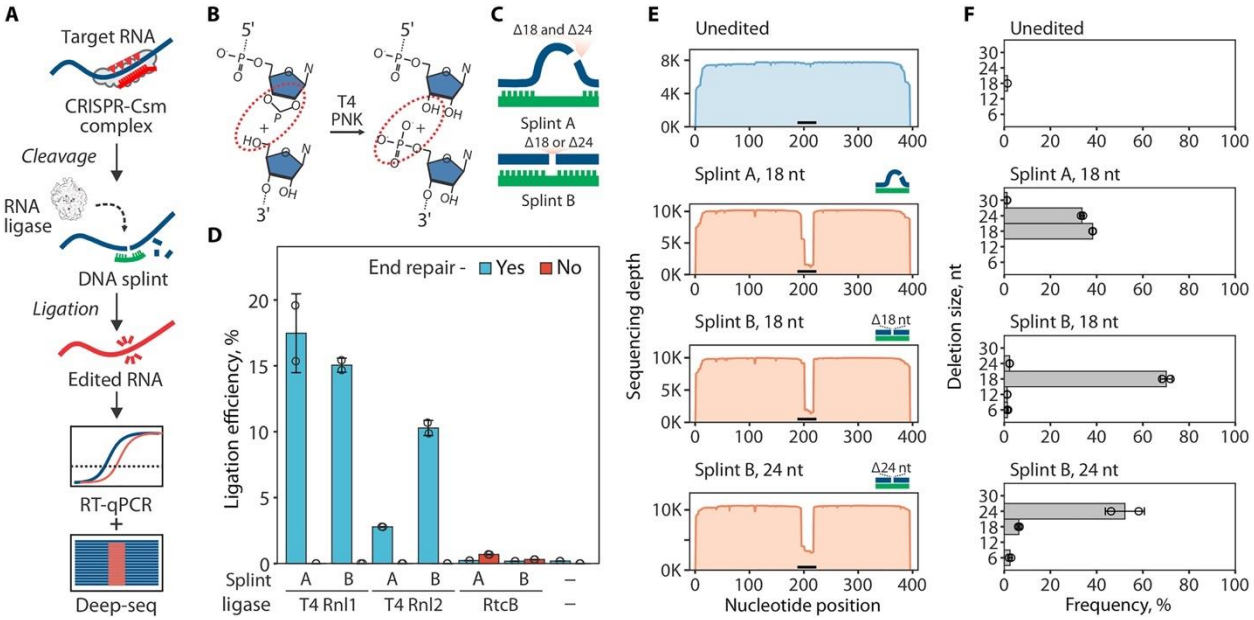


Fig. 4.1. RNA editing with CRISPR-based cleavage and RNA ligation. (A) Diagram of the RNA editing with type III CRISPR nucleases and splinted ligation. Type III CRISPR complexes cleave RNA in 6-nt intervals (red triangles), cutting out a portion of the target sequence. The resulting fragments are splint-ligated to introduce edits. (B) SthCsm-mediated cleavage generates 2′3′-cyclic phosphate and 5′-hydroxyl ends (left, substrate for RtcB ligase) that can be converted to 3′-hydroxyl and 5′-phosphate (right, substrate for T4 RNA ligases) using T4 PNK. (C) Two splint designs. Hybridization of the DNA splint A to complementary RNA leaves single-stranded RNA flaps (curved blue lines) that imitate a break in the tRNA anticodon loop, while splint B mimics nick in double-stranded RNA. (D) Comparison of RNA ligation efficiency with T4 RNA ligases (T4 Rnl) and RtcB ligase. Ligation efficiency was measured by performing RT-qPCR across the cut site and quantifying the signal relative to an uncut control (100%). (E) Deep sequencing of RNA ligated with splints depicted in (C). Horizontal black bar shows the target site of the SthCsm complex. (F) Quantification of editing outcomes in (E).

Programmed Deletions in the RNA Virus Genome

After validation with IVT RNA, we next sought to demonstrate how CRISPR-based RNA editing can be applied for the rapid engineering of RNA viruses. As a model, we used SINV, an alphavirus with a ~12.5-kb single-strand, positive-sense RNA genome that has a green fluorescent protein (GFP) reporter gene inserted at the 3′-end (see fig. 4.S2A for details). We designed a guide RNA to cut out a portion of the gfp gene that encodes for tyrosine in position 66 (Y66) of the protein, which would disrupt fluorescence in the encoded protein, such that the edited virus will produce a "dead" GFP (Fig. 4.2A). First, we in vitro transcribed the gfp gene and tested the cleavage activity of the gfp-targeting SthCsm complex. We sequenced the resulting cleavage products, which identified preferential excision of 6 or 12 nt at the target site (fig. 4.S3).

Next, we moved on to edit infectious viral RNA extracted from the supernatant of infected BHK-21 cells (fig. 4.S2B). We incubated viral RNA with a GFP-targeting SthCsm complex and measured the efficiency of cleavage using quantitative polymerase chain reaction (qPCR) (97.8 ± 0.4%; Fig. 4.2B). Cleaved viral RNA was treated with T4 PNK, annealed to a DNA splint designed to select for 12-nt deletion, and ligated with T4 RNA ligase (Fig. 4.2B). Melt curve analysis of the qPCR products spanning the target site identifies a ~1.5°C shift in the melting temperature of the edited sample, which is consistent with a deletion of the expected size and the faster migration of the PCR product on a polyacrylamide gel (Fig. 4.2C and fig. 4.S4, A and B). Edited or unedited viral RNA was transfected into BHK-21, and cells were imaged the next day. Cells transfected with the edited viral RNA produced significantly less GFP signal compared to the unedited control ($P = 0.0097$; Fig. 4.2D and fig. 4.S4, C and D), while viral RNA concentrations in cell supernatants were nearly identical for edited and unedited viral RNA

(P = 0.426 for nsP gene and P = 0.272 for gfp gene; fig. S4E). Last, the PCR performed using a template from the replicating virus suggests programmed deletions at the target site in the gfp gene (Fig. 4.2E).

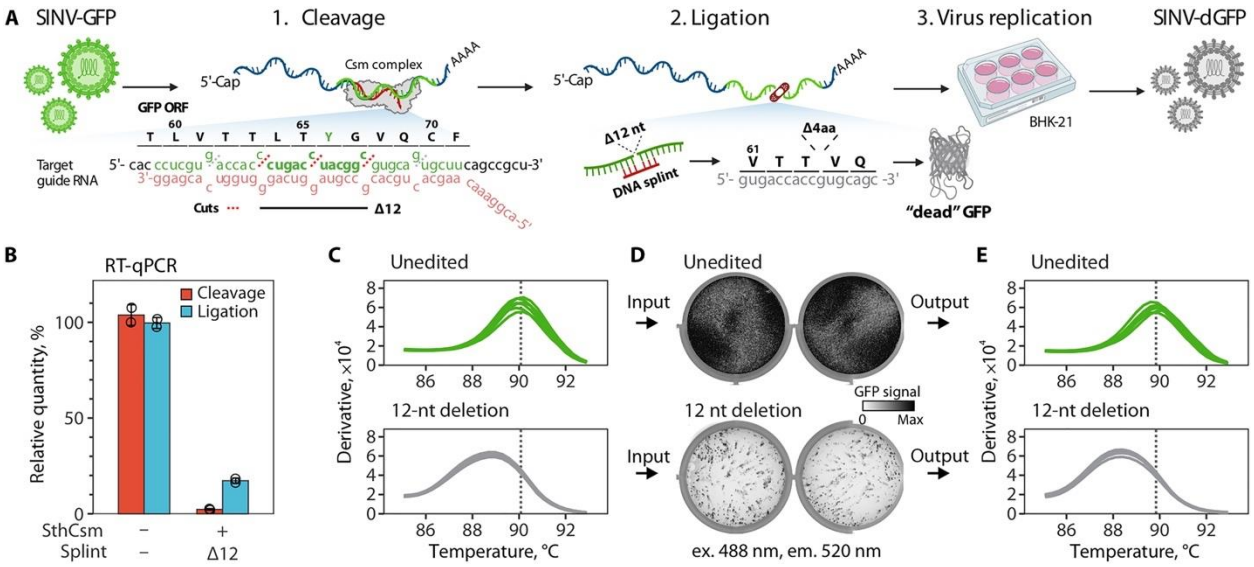


Fig. 4.2. Programmed deletion of specific RNA sequence from the SINV-GFP genome. (A) Diagram showing a pipeline for deleting 12 nt from the GFP open reading frame (ORF) of the Sindbis-GFP (SINV-GFP) virus. See fig. S2A for annotated genome map and fig. S4A for additional experimental details. The deletion eliminates four codons encoding amino acids (L64, T65, Y66, and G67) forming the chromophore of the fluorescent protein, which is expected to ablate fluorescence ["dead" GFP (dGFP)]. Icons for viruses, nucleic acids, 6-well plate, and dead GFP were created with BioRender (www.biorender.com). (B) RNA aliquots collected after the cleavage with the Csm complex (red) and ligation (blue) with T4 RNA ligase were reverse-transcribed and quantified with qPCR. Primers were designed to amplify cDNA across the target site. Relative quantities were calculated by normalizing to the uncut RNA control. (C) Melt curve analysis of the qPCR products generated with RNA after ligation in (B). Peaks indicate the melting temperature of the qPCR products. (D) BHK-21 cells were transfected with unedited or edited RNA of the SINV-GFP (two replicates each), seeded in six-well plates, and imaged 24 hours later to capture the GFP signal (excitation, 488 nm; emission, 520 nm). (E) Melt curve analysis of the RT-qPCR products that were generated with RNA extracted from supernatants of BHK-21 cells 24 hours after transfection. Primers were the same as in (B) and (C).

Sequencing and Isolation of Edited Viruses

To confirm that CRISPR-based RNA editing produces programmed deletions, we sequenced viral genomes extracted from the cell supernatants (fig. 4.S5). Sequencing identified a fraction of viral genomes in the unedited control that eliminated the entire gfp gene (~20%; fig. 4.S5B). These deletions are also present in the edited sample and accumulate during viral replication in the host because there is no selective pressure on the viral genome to retain the nonessential gfp gene. We filtered genomes that retained the gfp and quantified deletions of the sequence targeted by the SthCsm complex (Fig. 4.3A and fig. 4.S5C). The 12-nt deletion, which our protocol was designed to select, was the most prevalent editing outcome (34.5 ± 0.9%), but deletions of 18 nt (17.8 ± 2.5%), 24 nt (7.0 ± 2.1%), 30 nt (1.2 ± 0.5%), and 6 nt (2.2 ± 0.8%) could also be identified in the deep sequencing data.

To isolate edited viral clones, we picked plaques from cells transfected with infectious viral RNA and used this as an inoculum to infect individual wells of naïve cells (Fig. 4.3B and fig. 4.S6). We isolated RNA from 11 wells that were positive for cytopathic effect but negative for GFP. The viral RNA was sequenced using amplicon-based genome sequencing (fig. 4.S7A). All 11 viral genomes contain programmed deletions at the target site. Four of these genomes contain a 12-nt deletion (Fig. 4.3C), while the remaining seven are a mixture of 12-, 18-, or 24-nt deletions (fig. 4.S7).

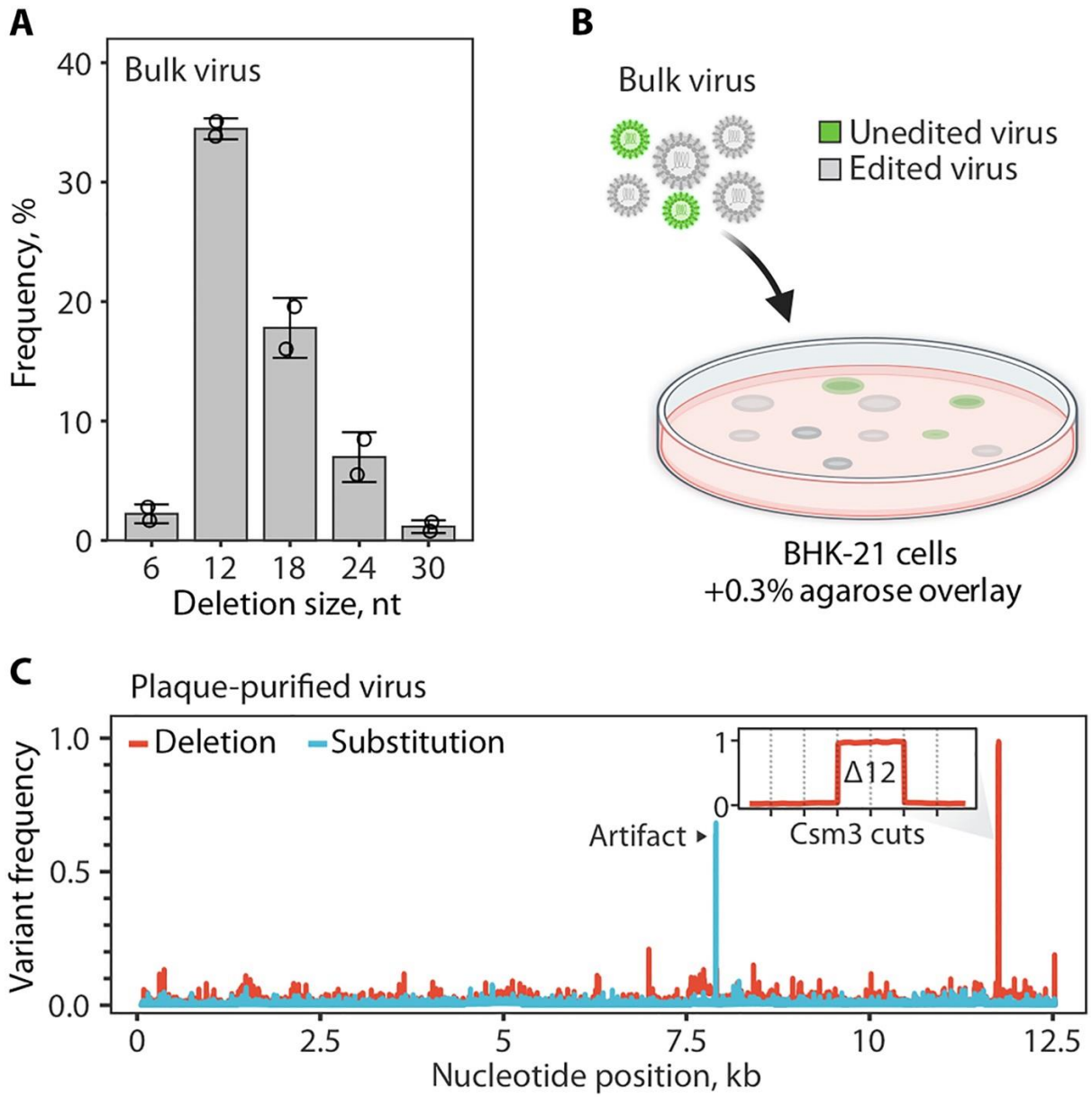


Fig. 4.3. Plaque purification of edited viruses. (A) Distribution of deletion sizes in sequencing reads that span the SthCsm target site in bulk edited virus. Data are shown as mean (n = 2) ± SD. Dots show individual replicates. (B) Diagram of plaque purification approach used for isolating edited virus clones. Icons for viruses and the Petri dish were created with BioRender (www.biorender.com). (C) Frequency of nucleotide variants identified with amplicon-seq of plaque-purified virus. See fig. S7 for details on sequencing strategy and analysis of other plaque-purified viruses. Black arrow indicates a sequencing artifact associated with amplicon-seq (see fig. S7). The inset is a close-up view of the targeted genomic region. Vertical dotted lines mark predicted cut sites of the SthCsm backbone subunit (Csm3) positioned in 6-nt intervals. The 12-nt deletion was intentionally selected using a DNA splint.

Programmable Insertion of New RNA in the Viral Genome

To demonstrate the versatility of this technology, we designed synthetic RNA and complementary DNA splint to insert a new sequence at the SthCsm cleavage site. Instead of SINV-GFP, this time, we used recombinant SINV with a blue fluorescent protein (BFP) gene (SINV-BFP). The nucleotide sequence of SINV-BFP is identical to SINV-GFP except for a single-nucleotide substitution (11,762U>C) in the reporter gene that changes tyrosine 66 to histidine (Y66H), resulting in a change in the emission spectrum (i.e., color) of the fluorescent protein. In the synthetic RNA insert, we included a single-nucleotide substitution to convert BFP to GFP (11,762C>U) and two additional silent mutations to distinguish the engineered gfp from wild-type gfp used in other experiments (11,761C>A and 11,764C>U; Fig. 4.4A and fig. 4.S8). We edited the SINV-BFP genome using an SthCsm complex targeting the bfp gene, a DNA splint, and a synthetic RNA insert. We transfected the edited RNA and imaged cells the next day. We detected cells that are either blue or green, indicating a mixture of edited and unedited RNA (Fig. 4.4, B and C). To quantify the editing efficiency and to verify that the GFP-positive cells are infected with the edited virus, we sequenced bulk and plaque-purified viruses. These results indicated that 5.1 ± 0.7% of the bulk virus is edited, and the GFP-positive clones contain the intended modification (figs. 4.S9 and 4.S10).

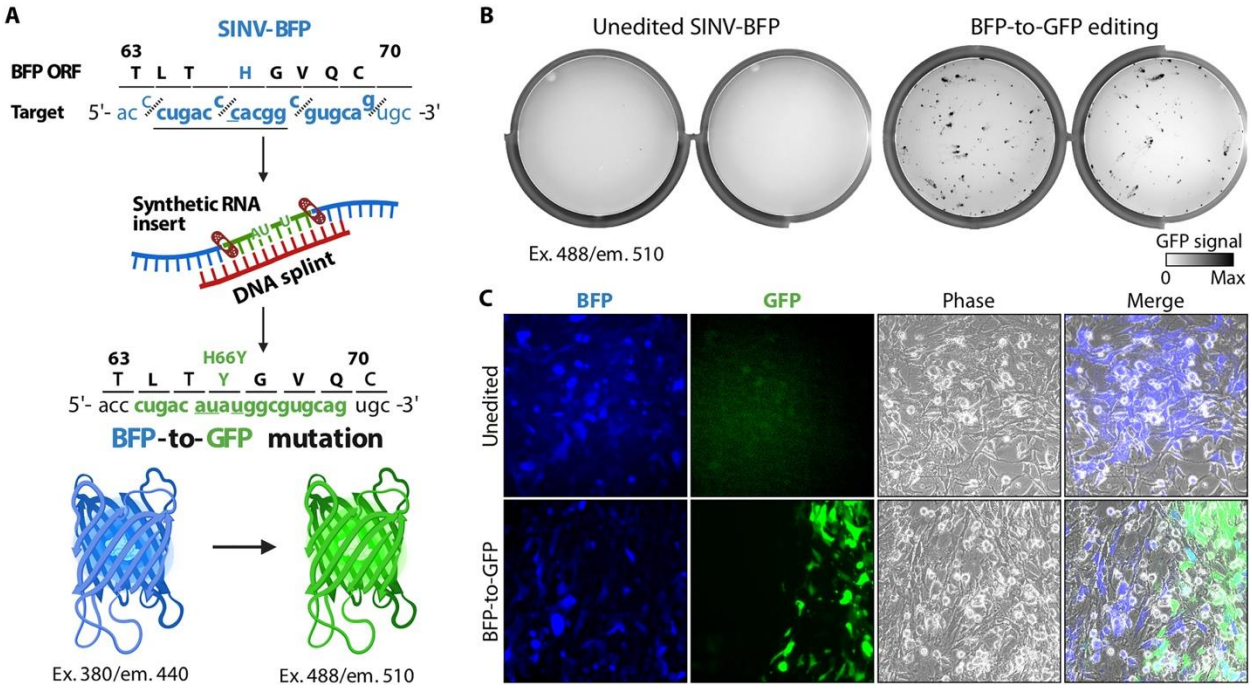


Fig. 4.4. Programmed substitutions in viral genomes. (A) Diagram showing editing strategy for substituting 12 nt (underlined) in the BFP ORF of the recombinant Sindbis-BFP (SINV-BFP) virus. The C to U substitution at 11,762 recodes histidine residue for tyrosine (H66Y), which changes the fluorescence from blue to green (BFP to GFP). Additional silent substitutions (underlined, 11,761C>A and 11,764 C>U) were included to distinguish edited RNA from the SINV-GFP used for experiments shown in Fig. 2. Icons for nucleic acids, band-aid, BFP, and GFP were created with BioRender (www.biorender.com). (B) BHK-21 cells were transfected with the unedited or edited viral RNA (two replicates each), seeded on six-well plates, and imaged after 24-hour incubation using 488-nm excitation laser and 520-nm emission filter. This laser and filter setting does not detect BFP (excitation, 380 nm; emission, 440 nm) in the unedited control, while replication of the edited BFP-to-GFP viruses is visible. (C) Transfected BHK-21 cells shown in (B) were imaged with an inverted fluorescent microscope at ×20 magnification.

Discussion

Several years after the Asimolar conference on recombinant DNA in 1975, the first complementary DNA (cDNA) clones of RNA viruses were created (25, 26). Recombinant viruses are a cornerstone of basic research science, leading to fundamental new discoveries in biology, new therapeutics, and vaccines for viral diseases (27, 28). Following the outbreak of SARS-CoV-2, the viral RNA genome (~30 kb) was reverse-transcribed, and 5 to 12 fragments of the viral genome were cloned into multiple plasmids or an artificial chromosome, which are modified, stitched back together to reconstruct the genome, reverse-transcribed, and transfected into permissive cells (29–31). While recombinant SARS-CoV-2 clones were instrumental in screening antivirals (32), studying viral pathogenesis (31), and linking genotypes to phenotypes (33–35), the reconstitution of mutant viruses is slow and technically challenging, which limits our ability to test new hypothesis (36).

Here, we establish an approach for RNA editing that enables fast and programmable deletions, insertions, and substitutions in any RNA by design (fig. 4.S11A). Unlike DNA editing, which often relies on protein-mediated recognition of a specific sequence motif, type III CRISPR systems only require complementary base pairing between the RNA guide and the RNA target (16), which eliminates additional sequence requirements and improves target site versatility. While direct RNA editing significantly accelerates and simplifies the process, the current protocol requires the expression and purification of a new SthCsm:crRNA ribonucleoprotein complex for each new RNA target. However, a single subunit type III-E CRISPR complex can be loaded with crRNA guides in vitro, which would simplify this process and make it readily accessible to laboratories that do not routinely purify recombinant proteins (19, 37).

Furthermore, type III-E CRISPR complexes have the potential to produce a single deletion rather than a series of cleavage events, which might be advantageous in some editing scenarios.

The utility of this RNA editing technology is currently restricted to applications where the editing is performed in vitro. Because most of the plant and animal viruses contain RNA genomes, we anticipate that this is where these tools will have the most immediate impact. Much like DNA editing in cells, edits made in the genomes of RNA viruses are propagated during the infection and inherited by the viral progeny. Edited viruses are plaque-purified and sequence-verified, which altogether can be performed in 3 to 4 days (fig. 4.S11B).

We anticipate that in vivo RNA editing that relies on the cut-and-repair reaction, similar to those routinely used for DNA engineering, will emerge over time, but applications for base editing might be more immediate. Mutations in type III CRISPR-Cas complexes that eliminate target RNA cleavage but preserve target RNA binding have been well characterized, and the architecture of type III complexes may facilitate the multivalent display of base editors or repair enzymes. Moreover, we expect that diverse type III systems will each have unique properties that are beneficial for applications in editing. Like type II (i.e., Cas9), type V (i.e., Cas12), and type VI (i.e., Cas13), the type III-E systems are composed of a single polypeptide (38), and the purified protein can be loaded with synthetic crRNAs in vitro (19, 37). Loading Cas effectors with synthetic RNA guides was an important advance for DNA editing, and similar advances are anticipated for RNA editing.

Materials and Methods

Plasmids

E. coli codon-optimized genes for expression of the type III-A Csm complex from S. thermophilus (SthCsm) were synthesized by GenScript and cloned into three different vectors. Genes for SthCsm3 with an in-frame N-terminal His6-TwinStrep-SUMO affinity tag, SthCsm4, and SthCsm5 were cloned into pRSF-1b (pRSF-1b-SthCsm3,4,5) using Nco I and Pac I restriction sites. SthCas10 and SthCsm2 were cloned into pACYCDuet1 (pACYCDuet1-SthCas10-SthCsm2) using Nco I and Sac I restriction sites. SthCas6 and CRISPR array containing five repeats and four identical spacers, designed to target the N-gene of SARS-CoV-2, were expressed in the pMA vector (pMA-SthCas6-4xcrRNA-N-gene, GenScript) (39). Cas10 gene with 15HD>HA and 573GGDD>GGAA mutations was synthesized and cloned in pACYCDuet1-SthCas10-SthCsm2 plasmid using Nco I and Not I restriction sites. These mutations inactivate the nuclease and polymerase domains of Cas10, which markedly improves protein yields (pACYCDuet1-SthCas10HD/DDSthCsm2; fig. 4.S3, A and B). A CRISPR array containing two repeats and one spacer targeting the gfp gene of recombinant SINV (SINV-GFP) was generated by annealing and extending two partially cDNA oligos with Q5 polymerase [New England Biolabs (NEB)] (table 4.S1). The resulting fragment was cloned using Nco I and Mfe I restriction enzymes into pMA-SthCas6-4xcrRNA-N-gene to substitute the CRISPR array. To generate a CRISPR array targeting the bfp gene, the plasmid encoding CRISPR targeting gfp (pMA-SthCas6-1xcrRNA-GFP) was mutagenized using Q5 site-directed mutagenesis (table 4.S1). All plasmid sequences were confirmed with whole-plasmid sequencing at Plasmidsaurus (https://plasmidsaurus.com/).

Nucleic Acids

DNA oligos were purchased from Eurofins, and RNA oligos were purchased from Integrated DNA Technologies (tables 4.S1 to 4.S5). The RNA donor was annealed to a DNA splint at a 1:1 ratio in buffer [10 mM tris-HCl (pH 7.8) and 50 mM NaCl]. Annealing was performed by incubating samples at 95°C for 5 min and then cooling down to 25°C at −0.1°C/s rate (table 4.S2). In vitro transcribed RNA of SARS-CoV-2 N-gene was synthesized with MEGAscript T7 (Thermo Fisher Scientific) using PCR product generated from SARS-CoV-2 WA1 cDNA and primers N0.5_F and N3-5_R (Eurofins) (table 4.S3). In vitro transcribed GFP RNA was synthesized with MEGAscript T7 (Thermo Fisher Scientific) using PCR product generated from pTE3′2 J-GFP plasmid and primers T7_SINV_sgmRNA_F and SINV_polyA_R (Eurofins) (table S3). Transcribed RNAs were purified using the Monarch RNA Cleanup Kit (NEB).

Viruses

Recombinant SINV expressing GFP (SINV-GFP) was reconstituted from a cDNA copy encoded in pTE3′2 J-GFP vector provided by M. Flenniken (40, 41). RNA (500 ng), in vitro transcribed from the pTE3′2 J-GFP template, was electroporated into 1 million BHK-21 cells, and the supernatant was harvested 24 to 48 hours after electroporation (passage 0). To generate recombinant SINV expressing BFP (SINV-BFP), the gfp gene in the DNA template was mutagenized with spliced PCR to make 11,762T>C mutation (table 4.S4). The spliced PCR fragment was cloned in the pTE3′2 J-GFP vector using the Xba I restriction enzyme.

Cell cultures

BHK-21 cells [American Type Culture Collection (ATCC), CCL-1] were provided by M. Jutila (Montana State University) and maintained at 37°C and 5% CO2 in Eagle’s modified

Eagle medium (EMEM; ATCC, catalog no.30-2003) supplemented with 10% fetal bovine serum (FBS; ATLAS Biologicals, lot no. F31E18D1) and penicillin (50 IU/ml) and streptomycin (50 mg/ml). BL21(DE3) protein expression strain was obtained from Thermo Fisher Scientific.

Protein Expression and Purification

Expression and purification of the SthCsm complexes were performed according to the following protocol. The expression vectors pRSF-1b-SthCsm3,4,5, pACYCDuet1-SthCas10-SthCsm2, pMA-SthCas6-4xcrRNA-N-gene, and pMA-SthCas6-1xcrRNA-GFP or pMA-SthCas6-1xcrRNA-BFP were transformed in BL21(DE3) cells and grown in LB broth (Lennox) (Thermo Fisher Scientific) supplemented with kanamycin (50 μg/ml), chloramphenicol (34 μg/ml), and ampicillin (100 μg/ml) at 37°C to an OD600 (optical density at 600 nm) of 0.5. Cultures were then incubated on ice for 1 hour and then induced with 0.5 mM isopropyl-β-d-thiogalactopyranoside (IPTG) for overnight expression at 16°C. Cells were lysed with sonication in lysis buffer [20 mM tris-HCl (pH 8), 500 mM NaCl, 1 mM dithiothreitol (DTT), and 1 mM EDTA], and the lysate was clarified by centrifugation at 10,000g for 25 min at 4°C. His6-TwinStrep–tagged protein was bound to a StrepTrap HP column (Cytiva) and washed with lysis buffer. The protein was eluted with lysis buffer supplemented with 2.5 mM desthiobiotin and concentrated (10,000 molecular weight cutoff Corning Spin-X concentrators) at 4°C. Last, the protein was purified using a HiLoad Superdex 200 10/300 column (Cytiva) in storage buffer [10 mM tris-HCl (pH 8), 1 mM DTT, 0.1 mM EDTA, 300 mM monopotassium glutamate, and 5% glycerol]. Fractions containing the target protein were pooled, concentrated, aliquoted, flash-frozen in liquid nitrogen, and stored at −80°C.

SINV Genomic RNA Purification

Six T300 flasks of confluent BHK-21 cells were infected with SINV-GFP or SINV-BFP viruses (passage 0). Viral stock (50 µl) was diluted in 50 ml of phosphate-buffered saline (PBS) supplemented with 3% FBS, 0.5 mM MgCl2, and 0.9 mM CaCl2. The medium was discarded, and 8 ml of diluted viral stock was added to each flask. Cells were incubated for 1 hour (37°C, 5% CO2), PBS was removed, and 40 ml of fresh EMEM supplemented with 5% FBS, penicillin (50 IU/ml), and streptomycin (50 mg/ml) was added. Supernatants were collected 24 hours after infection, combined, and spun down for 10 min at 3000g to remove cell debris. The supernatant was filtered through a 0.45-µm filter and ultracentrifuged in SW32 rotor at 100,000g for 1 hour at 4°C. The supernatant was discarded, and the pellet was dissolved in 300 µl of PBS. Ten volumes of TRIzol Reagent (Invitrogen) were added to one volume of viral suspension in PBS, mixed, and incubated for 10 min at room temperature. Then, 0.2 volume of chloroform per 1 volume of TRIzol was added, mixed, and centrifuged at 5000g for 10 min at 4°C to separate aqueous phase. The aqueous phase containing RNA was transferred into a new tube, mixed with one volume of 98% ethanol, and purified with the Monarch RNA Cleanup Kit (NEB). RNA concentration was measured using NanoDrop. This protocol yields ~80 to 90 µg of RNA, which is 0.3 to 0.35 µg per milliliter of supernatant. The extracted viral RNA is infectious, with 8.3 ± 0.6 × 104 plaque-forming units per microgram of the stock (fig. 4.S2B).

Editing of in vitro Transcribed RNA

For the experiment shown in Fig. 1C, IVT RNA of SARS-CoV-2 N-gene (0.5 µM) was mixed with SthCsm (1 µM) in a reaction buffer [20 mM Hepes (pH 8.0), 50 mM KCl, 1 mM Mg(CH3COO)2, and bovine serum albumin (0.1 mg/ml)] and incubated for 1 hour at 37°C. The

RNA cleavage products were purified using the Monarch RNA Cleanup Kit (NEB) and divided into two parts. One part of RNA was treated with T4 PNK (NEB) in a reaction buffer [50 mM tris-HCl (pH 7.5), 10 mM MgCl2, 1 mM adenosine triphosphate (ATP), and 10 mM DTT] for 1 hour at 37°C. The remaining RNA was incubated in the same buffer but without the enzyme. Next, RNA was purified with the Monarch RNA Cleanup Kit (NEB). PNK-treated and nontreated RNAs were annealed to splints (1:1 ratio) in a reaction buffer [50 mM tris-HCl (pH 7.5), 10 mM MgCl2, and 1 mM DTT] for 5 min at 95°C, followed by a ramp down of 0.1°C/s to 25°C. Then, three different ligases were tested to repair RNA. T4 Rnl1 was mixed with splinted RNA in a 1× T4 Rnl buffer (NEB) supplemented with ATP (1 mM) and polyethylene glycol (PEG) 8000 (10%) and run for 1 hour at 25°C. The reaction with T4 Rnl2 was performed in 1× T4 Rnl2 (NEB) supplemented with PEG 8000 (10%) and run for 1 hour at 25°C. The reaction with RtcB was performed in 1× RtcB reaction buffer (NEB) supplemented with guanosine triphosphate (GTP) (0.1 mM), MnCl2 (1 mM), and PEG 8000 (10%) and run for 1 hour at 37°C. The reactions were diluted 10 times and used for reverse transcription qPCR (RT-qPCR).
The experiment shown in Fig. 1 (D to F) was performed as described above with slight modifications. All RNA was T4 PNK–treated after the cleavage with SthCsm. T4 Rnl1 was used to ligate RNA annealed to splint A. To ligate RNA hybridized with splint B, T4 Rnl2 was used. Tenfold diluted reactions were used for RT-qPCR and amplicon sequencing (amplicon-seq) with Oxford Nanopore.

Deletion of Viral RNA

For viral RNA editing (Figs. 2 to 4), 1.64 μg of viral genomic RNA (gRNA; ~0.04 pmol based on the molecular weight of the full genome) was mixed with SthCsm-Cas10HD/DD (4

pmol) in 10 µl of reaction buffer [50 mM tris-HCl (pH 7.5), 1 mM DTT, and ribonuclease (RNase) inhibitor (1 U/µl) (NEB)] and incubated for 1 hour at 37°C. The RNA cleavage products were purified using the Monarch RNA Cleanup Kit (NEB) and mixed with 5 U of T4 PNK (NEB), DNA splint (0.4 pmol) in 10 µl of reaction buffer [50 mM tris-HCl (pH7.5), 1 mM DTT, and 1 mM MgCl2] supplemented with 1 mM ATP and RNase inhibitor (1 U/µl; NEB). The reaction was incubated for 1 hour at 37°C. After the incubation, 5 µl of a mix containing 5 U of T4 Rnl2 (NEB), 1 mM ATP, 3 µl of 50% PEG 8000, and RNase inhibitor (1 U/µl; NEB) in reaction buffer [50 mM tris-HCl (pH7.5), 1 mM DTT, and 1 mM MgCl2] was added directly to the previous reaction. Ligation reactions were incubated for 1 hour at 25°C.

Substitution of Viral RNA

The RNA insert (10 µM) was hybridized to the DNA splint (10 µM) by heating the mixture to 95°C and slow cooling to 25°C at 0.1°C/s rate in 10 mM tris-HCl (pH 7.8) buffer with 100 mM NaCl. Cleavage of viral RNA was performed as described above. After cleanup, cleaved viral genome was mixed with the RNA insert: DNA splint duplex (0.4 pmol) and 5 U of T4 PNK (NEB) in 10 µl of reaction buffer [50 mM tris-HCl (pH7.5), 1 mM DTT, and 1 mM MgCl2] supplemented with 1 mM ATP and RNase inhibitor (1 U/µl; NEB). After 1 hour at 37°C, 5 µl of a mix containing 5 U of T4 Rnl2 (NEB), 1 mM ATP, 3 µl of 50% PEG 8000, and RNase inhibitor (1 U/µl; NEB) in reaction buffer [50 mM tris-HCl (pH7.5), 1 mM DTT, and 1 mM MgCl2] was added and incubated for 1 hour at 25°C to perform ligation.

Electroporation of Edited Viral RNA

Ligation reaction (10 µl) was mixed with 106 BHK-21 cells in 100 µl of Ingenio Electroporation Solution (Mirus) and transferred to a 0.2-cm cuvette (Mirus). The electroporation

was performed using the Ingenio EZporator Electroporation System (Mirus) by pulsing with 150 V and 950 to 1050 μF. Immediately after electroporation, 500 μl of EMEM (ATCC) medium supplemented with 10% FBS (no antibiotics) was added directly in the cuvette. Cell suspensions were transferred to tissue culture plates with the addition of more media and incubated with 5% CO2 at 37°C.

Fluorescence Imaging

Viral infections were imaged with an Amersham Typhoon 5 scanner (Cytiva) with the default Cy2 (excitation, 488 nm; emission, 520 nm) setting. Images of cells infected with SINV BFP-to-GFP edited virus were captured with Nikon Ti-Eclipse inverted microscope (Nikon Instruments) equipped with a SpectraX light-emitting diode (LED) excitation module (Lumencor) and emission filter wheels (Prior Scientific). Fluorescence imaging used excitation/emission filters and dichroic mirrors for GFP and 4′,6-diamidino-2-phenylindole (DAPI) (Chroma Technology Corp.). Images were acquired with Plan Fluor 20Ph objective and an iXon 896 electron-multiplying charge-coupled device (EM-CCD) camera (Andor Technology Ltd.) in NIS-Elements software. Nucleic acid gels were stained with SYBR Gold (Thermo Fisher Scientific) and imaged using the Cy2 setting with an Amersham Typhoon 5 scanner (Cytiva). Protein gels were stained with Coomassie stain (in-house), washed, and imaged with infrared-short default setting with an Amersham Typhoon 5 scanner (Cytiva).

Plaque Purification

The cells electroporated with edited viral RNA were serially diluted, and 10,000 cells were seeded on a monolayer of BHK-21 cells in a 100-mm petri dish. The cells were incubated (37°C, 5% CO2) for 1 hour. Next, the medium was removed, and the medium supplemented with

0.3% agarose was overlaid. Petri dishes were imaged 16 to 18 hours after infection using an Amersham Typhoon 5 scanner (Cytiva). The individual plaques were picked using sterile tips under a light microscope and inoculated in naive BHK-21 cells for propagation.

Reverse Transcription qPCR

RNA was extracted from 140 µl of cell supernatants with the QIAamp Viral RNA Mini Kit (QIAGEN, #52906) and eluted in 60 µl of buffer AVE (QIAGEN). RNA ligations were diluted 10-fold and used for RT without extraction. The RT reactions were performed with the 2× LunaScript RT SuperMix Kit (NEB) with 5 µl of RNA input in 10 µl of total volume. The RT reactions containing cDNA were diluted 100-fold, and 5 µl was used for qPCRs with 2× Universal SYBR Green Fast qPCR Mix (ABclonal). Each qPCR contained 5 µl of cDNA, 4.2 µl of nuclease-free water, 0.4 µl of 10 mM primers (table S4), and 10 µl of 2× mastermix. Nuclease-free water was used as negative template control. Three technical replicates were performed for each sample. Amplification was performed in a QuantStudio 3 Real-Time PCR System instrument (Applied Biosystems) as follows: 95°C for 3 min, 40 cycles of 95°C for 5 s, and 60°C for 30 s, followed by melt curve analysis. Results were analyzed in Design and Analysis app at Thermo Fisher Connect Platform.

Nanopore Sequencing

Edited IVT RNA of SARS-CoV-2 N-gene was sequenced using the Oxford Nanopore with Ligation Sequencing Kit (SQK-LSK109). RNA was reverse-transcribed using the LunaScript RT SuperMix Kit (NEB), and cDNA was used for PCR with primers nCoV-96_L/nCoV-96_R (table S4) and Q5 polymerase (NEB). DNA was purified using magnetic beads (Omega Bio-tek, M1378-01) and used to prepare sequencing libraries as described in SQK-

LSK109 protocol using Native Barcoding Expansion 1-12 (EXP-NBD104). Resulting library (~20 ng) was sequenced on the Nanopore MinION with R9.4.1 flow cell. The sequencing run was performed in the high-accuracy base-calling mode in the MinKNOW software.

Products of RNA cleavage with gfp-targeting SthCsm were sequenced with Oxford Nanopore using a direct cDNA sequencing kit (SQK-DCS109). After incubation with SthCsm complex, RNA cleavage fragments were purified with the Monarch RNA Cleanup Kit (NEB). Then, RNA was treated with T4 PNK (NEB) for 1 hour at 37°C and purified with RNAClean XP beads (Beckman Coulter). The RNA ends were extended using E. coli poly(A) polymerase (NEB), and RNA was cleaned up again using RNAClean XP beads (Beckman Coulter). About 100 ng of poly(A) RNA was used as an input to prepare cDNA for nanopore sequencing as instructed in SQK-DCS109 protocol using Native Barcoding Expansion 1-12 (EXP-NBD104). About 200 fM of the library was loaded on Nanopore MinION (R9.4.1 flow cell). The flow cell was primed, and the library was loaded according to the Oxford Nanopore protocol (SQK-LSK109 kit). The sequencing run was performed in the high-accuracy base-calling mode in the MinKNOW software.

The bulk-edited SINV genomes were sequenced with Oxford Nanopore using a direct cDNA sequencing kit (SQK-DCS109). Plaque-purified clones were sequenced with amplicon-seq using the Ligation Sequencing Kit (SQK-LSK114). To perform amplicon-seq, RNA was reverse-transcribed using the LunaScript RT SuperMix Kit (NEB). Diluted cDNA (100-fold) was used for two multiplex PCRs with Q5 polymerase (NEB) with two primer pools (table S5). The primers were designed using the Primal Scheme web tool (http://primalscheme.com), synthesized (Eurofins), and diluted as previously described (42). After amplification, two

multiplex PCRs for each sample were combined and purified using magnetic beads (Omega Bio-tek, M1378-01). About 130 ng of DNA was used to prepare sequencing libraries as described in SQK-LSK114 protocol using Native Barcoding Kit 24 V14 (SQK-NBD114.24). About 20 fM of the library was loaded on Nanopore MinI[ON (R10.4.1 flow cell, in accurate mode (260 bases per second)]. For direct cDNA sequencing, 420 µl of viral supernatant was extracted with the QIAamp Viral RNA Mini Kit (QIAGEN, #52906) and eluted in 30 µl of buffer AVE (QIAGEN). Extracted RNA (7.5 µl) was reverse-transcribed, and the second strand was synthesized as instructed by the manufacturer's protocol (SQK-109). The resulting double-stranded DNA was used for sequencing library preparation with Native Barcoding Kit 24 V14 (SQK-NBD114.24). The flow cell was primed, and sequencing libraries were loaded according to the Oxford Nanopore protocol (SQK-LSK114). The sequencing run was performed in the fast base-calling mode in the MinKNOW software. Sequencing data (.fast5 files) were base-called with the super-accuracy model (dna_r10.4.1_e8.2_260bps_sup.cfg) using guppy_basecaller and demultiplexed with guppy_barcoder (both from Oxford Nanopore). Base calling was performed on the Tempest High Performance Computing System, operated and supported by University Information Technology Research Cyberinfrastructure at Montana State University.

Sequencing Data Analysis

Nanopore sequencing reads generated with R9.4.1 flow cells were base-called in high-accuracy mode in MinKNOW, and sequencing errors were corrected using the isONcorrect package (43). Nanopore sequencing reads generated with R10.4.1 flow cells were collected at a 260-bps speed and base-called in super-accuracy mode with guppy_basecaller using dna_r10.4.1_e8.2_260bps_sup.cfg configuration. Sequencing reads were aligned to the reference

using minimap2 (v2.17 -r954-dirty) with Nanopore preset (-ax map-ont setting). Alignments were converted to BAM format, sorted, and indexed using samtools v1.13. Sequencing depth was computed with the samtools depth program. Sequencing reads generated with amplicon-seq were trimmed to remove primer binding regions using the samtools ampliconclip. Genome-wide deletion and substitution frequencies were quantified and normalized by calculating frequency of each nucleotide variant at every position in the viral genome using the pileup function in Rsamtools v2.8.0 package in R. Deletion sizes were identified using the extract-junctions.py script (github.com/hyeshik/sars-cov-2-transcriptome/blob/master/nanopore/_scripts/extract-junctions.py) (44). Deletions were quantified and filtered using custom R scripts. Reads with complete deletions of gfp or bfp were identified, and the resulting read ID lists were used to filter alignment files (.BAM) using the samtools view. Deletion counts at the SthCsm cleavage site were normalized to the number of reads that span the target region, which was computed by extracting read information from BAM files using the bamtobed function in bedtools v2.30.0 package and filtering reads by start and end coordinate. All plots were created using ggplot2 v3.3.5 package in RStudio software v2023.03.0.

Statistical Analysis

RNA editing experiments were performed in two replicates. All RT-qPCRs were performed in two technical replicates. Statistical comparisons in fig. S4 (D and E) were performed in R with Welch's unequal variances t test using t.test function from stats package. Significance levels are as follows: *P < 0.05; **P < 0.01; ***P < 0.001; ns, nonsignificant.

Acknowledgments

We thank M. Flenniken for providing the cDNA clone of the Sindbis-GFP virus (pTE3′2 J-GFP) and the related protocols. We thank M. Jutila for providing BHK-21 cells. We thank M. Taylor for the use of the fluorescent microscope. Some of the diagrams and schematics were created with BioRender (www.biorender.com). Base calling of Nanopore sequencing data was performed with the Tempest High Performance Computing System, which is maintained by University Information Technology Research Cyberinfrastructure at Montana State University. Funding: This work was supported by NIH grants 1K99AI171893-01 (A. Nemudryi) and R35GM134867 (to B.W.).

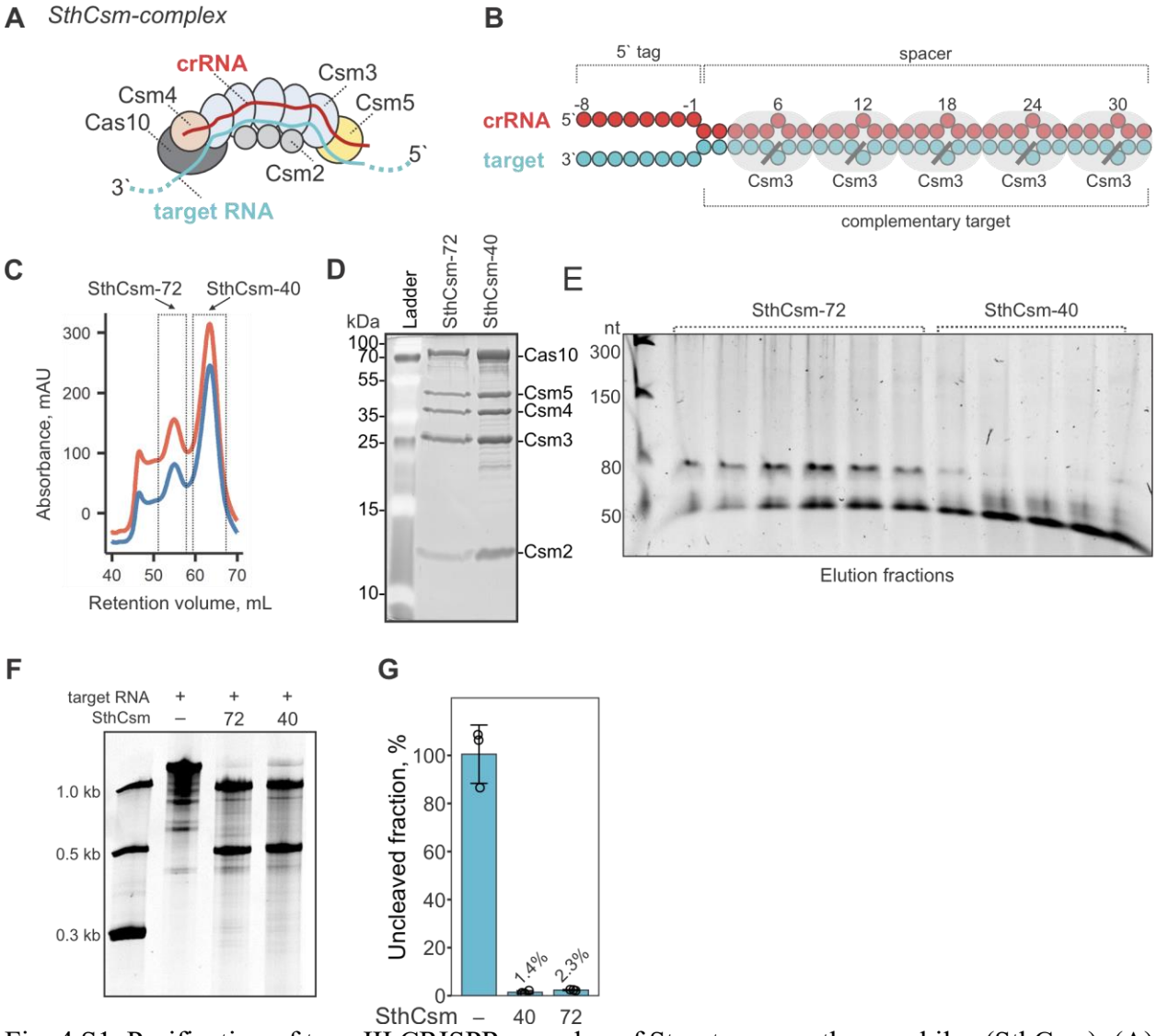


Fig. 4.S1. Purification of type III CRISPR complex of Streptococcus thermophilus (StCsm). (A) Diagram of the StCsm complex with crRNA and target RNA. The complex comprises five protein subunits (i.e., Csm2-5 and Cas10), which self-assemble along the crRNA (red). Csm3 subunits in the backbone of the complex cleave target RNA (blue).
(B) Diagram of the crRNA (red) bound to the target RNA (blue). Circles depict ribonucleotides. Multiple copies of the Csm3 subunit in the complex backbone cut bound RNA at six-nucleotide intervals (diagonal black lines).
(C) Size-exclusion chromatography (SEC) of affinity-purified StCsm complex. SEC was performed using a S200 10/300 size-exclusion column (Cytiva). StCsm complex eluted in two peaks that correspond to different crRNA lengths. StCsm-72 (fractions #6-15) contains an RNA that is 72 nts and StCsm-42 (fractions #20-32) contains an RNA that is 42 nts (16).
(D) SDS-PAGE of purified StCsm-complexes (72 and 42) (~3 ug/lane).
(E) RNAs were extracted from SEC fractions containing StCsm complex shown in (C) and loaded on 12% UREA-PAGE.

(F) Cleavage of IVT target RNA with SthCsm-complexes. Reactions were incubated for 1 h at 37°C, purified with Monarch RNA Cleanup kit (NEB) and loaded on 10% UREA-PAGE.
(G) Cleavage efficiency was quantified with RT-qPCR across the cut site. Data is shown as the mean of two replicates ± one standard deviation.

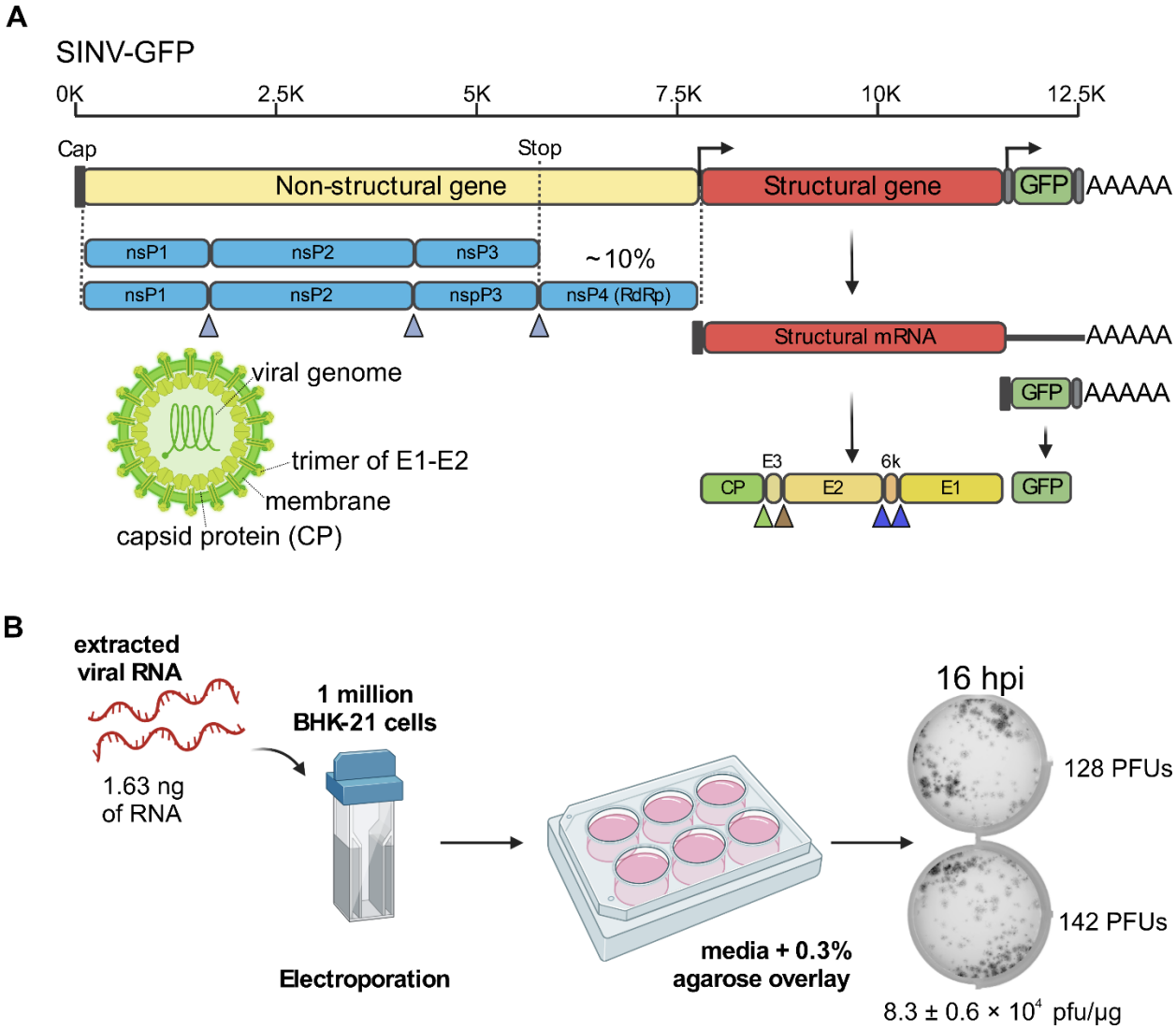


Fig. 4.S2 Architecture of recombinant Sindbis GFP virus (SINV-GFP) and tittering viral RNA. (A) Sindbis virus (alphavirus genus, Togaviridae family) has a ~12.5-kilobase single strand, capped and polyadenylated positive-sense RNA genome with two open reading frames that encode non-structural (nsP1-nsP4) and structural proteins (CP, E1-E3, 6K). Triangles indicate proteolysis sites. The structural gene is expressed from an internal promoter, which generates a subgenomic messenger RNA (sgmRNA). The recombinant Sindbis-GFP virus (SINV-GFP) used in this work has a gfp gene integrated at the 3'-end of the genome, which is expressed from an internal promoter and makes additional sgmRNA. Diagram of the viral genome and viral particle were created with BioRender (https://biorender.com/).(B) Viral RNA, concentrated and extracted from supernatants of infected cells, was diluted andtransfected in BHK-21 cells (see Methods). Cells were seeded in a 6-well plate, and after 1 hincubation at 37°C with 5% CO2 media was replaced with fresh media supplemented with 0.3%agarose. After 16 h incubation, cells were imaged using an Amersham Typhoon 5 scanner. Viral plaque forming units (PFUs) were quantified and used to determine viral titer in the undiluted RNA stock.

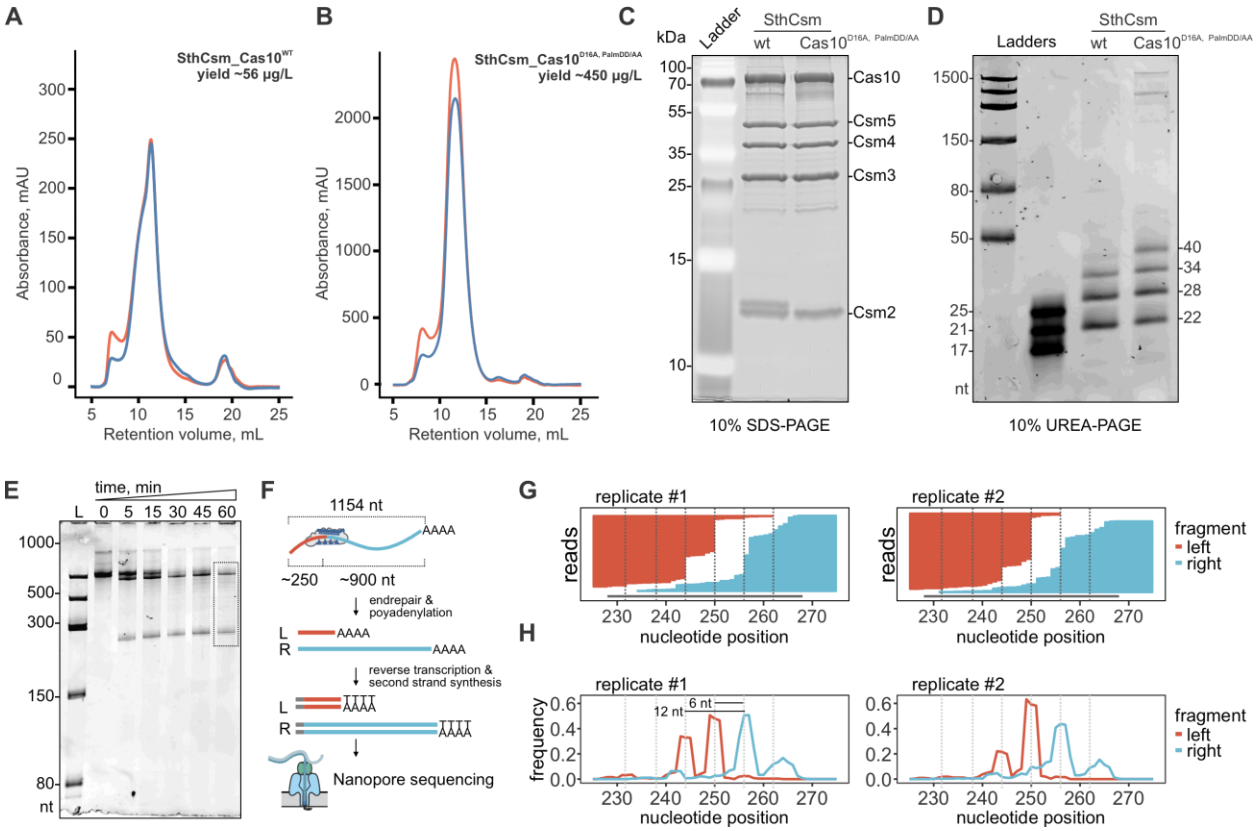


Fig. 4.S3. Purification and characterization of SthCsm-complexes targeting gfp gene in Sindbis virus genome. (A) Size-exclusion chromatography (SEC) of affinity purified wild-type SthCsm. (B) SthCsm complex containing a mutated Cas10 HD-nuclease (D16 to A16, Cas10D16A) and mutated polymerase (D575 and D576 to A575 and A576, Cas10PalmDD/AA) active sites. (C) SDS-PAGE of purified SthCsm complexes in (A) and (B) (~10 ug/lane).(D) RNAs were extracted from SthCsm complexes and were separated in 12% UREA-PAGE.(E) Cleavage assay with in vitro transcribed target RNA. Cleavage reactions were stopped atdifferent time points by adding 2 volumes of phenol:chloroform, extracted, and separated with8% UREA-PAGE. (F) Diagram of Nanopore sequencing strategy used to determine cleavage sites. L – left cleavage fragment, R – right cleavage fragment (G) Mapping of reads generated by sequencing cleavage fragments produced with SthCsm after 60 min incubation. Red and blue lines show reads generated from left and right cleavage products, respectively. Horizontal black line shows target region complementary to the crRNA. Vertical dotted lines mark predicted Csm3 cut sites. (H) Sequencing reads that end (left fragments, red) or start (right fragments, blue) within the target region were quantified, and resulting frequency was normalized to sequencing depth. Vertical dotted lines mark predicted Csm3 cut sites.

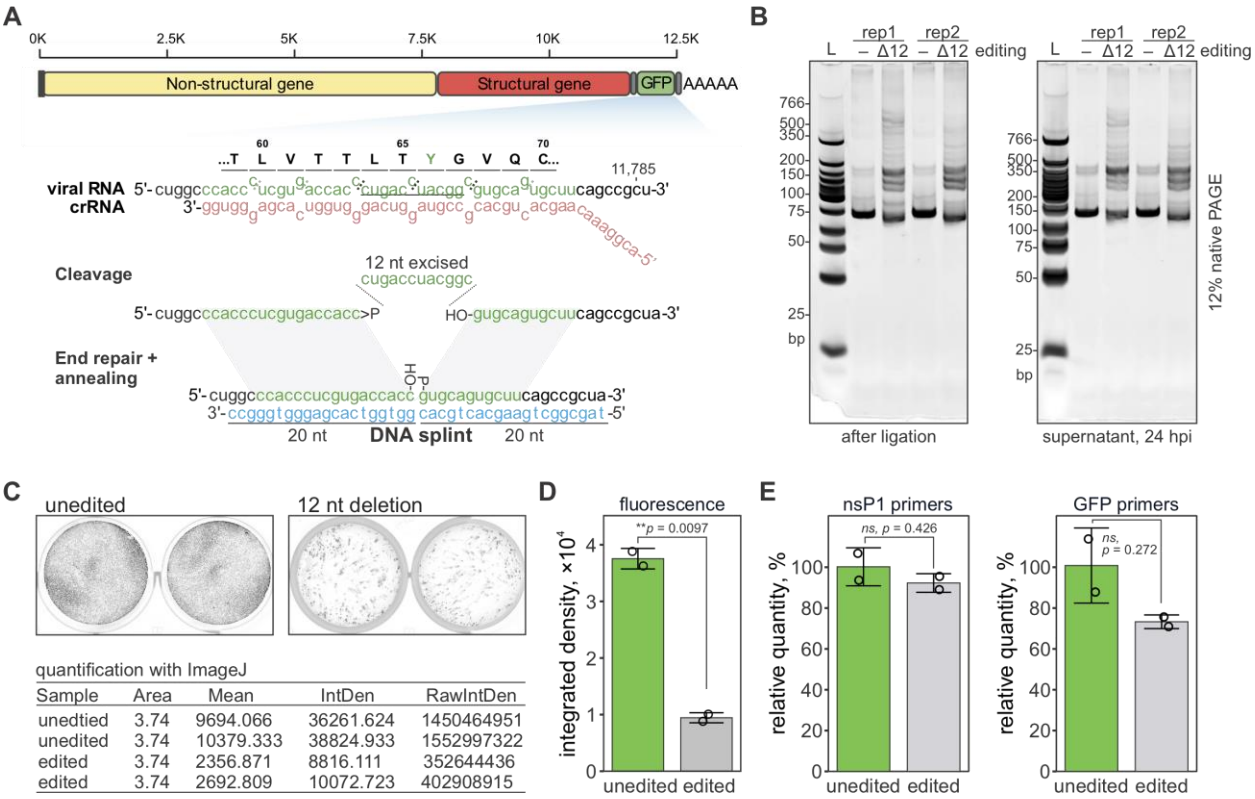


Fig. 4.S4. CRISPR-based programmed deletions in SINV-GFP. (A) Experimental design to delete 12 nucleotides in the gfp gene of SINV-GFP. The tyrosine in position 66 required for fluorescence is highlighted with green. (B) Related to main Fig. 2. Products generated in RT-qPCR shown in main Fig. 2B and E were analyzed with 12% native PAGE. Two replicates (rep1 and rep2) are shown for the unedited sample (-) and the sample with programmed deletion of 12 nucleotides (D12). L – molecular weight ladder. Diagram of the viral genome was created with BioRender (C) Related to main Fig. 2D. Fluorescence in cells transfected with unedited or edited viral RNA (top) was quantified in ImageJ software after background subtraction. Mean fluorescence, integrated density (IntDen), and raw integrated density (RawIntDen) were calculated. (D) Plot of fluorescence measurements in (C). Bars show the mean of two replicates. Error bars show the mean ± standard deviation. Samples were compared using Welch's unequal variances ttest; ** significant difference ($p < 0.01$) (E) Relative levels of RT-qPCR signal in edited vs. unedited sample generated with primers targeting nsP1 (left) or GFP (right) genes in the SINV-GFP genome. Bars show the mean of two replicates. Error bars show the mean ± standard deviation. Samples were compared using Welch's unequal variances t-test; ns – no significant difference.

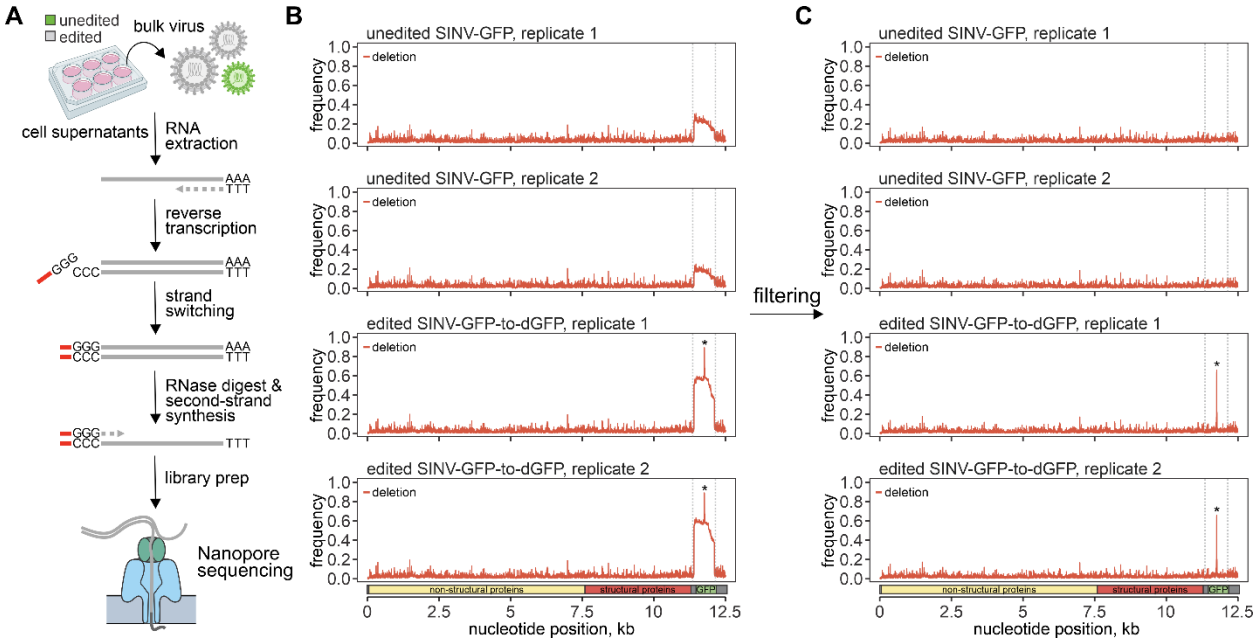


Fig. 4.S5. Direct cDNA sequencing with Nanopore. (A) Diagram of the direct cDNA sequencing protocol. RNA from cell supernatants containing bulk virus was extracted and reverse transcribed using poly-T primers. Then, the RNA copy was degraded with an RNase cocktail, and the remaining cDNA first strand was used as a template to synthesize the second DNA strand. The resulting double-stranded DNA was sequenced with Oxford Nanopore. (B) Sequencing reads were aligned to the reference, and deletion frequency was quantified across the entire viral genome. Vertical dotted lines indicate complete gfp deletion in the stock virus (unedited). Asterisk (*) marks programmed deletions at the site targeted by the SthCsm complex. (C) Sequencing data in (B) was filtered to remove reads with spontaneous deletions of the entire gfp gene.

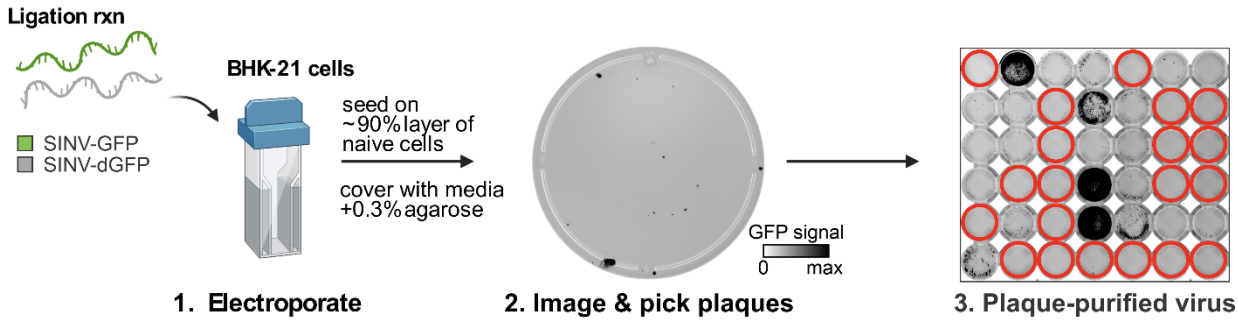


Fig. 4.S6. Plaque purification of the edited SINV-dGFP virus. 1) BHK-21 cells were transfected with cleaved, end-repaired, splinted, and ligated RNA. Transfected cells (~10,000 cells) were seeded on a ~90% layer of naïve cells in a 100 mm Petri dish. After 1 h incubation, media was changed for a fresh media supplemented with 0.3% agarose. 2) The next day (16 hpi) Petri dish was imaged with Amersham Typhoon 5 scanner to confirm infection. Then, ~20 hpi viral clones were picked by pushing sterile pipette tip through the agarose overlay. Collected agarose was then used to inoculate naïve BHK-21 cells (~90% confluency) in a 48-well plate. 3) After 24 h incubation, the 48-well plate was imaged with Amersham Typhoon 5 scanner to identify wells without GFP signal but with visible cytopathic effect (red circles). Diagrams of the ligated RNA and electroporation cuvette were created with BioRender.

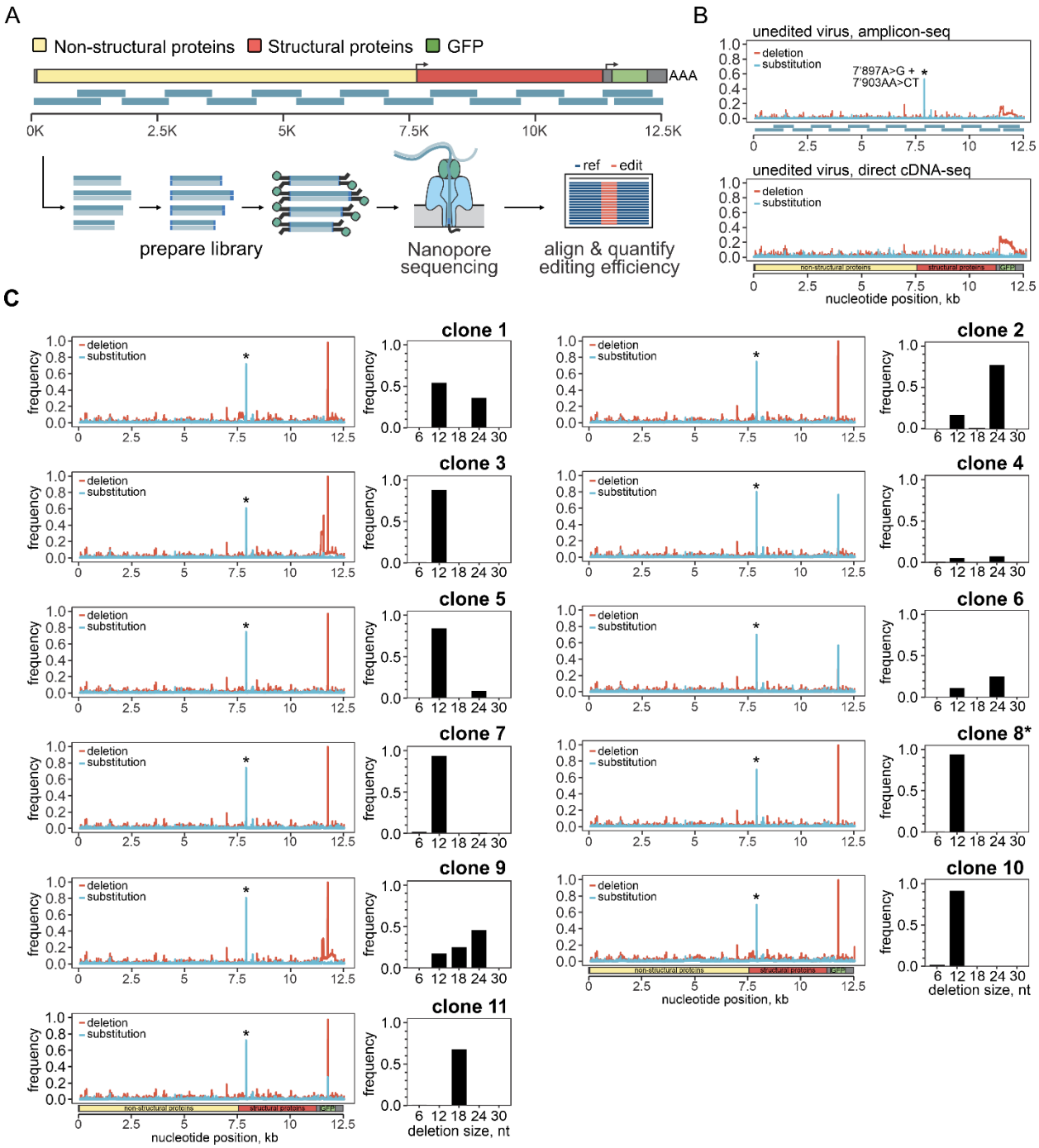


Fig. 4.S7. Amplicon sequencing of plaque-purified viruses to confirm programmed deletions in the viral genome. (A) Diagram of the amplicon sequencing strategy used to sequence genomes of plaque-purified SINV. Viral RNA was extracted from supernatants of the infected cell and reverse transcribed. A multiplex PCR was used to amplify ~1 kbp overlapping amplicons that span 97.8% of the viral genome (horizontal teal bars). Resulting amplicons were end-repaired, barcoded, and sequenced with Oxford Nanopore. (B) Unedited bulk virus was sequenced using two sequencing strategies. Sequence variants were identified by comparison to the reference

genome. Red shows the frequency of deletions, and blue – nucleotide substitutions. Asterisk (*) indicates 7,897A>G and 7,903AA>CT substitutions that 1) always coincide in the same sequencing reads (~60%), and 2) are not seen in direct cDNA sequencing of the same viral sample (bottom graph). Therefore, we conclude that this substitution is an artifact associated with amplicon sequencing rather than a bona fide genomic mutation. (C) Amplicon-based genome sequencing of 11 plaque-purified viruses (see Fig. 4.S6). Sequencing reads were compared to a reference, and all sequence variants were identified and quantified. Red indicates the frequency of deletions, and blue shows nucleotide substitutions. Asterisk (*) marks sequencing artifacts identified in (B) and presumably associated with PCR amplification. Peaks of deletion frequency in the gfp gene indicate programmed deletions introduced with CRISPR-based RNA editing (main Fig. 4.2 and Fig. 4.S4). Bar graphs show the distribution of deletion sizes at the region targeted by SthCsm complex. Sequencing data for clone 8 is shown in the main Fig 4.3C

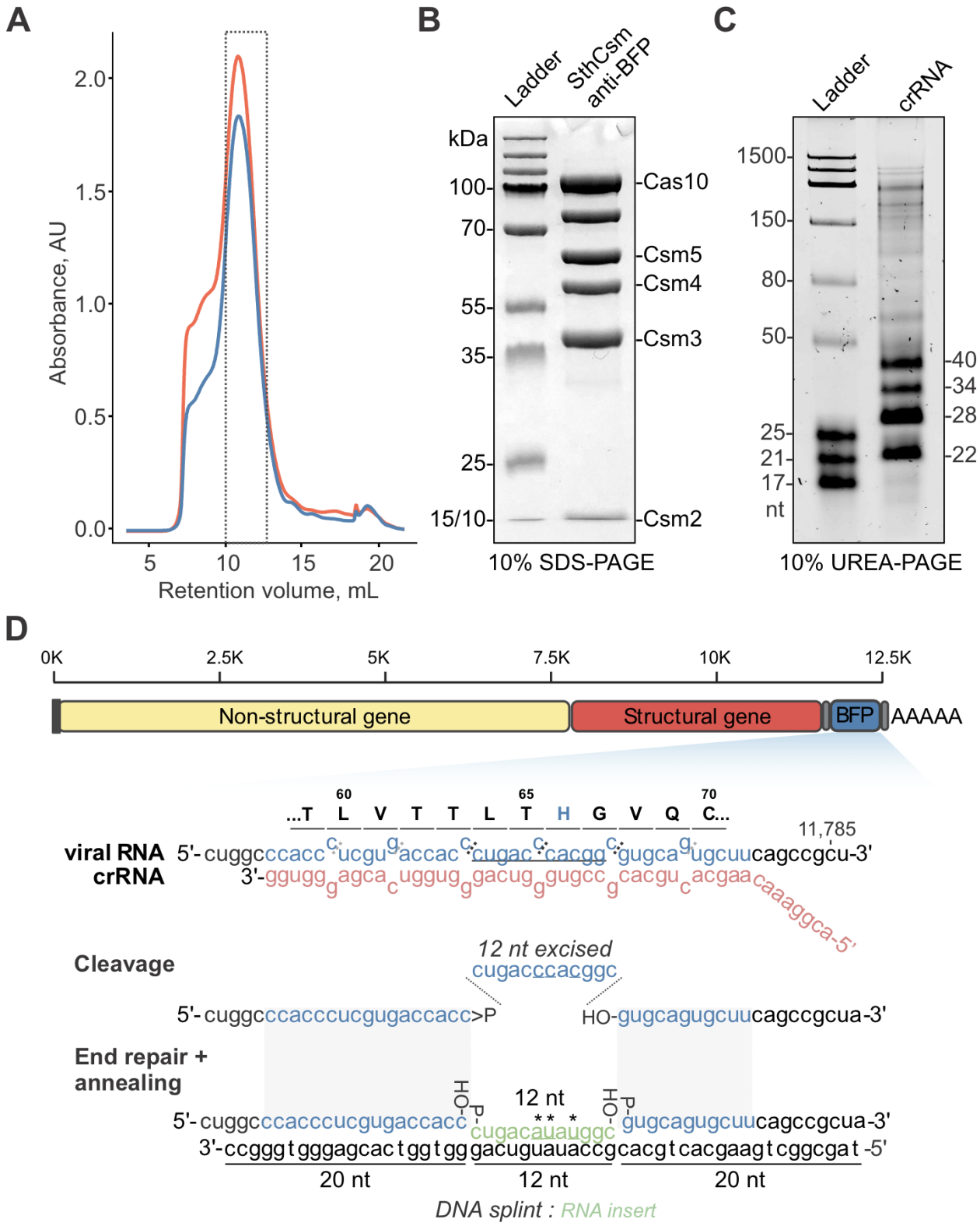

A
Absorbance, AU
Retention volume, mL
B
Ladder
SthCsm anti-BFP
kDa
Cas10
Csm5
Csm4
Csm3
Csm2
10% SDS-PAGE
C
Ladder
crRNA
nt
10% UREA-PAGE
D
0K
2.5K
5K
7.5K
10K
12.5K
Non-structural gene
Structural gene
BFP
AAAAA
viral RNA
crRNA
11,785
Cleavage
12 nt excised
End repair + annealing
12 nt
20 nt
20 nt
DNA splint : RNA insert

Fig. 4.S8. Targeting the bfp gene in SINV-BFP with SthCsm (A) Size-exclusion chromatography (SEC) of affinity purified wild-type SthCsm complex targeting bfp gene in SINV-BFP. The dotted box shows collected fractions that were pooled
together and used for SINV-BFP editing.
(B) SDS-PAGE of purified SthCsm complex shown in (A) (~10 ug/lane).
(C) RNA was extracted from SthCsm complex purification shown in (A) and separated in 10% Urea-PAGE.
(D) Experimental design to substitute 12 nucleotides in the bfp gene of SINV-BFP, converting it to GFP. The histidine in position 66 required for fluorescence is highlighted in blue. Asterisk (*) indicates nucleotide substitutions that are introduced into the viral genome through a synthetic RNA insert. The editing replaces the CAC codon with UAU, changing the encoded histidine to tyrosine (H66Y mutation).

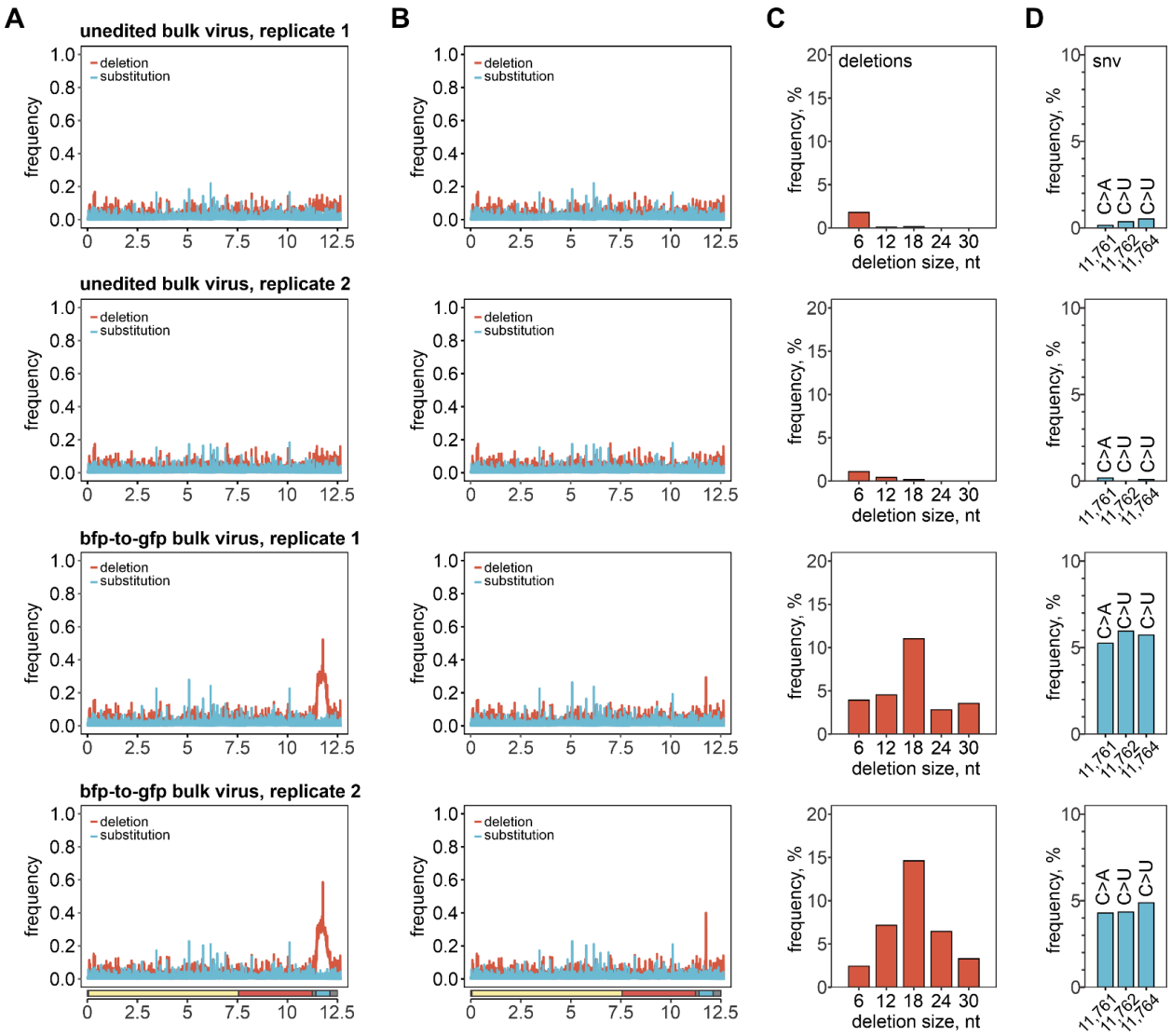


Fig. 4.S9. Genome sequencing of bulk SINV-BFP-to-GFP viruses (A) Direct cDNA sequencing of bulk SINV-BFP-to-GFP edited or unedited viruses. Sequence variants in sequencing reads were identified, quantified, and normalized to sequencing depth. Red shows deletions, blue – nucleotide substitutions. (B) Data shown in (A) was filtered to remove reads with complete deletions of the bfp gene that are both seen in edited and unedited viruses and accumulate independently of CRISPR targeting during viral passaging. (C) Distribution of deletion sizes at the region targeted by SthCsm. (D) Frequency of programmed substitutions (11,761C>A + 11,762C>U + 11,764C>U) at the target site in the bfp gene. Substitution 11,762C>U changes histidine in position 66 of the protein to tyrosine (H66Y mutation), while 11,761C>A and 11,764C>U are silent mutations.

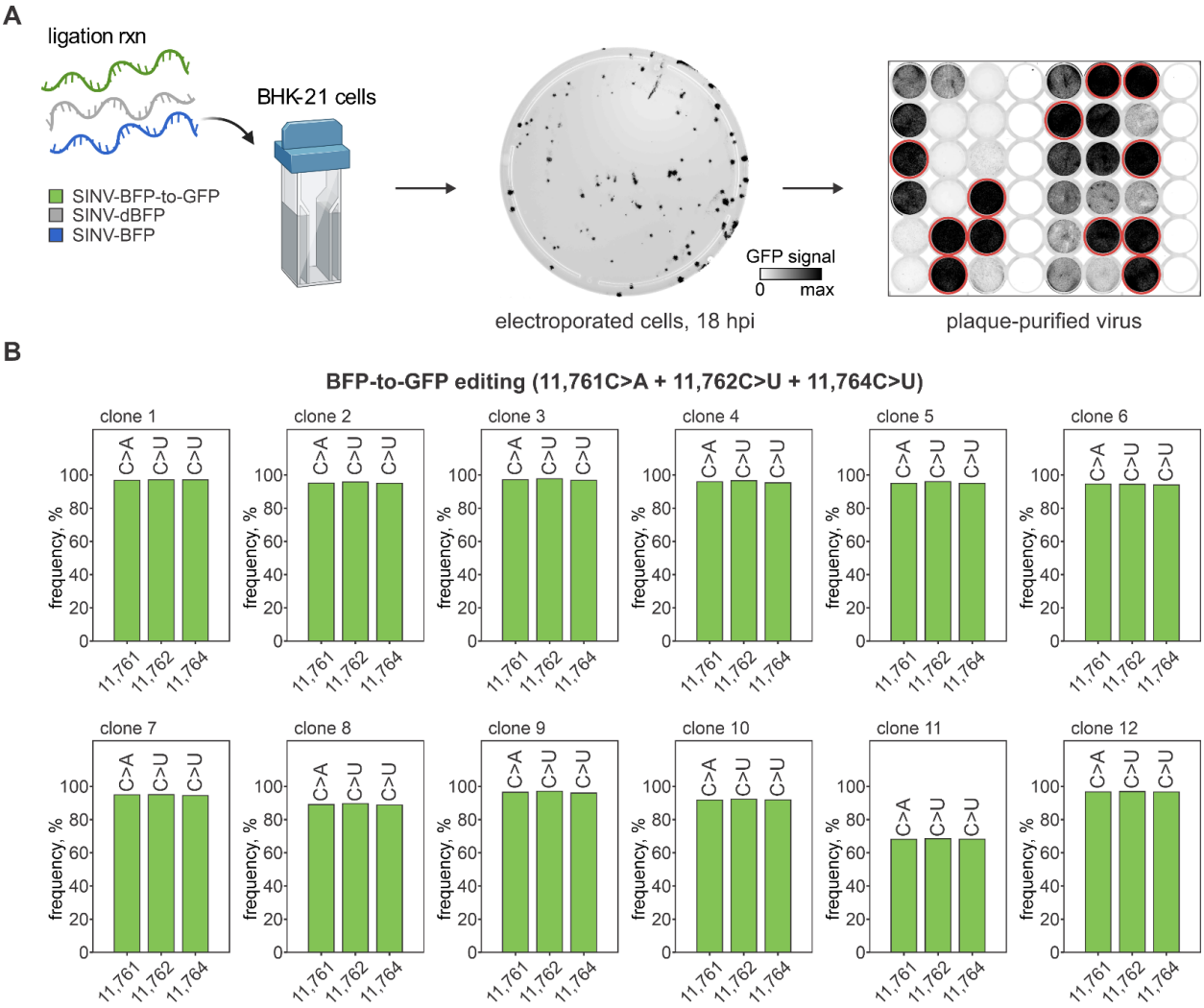


Fig. 4.S10. Genotyping of plaque-purified SINV-BFP-to-GFP viruses (A) Plaque purification of SINV-BFP-to-GFP edited virus. Transfected cells (~10,000 cells) were seeded on ~90% confluency layer of BHK-21 cells in a 100 mm Petri dish. After 1 h incubation (37°C, 5% CO_2), media was changed for fresh media supplemented with 0.3% agarose. Cells were imaged at 18 hpi. Viral plaques were picked under a standard light microscope and used to inoculate naïve cells in a 48-well plate. After imaging the 48-well plate, 12 wells producing high GFP signal (circled in red) were selected for genome sequencing. Diagrams of the ligated RNA and electroporation cuvette were created with BioRender (B) Frequency of programmed substitutions (11,761C>A + 11,762C>U + 11,764C>U) in the plaque-purified viruses. Substitution 11,762C>U changes histidine in position 66 of the protein to tyrosine (H66Y)

Table 4.S1. Oligonucleotides used for cloning anti-GFP and anti-BGP CRISPR loci

Name	Sequence 5`-3`
Sth_RSR_GFP_F	TAATCCATGGGATATAAACCTAATTACCTCGAGAGGGGACG GAAACAAGCACTGCACGCCGTAGGTCAGGGTGGTCACG
Sth_RSR_GFP_R	ATTACAATTGCCCGGGGTTTCCGTCCCCTCTCGAGGTAATTA GGTTTATATCCCACCCTCGTGACCACCCTGACCTACG
SDM_BFP_crRNA_F	CAGGGTGGTCACGAGGGTGG
SDM_BFP_crRNA_R	ACCCACGGCGTGCAGTGCTTG

Table 4.S2. DNA splints and RNA inserts used for RNA editing.

Name	Sequence 5`-3`
Programmed deletions in IVT RNA	
Splint A, 18 nt	TCCGAAGAACGCTGAAGCCAATGTTTGTAATCAGTTCC
Splint B, 18 nt	GAAGAACGCTGAAGCGATTTGCGGCCAATGTTT
Splint B, 24 nt	GAAGAACGCTGAAGCGGGCCAATGTTTGTAATCAGT
Programmed deletion in SINV-GFP	
Y66_d12_splint	TAGCGGCTGAAGCACTGCACGGTGGTCAC GGGGTGGGCC
Programmed substitution in SINV-BFP	
H66Y_splint	TAGCGGCTGAAGCACTGCACGCCATATGTCAGGGT GGTCAC GAGGGTGGGCC
H66Y_insert	/5Phos/CUGACAUAUGGC

Table 4.S3. Oligonucleotide primers used to generate DNA templates for in vitro transcription of RNA.

Name	Sequence 5`-3`
N0.5_F	TAATACGACTCACTATAGGGCGTGTTGTTTTAGATTTCATC
N3-5_R	CACACTGATTAAAGATTGCTATGTG
T7_SINV_sgmRNA_F	TAATACGACTCACTATAGTACATTTCATCTGACTAATACTA C
SINV_polyA_R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAATGTTAAAA ACAAAATTTTGTTG

Table 4.S4. Oligonucleotide primers used for PCR and qPCR.

Name	Sequence 5`-3`
11762T>C_F1	TAATACGACTCACTATAGTACATTTCATCTGACTAATACTAC
11762T>C_R1	GCACGCCGTGGGTCAGGGTGGTCACG
11762T>C_F2	GACCACCCTGACCCACGGCGTGCAGTGC
11762T>C_R2	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAATGTTAAAAA CAAAATTTTGTTG
CoV2_N_qPCR_F	GGGACCAGGAACTAATCAGAC
CoV2_N_qPCR _R	TGTGACTTCCATGCCAATG
qSindbisFW4495 (*41*)	TAGACAGAACTGACGCGGACGT
qSindbisREV4635 (*41*)	TCCATACTAACTCATCGTCGATCTC
qSINV_GFP_F1	GAGGGCGAGGGCGATG
qSINV_GFP_R1	GCTTCATGTGGTCGGGGTAG
nCoV-96_L	GCCAACAACAACAAGGCCAAAC
nCoV-96_R	TAGGCTCTGTTGGTGGGAATGT

Table 4.S5. PCR primers used for amplicon sequencing of SINV-GFP and SINV-BFP genomes.

Name	Sequence 5`-3`	Pool
sinv_1K_1v2_L	ACAGCCGACCAATTGCACTA	1
sinv_1K_1v2_R	AGAGGCTGGGACTTTTACGC	1
sinv_1K_3_L	CTGAAGAATGCCAAACTCGCAC	1
sinv_1K_3_R	ATATCCCCTGGCTTCGGCTT	1
sinv_1K_5_L	CTTGATTTGCAGACGGGGAGAA	1
sinv_1K_5_R	GCGGTTGTCAAGCAGTTAAGTG	1
sinv_1K_7_L	GCTTTAGCGGATCGGACAACT	1
sinv_1K_7_R	ACTGTCCCGTCTACCATATCCAA	1
sinv_1K_9_L	CGAGTGCGCCTTTGGAGAAATA	1
sinv_1K_9_R	TTGACGTCGAACAATCTGTCGG	1
sinv_1K_11v3_L	CTGCCACCATACTGAACCGT	1
sinv_1K_11v3_R	ATGGTGTACACAGGATGGCG	1
sinv_1K_13_L	GCGATGAAAGTAGGACTGCGTA	1
sinv_1K_13_R	GGCTTGGAATGCTCTTTTGCTC	1
sinv_1K_15_L	CACCTCTAGAATCGCCACCATG	1
sinv_1K_15_R	GCACCACGCTTCCTCAGAAATA	1
sinv_1K_2_L	TGTGATACAGTGGTGAGTTGCG	2
sinv_1K_2_R	GGCATCAGTACTTTAGCGTCGT	2
sinv_1K_4_L	GCACACAGCCAGTTACAGCTAT	2
sinv_1K_4_R	ACTGAGTGGTGTTTGAACTGGTT	2
sinv_1K_6_L	ACAAAACGCCTACCATGCAGT	2
sinv_1K_6_R	CTTTAGCCTTGGCGGTGGAATA	2
sinv_1K_8_L	GAACTACTCCGATCCACAGTTCG	2
sinv_1K_8_R	AGTACTCTGCTGGCGATAACGA	2
sinv_1K_10_L	CCGAAGAAACCAAAAACGCAGG	2
sinv_1K_10_R	TCCATGGTGCCTTCTTTAACGG	2
sinv_1K_12_L	TGGCCTGGAATACATATGGGGA	2
sinv_1K_12_R	ACGACCTTATGATCGAATGGCG	2
sinv_1K_14_L	CGCCTCGTCGCTATTAATTATAGGA	2
sinv_1K_14_R	TACATCGAGTTTTGCTGGTCGG	2

Manuscript References

1. D. A. Jackson, R. H. Symons, P. Berg, Biochemical method for inserting new genetic information into DNA of simian virus 40: Circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 69, 2904–2909 (1972).

2. P. Berg, M. F. Singer, The recombinant DNA controversy: Twenty years later. Proc. Natl. Acad. Sci. U.S.A. 92, 9011–9013 (1995).

3. P. Rouet, F. Smih, M. Jasin, Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. 14, 8096–8106 (1994).

4. B. L. Stoddard, Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. Structure 19, 7–15 (2011).

5. M. Bibikova, D. Carroll, D. J. Segal, J. K. Trautman, J. Smith, Y. G. Kim, S. Chandrasegaran,
Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol. Cell. Biol. 21, 289–297 (2001).

6. M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove,
D. F. Voytas, Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757–761 (2010).

7. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable
dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821 (2012).

8. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex
mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U.S.A. 109, E2579–E2586 (2012).

9. O. O. Abudayyeh, J. S. Gootenberg, P. Essletzbichler, S. Han, J. Joung, J. J. Belanto, V. Verdine, D. B. T. Cox, M. J. Kellner, A. Regev, E. S. Lander, D. F. Voytas, A. Y. Ting, F. Zhang, RNA targeting with CRISPR-Cas13. Nature 550, 280–284 (2017).

10. O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D. B. T. Cox, S. Shmakov, K. S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E. S.

Lander, E. V. Koonin, F. Zhang, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353, aaf5573 (2016).

11. A. J. Meeske, S. Nakandakari-Higa, L. A. Marraffini, Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature 570, 241–245 (2019).

12. S. Konermann, P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev, P. D. Hsu, Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell 173, 665–676.e14 (2018).

13. D. B. T. Cox, J. S. Gootenberg, O. O. Abudayyeh, B. Franklin, M. J. Kellner, J. Joung, F. Zhang, RNA editing with CRISPR-Cas13. Science 358, 1019–1027 (2017).

14. O. O. Abudayyeh, J. S. Gootenberg, B. Franklin, J. Koob, M. J. Kellner, A. Ladha, J. Joung,
P. Kirchgatterer, D. B. T. Cox, F. Zhang, A cytosine deaminase for programmable single-base RNA editing. Science 365, 382–386 (2019).

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

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

17. D. Colognori, M. Trinidad, J. A. Doudna, Precise transcript targeting by CRISPR-Csm complexes. Nat. Biotechnol., (2023).

18. W. T. Woodside, N. Vantsev, R. J. Catchpole, S. C. Garrett, S. Olson, B. R. Graveley, M. P. Terns, Type III-A CRISPR systems as a versatile gene knockdown technology. RNA 28, 1074–1088 (2022).

19. A. Ozcan, R. Krajeski, E. Ioannidi, B. Lee, A. Gardner, K. S. Makarova, E. V. Koonin, O. O. Abudayyeh, J. S. Gootenberg, Programmable RNA targeting with the single-protein CRISPR effector Cas7-11. Nature 597, 720–725 (2021).

20. T. Fricke, D. Smalakyte, M. Lapinski, A. Pateria, C. Weige, M. Pastor, A. Kolano, C. Winata, V. Siksnys, G. Tamulaitis, M. Bochtler, Targeted RNA knockdown by a type III CRISPR-Cas complex in zebrafish. CRISPR J. 3, 299–313 (2020).

21. N. Tanaka, S. Shuman, RtcB is the RNA ligase component of an Escherichia coli RNA repair
operon. J. Biol. Chem. 286, 7727–7731 (2011).

22. V. Cameron, O. C. Uhlenbeck, 3’-Phosphatase activity in T4 polynucleotide kinase. Biochemistry 16, 5120–5126 (1977).

23. R. Silber, V. G. Malathi, J. Hurwitz, Purification and properties of bacteriophage T4-induced
RNA ligase. Proc. Natl. Acad. Sci. U.S.A. 69, 3009–3013 (1972).

24. M. R. Stark, S. D. Rader, Efficient splinted ligation of synthetic RNA using RNA ligase. Methods Mol. Biol. 1126, 137–149 (2014).

25. V. R. Racaniello, D. Baltimore, Cloned poliovirus complementary DNA is infectious in mammalian cells. Science 214, 916–919 (1981).

26. T. Taniguchi, M. Palmieri, C.Weissmann, Qβ DNA-containing hybrid plasmids giving rise to
Qβ phage formation in the bacterial host. Nature 274, 223–228 (1978).

27. B. D. Lindenbach, in Advances in Virus Research, M. Kielian, T. C. Mettenleiter, M. J. Roossinck, Eds. (Academic Press, 2022), vol. 122, pp. 1–29.

28. C. Stobart, M. Moore, RNA virus reverse genetics and vaccine design. Viruses 6, 2531–2550 (2014).

29. T. T. N. Thao, F. Labroussaa, N. Ebert, P. V’kovski, H. Stalder, J. Portmann, J. Kelly, S. Steiner, M. Holwerda, A. Kratzel, M. Gultom, K. Schmied, L. Laloli, L. Hüsser, M. Wider, S. Pfaender, D. Hirt, V. Cippà, S. Crespo-Pomar, S. Schröder, D. Muth, D. Niemeyer, V. M. Corman, M. A. Müller, C. Drosten, R. Dijkman, J. Jores, V. Thiel, Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. Nature 582, 561–565 (2020).

30. X. Xie, A. Muruato, K. G. Lokugamage, K. Narayanan, X. Zhang, J. Zou, J. Liu, C. Schindewolf, N. E. Bopp, P. V. Aguilar, K. S. Plante, S. C. Weaver, S. Makino, J. W. LeDuc, V. D. Menachery, P. Y. Shi, An infectious cDNA clone of SARS-CoV-2. Cell Host Microbe 27, 841–848.e3 (2020).

31. Y. J. Hou, K. Okuda, C. E. Edwards, D. R. Martinez, T. Asakura, K. H. Dinnon III, T. Kato, R. E. Lee, B. L. Yount, T. M. Mascenik, G. Chen, K. N. Olivier, A. Ghio, L. V. Tse, S. R. Leist, L. E. Gralinski, A. Schäfer, H. Dang, R. Gilmore, S. Nakano, L. Sun, M. L. Fulcher, A. Livraghi-Butrico, N. I. Nicely, M. Cameron, C. Cameron, D. J. Kelvin, A. de Silva, D. M. Margolis, A. Markmann, L. Bartelt, R. Zumwalt, F. J. Martinez, S. P. Salvatore, A. Borczuk, P. R. Tata, V. Sontake, A. Kimple, I. Jaspers, W. K. O’Neal, S. H. Randell, R. C. Boucher, R. S. Baric, SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. Cell 182,
429–446.e14 (2020).

32. X. He, S. Quan, M. Xu, S. Rodriguez, S. L. Goh, J.Wei, A. Fridman, K. A. Koeplinger, S. S. Carroll, J. A. Grobler, A. S. Espeseth, D. B. Olsen, D. J. Hazuda, D. Wang, Generation of SARS-CoV-2 reporter replicon for high-throughput antiviral screening and testing. Proc. Natl. Acad. Sci. U.S.A. 118, e2025866118 (2021).

33. J. A. Plante, Y. Liu, J. Liu, H. Xia, B. A. Johnson, K. G. Lokugamage, X. Zhang, A. E. Muruato, J. Zou, C. R. Fontes-Garfias, D. Mirchandani, D. Scharton, J. P. Bilello, Z. Ku, Z. An, B. Kalveram, A. N. Freiberg, V. D. Menachery, X. Xie, K. S. Plante, S. C.Weaver, P.-Y. Shi, Spike mutation D614G alters SARS-CoV-2 fitness. Nature 592, 116–121 (2021).

34. D.-Y. Chen, C. V. Chin, D. Kenney, A. H. Tavares, N. Khan, H. L. Conway, G. Q. Liu, M. C. Choudhary, H. P. Gertje, A. K. O'Connell, S. Adams, D. N. Kotton, A. Herrmann, A. Ensser, J. H. Connor, M. Bosmann, J. Z. Li, M. U. Gack, S. C. Baker, R. N. Kirchdoerfer, Y. Kataria, N. A. Crossland, F. Douam, M. Saeed, Spike and nsp6 are key determinants of SARS-CoV-2 Omicron BA.1 attenuation. Nature 615, 143–150 (2023).

35. T. Y. Taha, I. P. Chen, J. M. Hayashi, T. Tabata, K.Walcott, G. R. Kimmerly, A. M. Syed, A. Ciling, R. K. Suryawanshi, H. S. Martin, B. H. Bach, C.-L. Tsou, M. Montano, M. M. Khalid, B. K. Sreekumar, G. R. Kumar, S. Wyman, J. A. Doudna, M. Ott, Rapid assembly of SARS-CoV-2 genomes reveals attenuation of the Omicron BA.1 variant through NSP6. Nat. Commun. 14, 2308 (2023).

36. C. Kurhade, X. Xie, P.-Y. Shi, Reverse genetic systems of SARS-CoV-2 for antiviral research. Antiviral Res. 210, 105486 (2023).
37. S. P. B. van Beljouw, A. C. Haagsma, A. Rodríguez-Molina, D. F. van den Berg, J. N. A. Vink, S. J. J. Brouns, The gRAMP CRISPR-Cas effector is an RNA endonuclease complexed with a caspase-like peptidase. Science 373, 1349–1353 (2021).

38. K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Č. Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, E. V. Koonin, Evolutionary classification of CRISPRCas systems: A burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67–83 (2019).

39. T. Wiegand, R. Wilkinson, A. Santiago-Frangos, M. Lynes, R. Hatzenpichler, B. Wiedenheft,Functional and phylogenetic diversity of Cas10 proteins. CRISPR J. 6, 152–314 (2023).

40. C. S. Hahn, Y. S. Hahn, T. J. Braciale, C. M. Rice, Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. Proc. Natl. Acad. Sci. U.S.A. 89, 2679–2683 (1992).

41. M. L. Flenniken, R. Andino, Non-specific dsRNA-mediated antiviral response in the honey bee. PLOS ONE 8, 10 (2013).

42. J. Quick, N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara Jr., E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen N. J. Loman, Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. 12, 1261–1276 (2017).

43. K. Sahlin, P. Medvedev, Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. Nat. Commun. 12, 2 (2021).

44. D. Kim, J. Y. Lee, J. S. Yang, J. W. Kim, V. N. Kim, H. Chang, The architecture of SARS-CoV-2 Transcriptome. Cell 181, 914–921.e10 (2020).

CHAPTER FIVE

VIRAL PROTEINS ACTIVATE PARIS-MEDIATED TRNA DEGRADATION AND VIRAL TRNAS RESCUE INFECTION

Contribution of Authors and Co-Authors

Manuscript in Chapter 5

Author: Nathaniel Burman

Contributions: All authors contributed significantly to this work

Co-Author: Svetlana Belukhina

Contributions: All authors contributed significantly to this work

Co-Author: Florence Deoardieu

Contributions: All authors contributed significantly to this work

Co-Author: Royce A. Wilkinson

Contributions: All authors contributed significantly to this work

Co-Author: Mikhail Skutel

Contributions:

Co-Author: Andrew Santiago-Frangos

Contributions: All authors contributed significantly to this work

Co-Author: Ava B. Graham

Contributions: All authors contributed significantly to this work

Co-Author: Alexei Livinskyi

Contributions: All authors contributed significantly to this work

Co-Author: Anna Chechenina

Contributions: All authors contributed significantly to this work

Co-Author: Natalia Morozova

Contributions: All authors contributed significantly to this work

Co-Author: Trevor Zahl

Contributions: All authors contributed significantly to this work

Co-Author: William S Hendriques

Contributions: All authors contributed significantly to this work

Co-Author: Murat Buyukyoruk

Contributions: All authors contributed significantly to this work

Co-Author: Chistophe Rouillon

Contributions: All authors contributed significantly to this work

Co-Author: Lena Agnés Thierry

Contributions: All authors contributed significantly to this work

Co-Author: Romain Koszul

Contributions: All authors contributed significantly to this work

Co-Author: Florian Tesson

Contributions: All authors contributed significantly to this work

Co-Author: Aude Bernheim

Contributions: All authors contributed significantly to this work

Co-Author: Divid Bikard

Contributions: All authors contributed significantly to this work

Co-Author: Blake Wiedenheft

Contributions: All authors contributed significantly to this work

Co-Author: Artem Isaev

Contributions: All authors contributed significantly to this work

Manuscript Information

Nathaniel Burman[1]*, Svetlana Belukhina[2]*, Florence Depardieu[3]*, Royce A. Wilkinson[1], Mikhail Skutel[2], Andrew Santiago-Frangos[1], Ava B. Graham[1], Alexei Livenskyi[4,5], Anna Chechenina[2], Natalia Morozova[6], Trevor Zahl[1], William S. Henriques[1], Murat Buyukyoruk[1], Christophe Rouillon[3], Lena Shyrokova[7], Tatsuaki Kurata[7], Vasili Hauryliuk[7,8,9], Konstantin Severinov[10], Justine Groseille [11,12,13], Agnès Thierry[11], Romain Koszul[11], Florian Tesson[14], Aude Bernheim[14], David Bikard[3]**, Blake Wiedenheft[1]**, Artem Isaev[2]**

[1]Montana State University, Bozeman, Department of Microbiology and Cell Biology, 59715 Montana, USA

[2]Skolkovo Institute of Science and Technology, Center for Molecular and Cellular Biology, Moscow, Russia

[3]Institut Pasteur, Université Paris Cité, CNRS UMR 6047, Synthetic Biology, 75015 Paris, France

[4]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow 119334, Russia

[5]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia

[6]Peter the Great St Petersburg State Polytechnic University, St. Petersburg, Russia

[7]Department of Experimental Medical Science, Lund University, 221 00 Lund, Sweden

[8]Virus Centre, Lund University, Lund, Sweden

[9]Science for Life Laboratory, Lund, Sweden

[10]Waksman Institute of Microbiology, Piscataway, USA

[11]Institut Pasteur, CNRS UMR 3525, Université Paris Cité, Unité Régulation Spatiale des Génomes, 75015 Paris, France

[12]Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS, INSERM, Université PSL, Paris, France

[13]Sorbonne Université, College Doctoral

[14]Institut Pasteur, Université Paris Cité, Molecular Diversity of Microbes, 75015 Paris, France

* Equal contribution

** Corresponding Author: D.B. (david.bikard@pasteur.fr), B.W. (bwiedenheft@gmail.com), and A.I. (Artem.Isaev@skoltech.ru)

___ Prepared for submission to a peer-reviewed journal

_X_ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

___Published in a peer-reviewed journal

Abstract

Viruses compete with each other for limited cellular resources, and some viruses deliver defense mechanisms that protect the host from competing genetic parasites. PARIS is a defense system, often encoded in viral genomes, that is composed of a 53 kDa ABC ATPase (AriA) and a 35 kDa TOPRIM nuclease (AriB). Here we show that AriA and AriB assemble into a 425 kDa supramolecular immune complex. We use cryo-EM to determine the structure of this complex which explains how six molecules of AriA assemble into a propeller-shaped scaffold that coordinates three subunits of AriB. ATP-dependent detection of foreign proteins triggers the release of AriB, which assembles into a homodimeric nuclease that blocks infection by cleaving the host tRNALys. Phage T5 subverts PARIS immunity through expression of a tRNALys variant that prevents PARIS-mediated cleavage, and thereby restores viral infection. Collectively, these data explain how AriA functions as an ATP-dependent sensor that detects viral proteins and activates the AriB toxin. PARIS is one of an emerging set of immune systems that form macromolecular complexes for the recognition of foreign proteins, rather than foreign nucleic acids.

Introduction

Anti-viral defense systems in bacteria and archaea are extraordinarily diverse and many of these systems are mechanistically similar to immune responses in eukaryotic cells1. The recent expansion of bacterial and archaeal immune systems stems from the appreciation that defense systems tend to co-localize in the genome2 and that most systems are carried by mobile genetic elements, including prophages and satellite viruses3. The Phage Anti-Restriction Induced System (PARIS) is one of several defense systems recently identified within a hotspot of genetic diversity carried by P4 phage satellites4. The PARIS system of *Escherichia coli* strain B185 protects against phage T7 infection through a mechanism triggered by sensing the Ocr (overcoming classical restriction) protein of T7. Ocr is a DNA mimic that inhibits Type I Restriction-Modification and BREX defense systems5,6. Thus, PARIS is an 'anti-anti-restriction' system that senses Ocr, a viral immune suppressor.

The PARIS defense system consists of two proteins, AriA and AriB (**Fig 1A,B**). AriA contains a conserved ABC ATPase domain, while AriB is a domain of unknown function (DUF4435), which includes a TOPRIM nuclease7,8. AriA and AriB are usually separate proteins, though they are occasionally fused into a single polypeptide. Several defense systems share an ABC ATPase domain and a TOPRIM domain, including the OLD (Overcoming Lysogenization Defect) protein produced by the P2 prophage, which blocks infection by phage lambda9. OLD does not directly sense a phage protein but rather the inactivation of the RecBCD exonuclease10. Activation of OLD results in translation inhibition11, but the underlying mechanism has not been determined.

Association between ABC ATPase and TOPRIM domains is an emerging theme in recently discovered immune systems12. Examples include the Gabija system13,14, effector toxin proteins of retron immunity15, and *bona fide* toxin-antitoxin (TA) modules16. This implies a broad adaptability of the conserved ABC ATPase and TOPRIM nuclease architecture to a variety of antigenic stimuli, resulting in the diversification of the immune modules likely acting through "abortive" mechanisms. We currently have little understanding of the molecular mechanisms underpinning how these systems sense and respond to viral infections.

Here we use cryo-EM to show that *E. coli* B185 PARIS systems assembles into a 425 kDa propeller-shaped complex with six molecules of AriA and three subunits of AriB. We show that AriA senses the T7 Ocr anti-restriction protein leading to the release of the AriB nuclease effector. Release of AriB from the AriA scaffold is necessary to potentiate AriB, which forms a nuclease-active dimer that degrades host lysine tRNA (tRNALys), thereby blocking translation and leading to growth arrest or cell death. We find that phage T5 carries its own tRNALys isoacceptor, which rescue the phage from PARIS-mediated defense. Finally, we perform a phylogenetic analysis of PARIS systems, which provides insights into their evolutionary history and shows how they relate to other defense systems that employ ABC ATPases.

Results

PARIS Assembles into a Propeller Shaped Macromolecular Complex

The PARIS defense system consists of two proteins, AriA and AriB, which protect cells from phages through a previously undetermined mechanism4. Here, we focused on the system from the P4 satellite integrated into the *E. coli* (str. B185) genome (**Fig 5.1A,B**). Based on

published experimental structures of related proteins8,13 and AlphaFold217 predictions (**Fig 4.1C, Sup Fig 5.1A-D**), we anticipated that AriA:AriB would assemble into a multi-subunit immune complex. To determine if and how AriA and AriB assemble, we overexpressed AriA with a C-terminally tagged AriB (AriB-Strep), pulled down on AriB, and purified the complex using size exclusion chromatography (**Fig 4.1D**). The AriA and AriB proteins co-elute in a single peak with an estimated molecular weight of 450 kDa. Next, we used cryo-EM to determine structures of the PARIS complex (**Fig 5.1E**). The structure explains how six molecules of AriA come together to form a D3 symmetric scaffold decorated by three subunits of AriB. The helical domains, which typically provide a platform for DNA binding in related ABC ATPases18, instead function as additional dimerization interfaces that enable a trimeric assembly of AriA homodimers (**Fig 5.1C, Sup Fig 5.2**). Dimerization of the helical domains from adjacent AriA homodimers form three blades of the propeller (**Fig 5.1E, Sup Fig 5.2)**. Additional density is evident between each of the AriA blades, where AriB attaches to AriA near the ATP binding sites.

However, density in this region of the reconstruction reveals conformational heterogeneity, consistent with AriB attaching to either, but not both AriA molecules at the dimer interface (**Sup Fig 5.1E**). Using multiple rounds of focused 3-D classification, we resolved two distinct isomers of the complex (**Fig 5.1E, Sup Fig 5.3D**). In one reconstruction, AriB attaches to three symmetrically equivalent AriA molecules (3.8 Å-resolution, *cis* configuration), while in the other reconstruction, one of the three AriB molecules binds the opposing AriA, which flips AriB 180-degrees relative to the other two AriB subunits (4.0 Å-resolution, *trans* configuration) (**Sup Fig 5.3**).

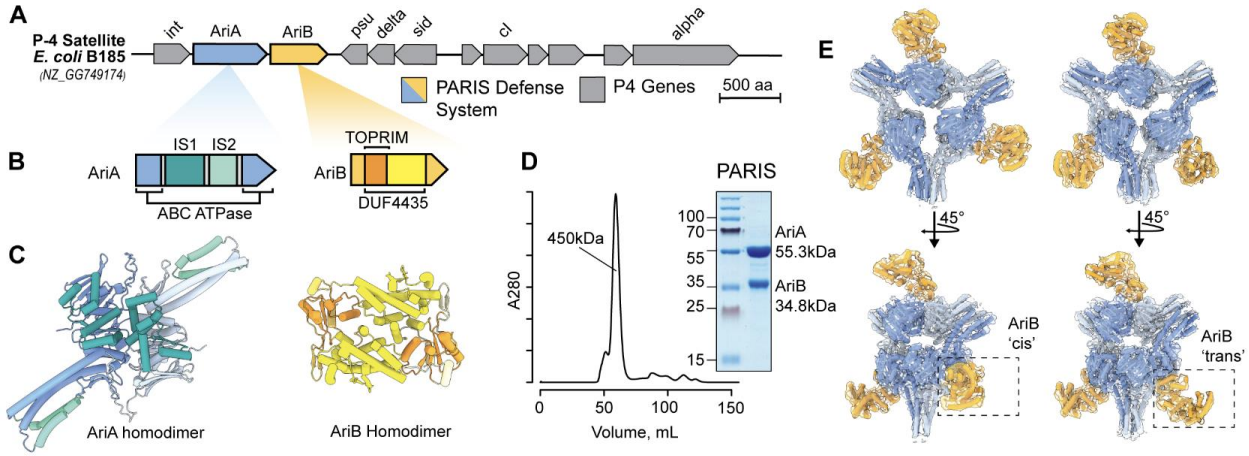


Figure 5.1: PARIS is a two-component system that assembles into a propeller-shaped supramolecular complex. A) Schematic of the integrated P4 satellite in the E. coli genome. B) PARIS genes colored according to domain organization. AriA is an ABC ATPase, while AriB a DUF4435 that includes TOPRIM nuclease. Insertion sequences 1 (IS1) is composed of three helices flanked by the ABC ATPase domain, while Insertion sequences 2 (IS2) adds two additional helices to the coiled-coil resulting in a helical bundle. C) AlphaFold2-predicted structures of AriA and AriB as homodimers colored according to structural features shown in panel B. D) Size exclusion chromatography reveals that AriA and AriB copurify as an ~450 kDa complex. E) Experimentally determined density maps of the PARIS complex reveals a homohexameric assembly of AriA with D3 symmetry. The AriA hexamer is decorated by AriB subunits that attach to AriA homodimers in one of two possible orientations, a 'cis' arrangement where the three AriB subunits are facing the same direction, or a 'trans' arrangement where one AriB subunits is rotated by 180° relative to the other two.

AriA is an ABC ATPase that Sequesters the AriB TOPRIM Nuclease

While the homohexameric assembly of AriA was apparent in the early stages of reconstruction, conformational heterogeneity limited the resolution (**Sup Fig 5.3**). To improve the resolution, we used local refinement to determine a 3.2 Å nominal resolution reconstruction of one asymmetric unit of the PARIS complex (AriA2:AriB1), into which we built an atomic model (**PDB ID: 8UX9, Sup Table 1**) (**Fig 5.2**). Each asymmetric unit of the PARIS complex is composed of two AriA molecules, which form a head to tail homodimer, and one AriB molecule that binds directly above the AriA dimer interface.

AriB is a metal-dependent TOPRIM nuclease and the predicted effector of the PARIS defense system4,7,8. The structure reveals a series of spatially conserved, negatively charged residues (E26, D30, D88, E90, and E122) that are characteristic of TOPRIM domains that bind two metal ions. The E26A mutation inactivates PARIS-mediated defense (**Fig 5.2B**)4,7. Glutamic acid 26 is just upstream of an N-terminal alpha helix (α2) on AriB, which includes two Arginine residues (R31 and R28) that form salt bridges with negatively charged residues on AriA that is positioned directly above the ATP binding sites (**Fig 5.2B, Sup Fig 5.2**) Mutations that interfere with the AriA:B interface (R31E, R28E) limit phage protection by PARIS (**Fig 5.2E**), which suggests that the protective activity of AriB requires an association with AriA.

The AriA ABC ATPase domain resembles that of Rad50 (RMSD 1.1 Å across 86 C-alpha atom pairs), a universally conserved protein involved in dsDNA break repair19–21. Structural comparison of AriA to Rad50 confirms that AriA contains the sequence motifs required for ATPase activity, along with two unique insertion sequences, IS1 and IS2 (**Fig 5.1B, Sup Fig 5.2**). Insertion sequence 2 contains a series of aromatic residues located on one face of helix α8

(F230, F234, F238, and Y304) that coordinate interactions between AriA homodimers, and explains how AriA homodimers assemble into a trimer of dimers (**Sup Fig 5.2**); a rare assembly state for this family of proteins19. Like other ABC ATPases, the head to tail assembly of the AriA homodimer results in two symmetrically positioned nucleotide binding domains (NBD) at the dimer interface (**Fig 5.2A**). In the presence of a nonhydrolysable ATP analog (i.e., ATPγS), we observe density for the ligand in both nucleotide binding domains (NBDs), but in different coordination states (**Fig 5.2C-D**). Previous studies of ABC ATPases revealed that ATP hydrolysis occurs at the two NBDs in an alternating fashion19,22, which is consistent with the observed "open" and "closed" conformations of NBDs (**Fig 5.2C-D).**

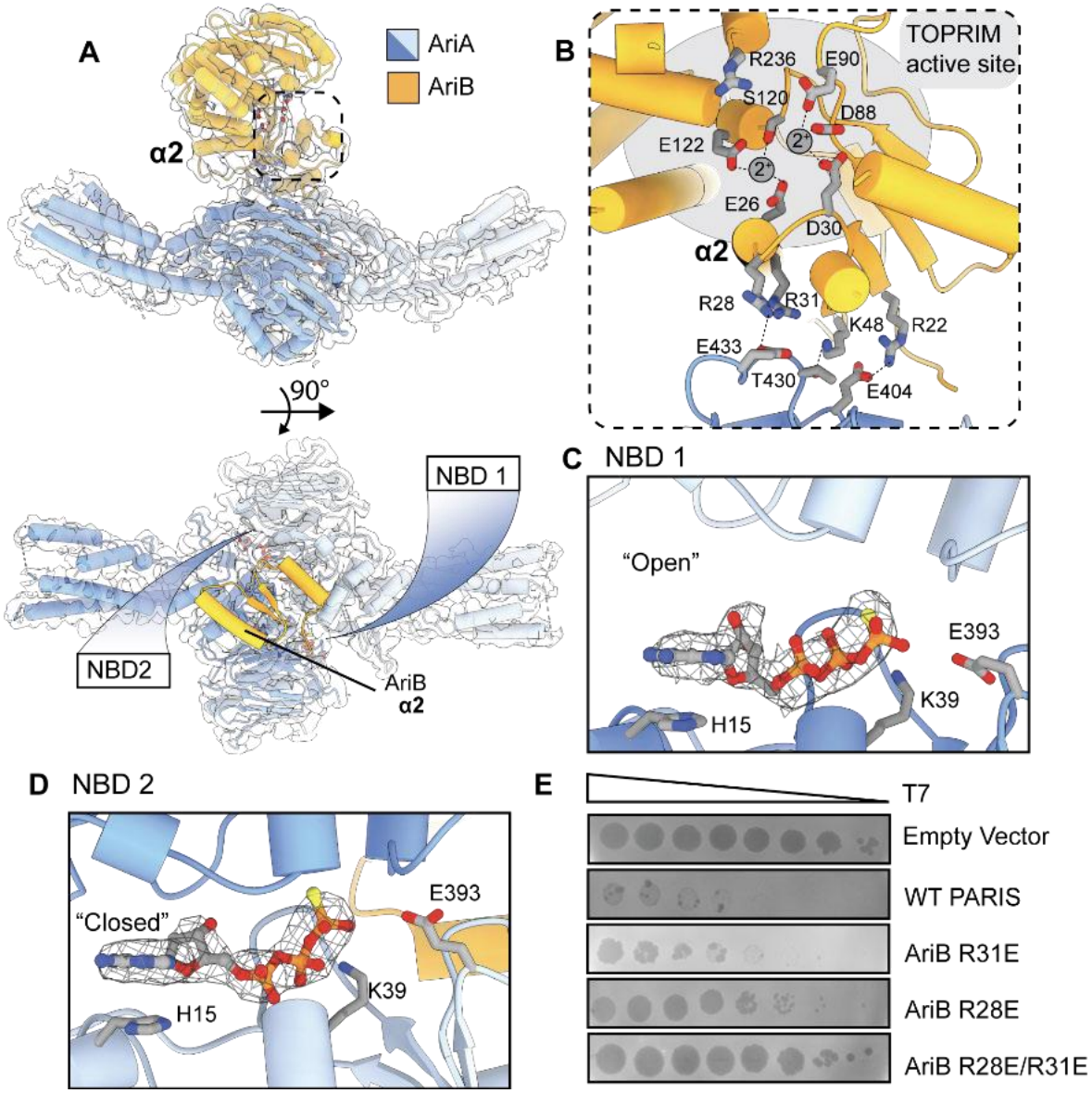


Figure 5.2: Interaction between the AriA ATPase and AriB are essential for PARIS-mediated defense. A) Each asymmetric unit of the PARIS complex contains an AriA homodimer with two nucleotide binding domains (NBD). A 90-degree rotation of the asymmetric unit, showing the N-terminus (res 1-53, yellow) of AriB positioned over the ATPase active sites of AriA. B) AriB attaches to AriA through electrostatic interactions near helix α2 of AriB. TOPRIM active site residues predicted to participate in metal binding (gray spheres) and catalysis are shown as sticks. C) Close-up of NBD1 showing ATPγS bound in an "open" conformation. D) Close-up of NBD2 showing bound ligand trapped in a "closed" pre-cleavage state. E) Plaque assays performed using a serial dilution of phage T7 on E. coli MG1655 carrying the WT or mutated PARIS system.

AriA Binds to the Ocr Anti-Restriction Protein Leading to AriB Release and Activation

The anti-restriction protein Ocr from phage T7 was previously identified as a trigger of the PARIS system4,6. To determine if AriA:B senses Ocr through a direct interaction, we co-expressed AriA with an active site mutant of AriB (E26A) and Ocr-Strep. AriA, but not AriB, co-purified with Ocr-Strep and the complex migrates according to the predicted mass of an AriA hexamer (**Fig 5.3A, Sup Fig 5.4A**). This is consistent with a direct interaction between AriA and Ocr, and release of AriB. A pull-down with AriB-Strep confirms that AriB dissociates from the complex in the presence of Ocr (**Fig 5.3B**).

To determine how Ocr triggers the release of AriB, we used the purified AriA:Ocr-Strep complex for structure determination. The 4.4 Å-resolution structure reveals a D3 symmetric, homohexamer of AriA with no AriB (**Fig 5.3C and Sup Fig 5.5**). At this resolution, the structure of AriA purified using Ocr-Strep is indistinguishable from the structure of AriA purified using AriB-Strep. We anticipated that the structure would include Ocr, since Ocr-Strep was used to pull down the AriA hexamer and the complex was stable during gel filtration. However, multiple data processing strategies failed to resolve Ocr, suggesting that Ocr has several binding sites on AriA and/or that the interaction is conformationally flexible.

Since Ocr is a negatively charged DNA mimic6, we hypothesized that the binding site on AriA would be positively charged. An electrostatic surface of the AriA hexamer reveals a central pore, flanked by a series of three positively charged radial pores (**Fig 5.3C**). The radial pores contain two flexible loops (one from each of the adjacent AriA homodimers) with 14 unmodelled residues (161-174, SSDVGYERRVIRSS), while the central pore contains six loops (one from each AriA) that contain 12 unmodelled residues (113-124, ESERHLRERDVK). To determine the

role of these loops in phage defense, we tested the effects of alanine substitutions and charge swap mutations on positively charged residues. Arginine to alanine substitutions (i.e., R116A and R119A) in the central pore result in a partial loss of phage protection (**Fig 5.3D**). Attempts to generate other mutations in the central (R116E, R119E) or radial pores (R168E, R168A, R172E, or R172A) of AriA consistently failed when we used a vector that also included wild-type AriB. We hypothesized that these mutations in AriA mimic trigger binding, which releases AriB, and results in toxicity. To test this hypothesis, we expressed the central pore mutant (R116E/R119E), with the catalytically inactive AriB (E26A), which is non-toxic. Pull down experiments performed with Ocr-Strep (**Fig 5.3E**) or AriB-Strep (**Fig 5.3F**) demonstrate that AriA (R116E/R119E) has limited interactions with both Ocr and AriB. Collectively, these results suggest that either the central, radial pore or both are involved in the interaction with the trigger. Expression of AriB alone is not toxic to the cells, nor does it provide phage defense4. This suggests that the toxicity of AriB depends on its initial association with AriA. The ATPase activity of AriA is also required for phage defense4, which suggests that the ATPase of AriA is necessary for loading AriB onto AriA, or release of activated AriB. To differentiate between loading or release of AriB, we performed pull down assays using AriB-Strep and WT or an ATPase defective mutant of AriA (K39A). AriB-Strep associates with mutant and wild-type AriA with similar efficiency (**Fig 5.3B**), demonstrating that the ATPase mutant of AriA loads, but does not release AriB. We reasoned that the lack of AriB release in the presence of AriA K39A is due to a defect in trigger recognition. To confirm the role of AriA ATPase activity in trigger recognition, we conducted pull downs using Ocr-Strep and WT or K39A AriA. The AriA K39A mutant fails to co-purify with Ocr, indicating trigger recognition is ATPase dependent (**Fig 3A**).

To determine if or how the trigger impacts the ATPase activity of AriA, we measured the rate of PARIS-mediated ATP hydrolysis in the presence or absence of Ocr and compared these results to an ATPase defective AriA mutant (**Fig 3G**). These results demonstrate that Ocr significantly ($p=0.0001$) reduces ATP turnover by AriA (**Fig 5.3G and Sup Fig 5.6**).

TOPRIM-containing proteins are known to form homodimers23,24, but a predicted homodimer of AriB generated using AlphaFold2 clashes with AriA (**Sup Fig 5.1F)**. Thus, we propose that AriB forms an active homodimer after trigger-mediated release from AriA. The toxicity of PARIS upon activation makes purification of active AriB challenging. To overcome this problem, we mixed cell lysates expressing AriA:B-Strep and Ocr, and then recovered activated AriB using affinity chromatography (**Sup Fig 5.4B-C**). Size exclusion chromatography indicates that activated AriB forms a homodimer (**Fig 5.3H**, **Sup Fig 5.4A**), which was further confirmed by glutaraldehyde cross-linking experiments (**Fig 5.3I**). Mutations in the AriB dimerization interface (E285R and F137A), predicted from the AlphaFold2 model (**Fig 5.1B, Sup Fig 5.1G**), resulted in the loss of phage defense (**Fig 5.3J**), highlighting that AriB dimerization is essential for PARIS activity. Collectively, these data show that the Ocr trigger interacts with AriA in an ATPase dependent manner, leading to the release and dimerization of AriB.

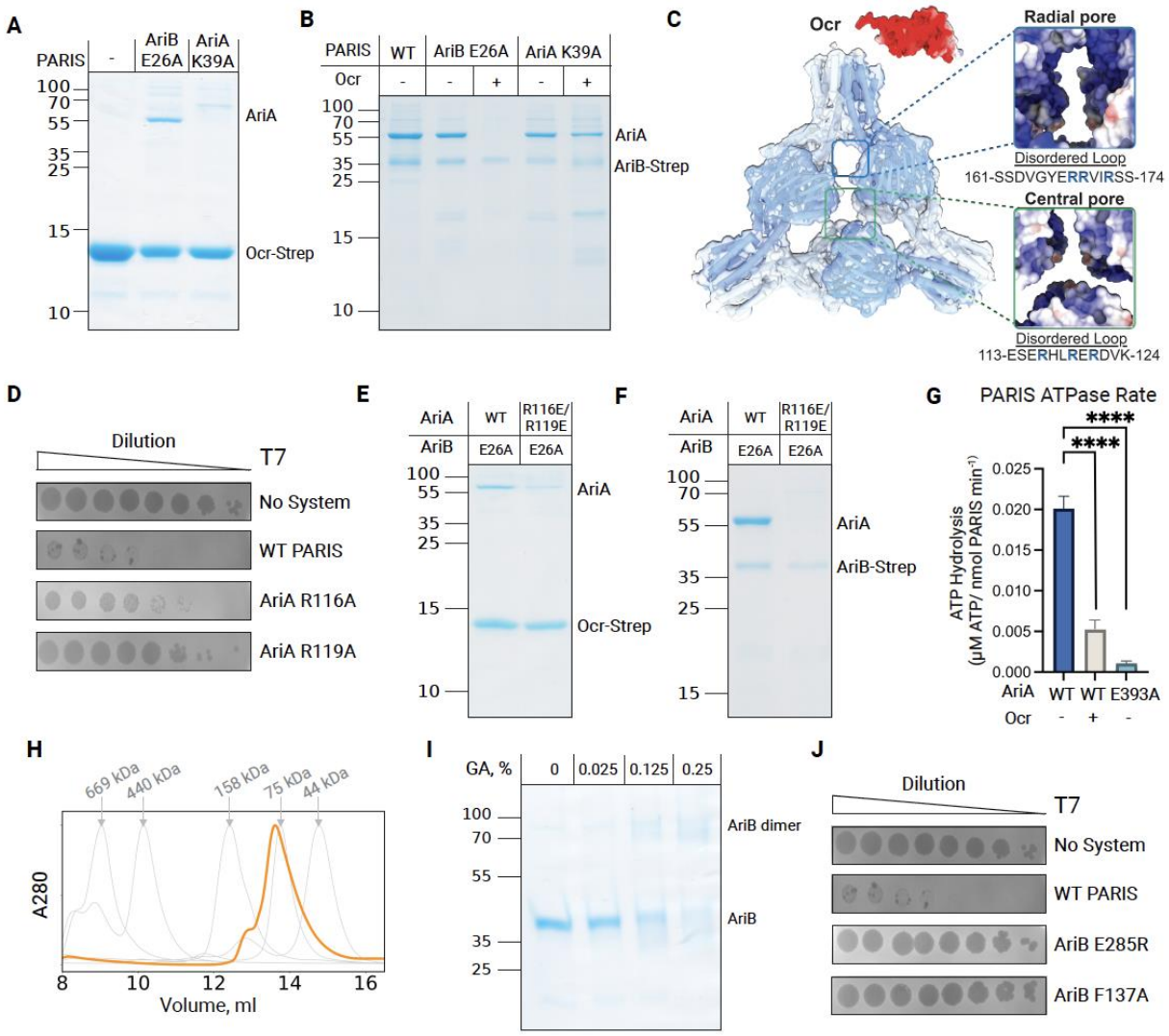


Figure 5.3: AriA interacts with the Ocr trigger in an ATPase-dependent manner, leading to the release of AriB from the PARIS complex. A) Ocr-Strep pulls down AriA, but not AriB. A mutation in the ATPase active site blocks binding of AriA to Ocr-Strep. B) Ocr releases AriB (E26A) from wild-type (WT) AriA (lane 3) but Ocr does not trigger AriB release from an ATPase-deficient variant of AriA (K39A) (lane5). C) A 4.5 Å-resolution reconstruction of AriA that was purified using Ocr-Strep as a bait. The structure reveals three radial pores that are symmetrically positioned around a central pore. The pores contain disordered loops with several positively charged residues. Ocr is shown as an electrostatic surface. D) Mutations in the central pore (i.e., R116A or R119A) reduce the efficiency of phage protection by several orders of magnitude. E) Ocr-Strep binds WT AriA and ejects AriB (E26A), while the charge swap mutations in the central pore of AriA (R116E/R119E) limit interaction with Ocr-Strep. F) AriB-Strep (E26A) pulls down WT AriA but not AriA (R116E/R119E). G) ATP and ADP were separated using thin layer chromatography and the rates of hydrolysis were measured by quantifying the accumulation of ADP (**Sup Fig 5.6)**. The rates of ATP hydrolysis for PARIS, PARIS mixed with ten-fold excess Ocr, or PARIS with an ATPase mutation in AriA (E393A).

The experiments were performed in triplicate. Statistical comparisons between groups were performed using a *post hoc* Dunnett's test ****$p<0.0001$. H) Size exclusion chromatography of AriB-Strep (36 kDa) after Ocr-mediated release from AriA. The column was calibrated using molecular weight standards (gray lines). AriB elutes in single peak with an estimated molecular weight of 81 kDa (**Sup Fig 5.4A**). I) Glutaraldehyde (GA) cross-linking assay with activated AriB (E26A). J) Mutations predicted to block AriB dimerization (E285R and F137A) prevent PARIS-mediated defense (**Sup Fig. 5.1G**).

PARIS Activation Results in Nucleoid Compaction and Translation Inhibition

We then investigated the consequences of PARIS activation by Ocr. The expression of Ocr from a plasmid in *E. coli* is sufficient to cause cell death in the presence of PARIS (**Fig 5.4A**). Monitoring of PARIS+ cells under the microscope using live/dead staining with propidium iodide revealed the accumulation of cells with permeabilized membrane within 20 min of Ocr induction (**Fig 5.4B)** The TOPRIM domain, likely responsible for AriB toxicity, is frequently associated with DNase or RNase activities25, but RNA or DNA preparations from AriB activated cells do not reveal evidence for indiscriminate nuclease activity (**Sup Fig 5.7A-C**). However, DAPI staining reveals that PARIS activation results in nucleoid compaction (**Fig 5.4B**), reminiscent of the translation inhibition phenotype27,28. To understand the nature of this compaction we performed chromosome conformation capture experiments (Hi-C). The results revealed a loss of chromosomal structures similar to that observed upon treatment with chloramphenicol, a known inhibitor of translation (**Sup Fig 5.8**).

To better understand the nature of PARIS-induced translational inhibition, we conducted metabolite labelling experiments and *in vitro* translation assays. Production of Ocr in PARIS+ cells resulted in rapid inhibition of 3H methionine incorporation, consistent with translation inhibition (**Fig 5.4C**), and in a moderate increase of 3H thymidine uptake. This increase is a known consequence of translation inhibition that can be explained by an enhanced transcription of rRNA genes16. The importance of activated AriB in translational arrest was further corroborated using a luciferase reporter in an *in vitro* translation assay. The presence of activated AriB in the reaction mixture interferes with the production of luciferase, while the active site

mutant (i.e., AriB E26A) has no impact (**Fig 5.4D**). Collectively, the results demonstrate that PARIS activation results in translational inhibition.

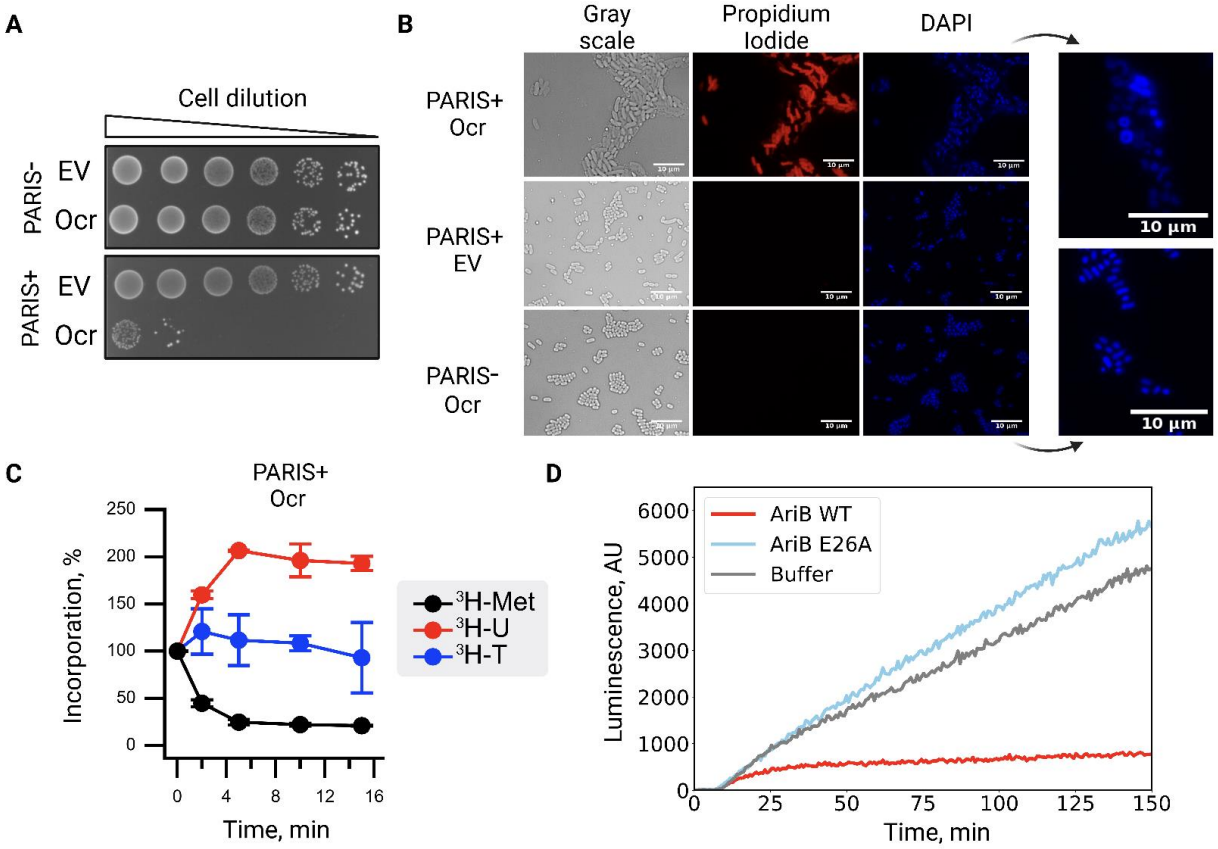


Figure 5.4: The activation of PARIS leads to cell death and inhibition of translation. A) Expression of Ocr from the pBAD vector in PARIS+ culture results in cell toxicity, while the empty vector (EV) has no impact. B) Ocr expression in PARIS+ cells induces cell death (propidium iodide staining, red) and DNA compactization (DAPI staining, blue). Arrows point to close-up images of the nucleoid structures from PARIS+ / Ocr (top) and PARIS- / Ocr (bottom). Cells were imaged 1 hour after induction of Ocr expression. C) Metabolic labelling experiments reveal a drop in 3H methionine incorporation following induction of Ocr, indicative of translation inhibition upon PARIS activation. D) *In vitro* translation in the presence of activated AriB (WT) or active site mutant (E26A). Translation of firefly luciferase mRNA was monitored by accumulation of luminescent signal. In the presence of activated AriB, luciferase signal is reduced dramatically, further indicating that PARIS functions through a mechanism of translational inhibition.

Phage T5 Encodes Two Protein Triggers and a Suppressor of PARIS

The analysis of related phages has provided considerable insight into defense and anti-defense strategies26. While testing the activity of PARIS against phages of the T5 family, we noticed distinct plaquing phenotypes among related phages (**Fig 5.5A**). PARIS does not protect cells from T5wt phage infection, but efficiently blocks infection by a variant of T5 from the Moscow phage collection (T5Mos). Consistent with translation arrest upon PARIS activation, infection of PARIS+ cells with T5Mos at high MOI results in rapid growth inhibition, which occurs faster than the time it takes to lyse PARIS- cells (**Sup Fig 5.9A**). Aside from a large deletion (*Δ30661-38625*), the T5Mos genome is nearly identical to the T5 archetype (**Fig 5.5D**). We hypothesized that both phages contain a trigger that activates PARIS, but the T5 architype also contains a suppressor of PARIS that has been lost in T5Mos. To determine how T5Mos triggers PARIS, we isolated and sequenced T5Mos mutants that escape PARIS immunity. All PARIS-escape mutants carried mutations in two genes encoding highly acidic proteins of unknown function: ORF094 and ORF103 (**Fig 5.5A-B**). Expression of these proteins from a plasmid resulted in the activation of PARIS toxicity (**Fig 5.5C and Sup Fig 5.9**), confirming that T5 carries two novel PARIS triggers. Mutated variants encoded by escape mutants did not trigger PARIS. We named these proteins PARIS trigger 1 and 2 (Ptr1 - ORF094; Ptr2 - ORF 103)27. These results reveal that structurally diverse, negatively charged viral proteins can activate the PARIS system. The presence of *ptr1* and *ptr2* genes in the genome of T5wt, which is not affected by PARIS, further indicates that this phage contains a suppressors of PARIS immunity.

AriB Cleaves the Host tRNALys and Phage T5 Encodes a Non-Cleavable tRNALys Substitute

The T5 genome encodes 23 tRNA, 16 of which have been lost in T5Mos. Given the inhibitory effect of activated AriB on translation, we hypothesized that phage-encoded tRNAs are responsible for overcoming PARIS toxicity. To narrow the search for a PARIS suppressor, we tested a collection of T5 deletion mutants for their sensitivity to PARIS28. Among the variants tested, T5123 carries the smallest deletion in the tRNA genes region (*Δ29191-32442*) and is still sensitive to PARIS (**Sup Fig 5.9B**). The DNA fragment missing from T5123 was divided into three segments, which were separately cloned on an expression plasmid and introduced in PARIS+ cells. A fragment encoding tRNAPro, tRNAfMet, tRNALys and tRNAVal, partially rescued the ability of phage T5123 to infect PARIS+ cells, while other fragments had no effect (**Sup Fig 5.9C**). Next, we cloned each of the tRNAs individually and tested their ability to rescue T5123 infection. Expression of the T5 tRNALys completely restored T5123 infectivity (**Fig 5E**), demonstrating that this phage-encoded tRNA neutralizes PARIS-induced toxicity, and suggests that the host tRNALys is targeted by AriB.

To confirm that the *E. coli* tRNALys is indeed degraded upon PARIS activation, we performed a Northern Blot with total RNA extracted from PARIS+ cells 15 or 30 min after induction of the Ocr trigger. tRNALys was degraded, while SYBR-Gold staining did not reveal degradation of other RNAs (**Fig 5.5F**). To determine how T5 tRNALys suppresses PARIS immunity, we analyzed RNA extracted from PARIS activated cells expressing phage tRNA from a plasmid. Unlike *E. coli* tRNALys, the T5 tRNALys remained intact (**Fig 5.5G).** The phage-encoded variant of tRNALys contains mutations in the anticodon stem-loop (**Fig 5.5H**), which could be responsible for the lack of AriB recognition. To test this hypothesis, we introduced

mutations from the T5 tRNA into the *E. coli* tRNALys counterpart. The resulting chimeric tRNA rescued infection by phage T5123 in the presence of PARIS, confirming that AriB recognizes the tRNA anticodon loop (**Fig 5.5I**). Collectively, these results indicate that phage T5 encodes a non-cleavable variant of tRNA, compensating for the loss of the host tRNA depleted upon PARIS activation.

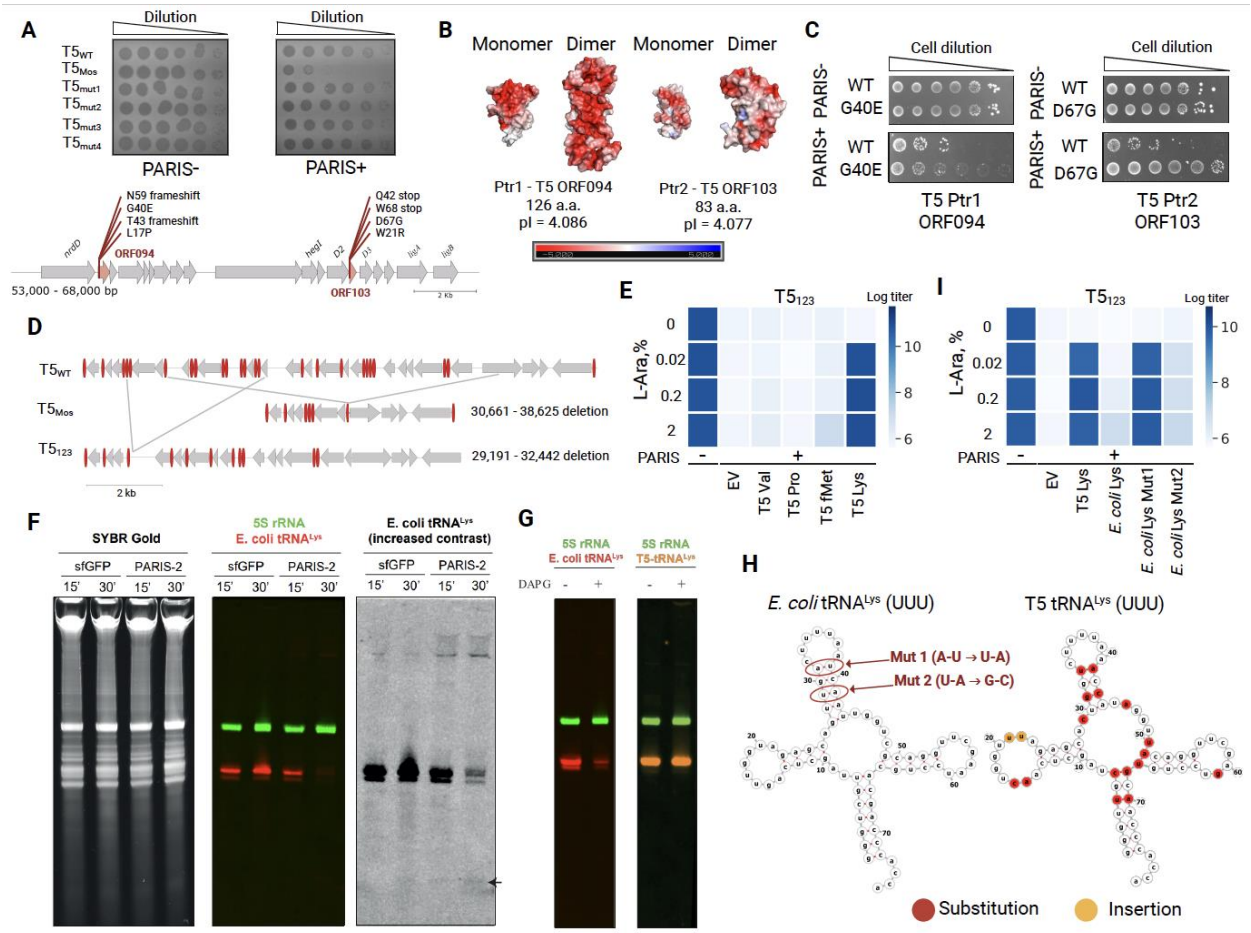


Figure 5.5: PARIS cleaves the *E. coli* tRNALys and T5 bypasses immunity by encoding its own non-cleavable tRNALys. A) Serial dilution plaque assays conducted in PARIS +/- cells against T5wt, T5Mos and T5Mos-derived PARIS escapers. T5Mos phages insensitive to PARIS defense accumulate mutations in ORF094 and ORF103. B) AlphaFold2-predicted structures for T5-encoded PARIS triggers shown as monomers or dimers with their surfaces colored by electrostatic potential. C) Expression of Ptr1 (T5 ORF094) and Ptr2 (T5 ORF103), but not of the selected mutated variants, triggers PARIS toxicity. D) Comparison of T5Mos and T5123 genomes with T5WT. T5Mos is missing an 8 kb fragment while T5123 has a smaller 3 kb deletion. Both deletions encode genes for tRNA, shown in red. E) Overexpression of T5 tRNALys in the PARIS+ cells rescues infection with T5123 phage that lacks tRNALys. F) Northern Blot performed on total RNA extracted from *E. coli* MG1655 carrying the PARIS system (pFR85) or a GFP control (pFR66), and *ocr* under the control of a PPhlF promoter (pFD250). RNA was extracted 15 min or 30min after induction with DAPG. The 5S rRNA is marked with probe B811 (FAM) and the *E. coli* tRNALys with probe B803 (IRDye700). A small black arrow shows a potential degradation product of the tRNALys. G) Northern Blot performed on total RNA extracted form *E. coli* carrying pFR85, pFD250 and the T5 tRNALys under the control of a PBAD promoter (pFD287). RNA was extracted after 30 min of incubation in LB with or without DAPG. The tRNALys with probe B803 (IRDye700) and the T5 tRNALys is marked with probe B806 (IRDye 800).

Evolutionary History of the PARIS Defense System

To inform our understanding of the diversity of PARIS systems and their relationship with other ABC ATPase powered defense systems, we conducted phylogenetic analyses. First, we generated trees for AriA and AriB, revealing a diverse set of systems that can be grouped into 11 distinct clades including two in which *ariA* and *ariB* merged into a single *ariAB* gene (**Fig 5.6A,B**). The *ariA* and *ariB* genes of a same system consistently fall in matching clades, showing they coevolve and are not frequently swapped between systems. A rooted version of the AriB tree shows that AriAB emerged once from clade 9 of two gene systems in a single event and was then split in two different clades (**Sup Fig. 5.10**). Interestingly, sequence alignments of AriA reveal that residues associated with the central and radial pores of the complex are poorly conserved (**Sup Fig 5.10**). This suggests that the signal transduction and effector parts of PARIS immunity are shared across homologs, while the central and radial pores have evolved to recognize different viral triggers.

PARIS was recently assigned to a family of OLD-like defense systems that share an ABC ATPase and a TOPRIM nuclease12. This analysis identified four classes, where the single gene systems such as P2 Old are categorized as class 1, Gabija are class 2, reverse-transcriptase containing systems are class 3, and PARIS systems are class 4. We built a tree of known anti-phage defense systems that carry an ABC ATPase of the AAA15/21 family (**Fig 5.6C**). Our analysis adds AbiL and MADS3-4 systems to the list of systems sharing an ABC ATPase and TOPRIM nuclease. It also reveals how this ABC ATPase can be found in association with a variety of other effector domains, revealing the evolutionary success of this domain as the likely sensor of viral infection in diverse immune systems.

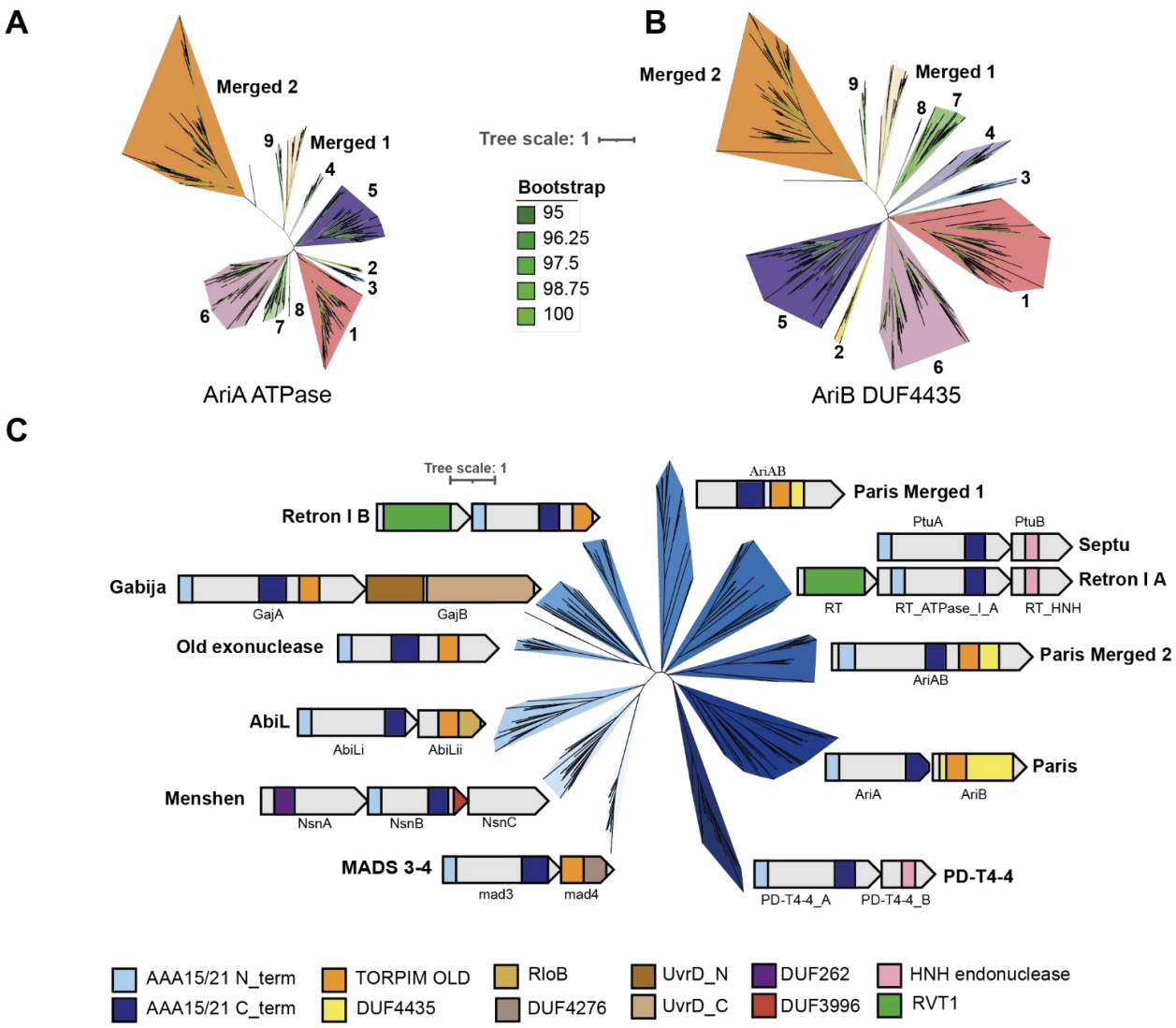


Figure 5.6: Phylogenies of PARIS and related ABC ATPase powered defense systems. A-B) Phylogenetic trees of AriA and AriB share similar branching patterns. Different colors represent different 11 subclades of PARIS including two clades in which *ariA* and *ariB* merged into a single *ariAB* gene (Merged 1 and Merged 2). C) Phylogenetic tree of the AAA15/21 ATPase containing defense proteins. Pfam accession: AAA15 (PF13175), AAA21 (PF13304), TOPRIM-Old (PF20469), DUF4435 (PF14491), RloB (PF13707), DUF4276 (PF14103), UvrD_N (PF00580), UvrD_C (PF13361), DUF262 (PF03235), DUF3996 (PF13161), HNH endonuclease (PF14279, PF10107), RVT1 (PF00078), were used for domain annotations.

Discussion

Our results demonstrate how AriA forms propeller-shaped hexamer capable of binding three AriB monomers. AriA is an ATP-dependent sensor that detects viral immune suppressors (e.g., Ocr), which trigger the release of AriB. Released, AriB assembles into a homodimer that cleaves specific host tRNAs resulting in a rapid interruption of translation followed by a loss of membrane integrity. As opposed to prototypical toxin-antitoxin systems, AriB isn't toxic when expressed alone. This suggests that AriA activates AriB through a structural modification that enables dimerization or an unidentified posttranslational modification. More work will be necessary to determine the mechanism of AriB release and activation.

Interestingly, we found that the PARIS system from *E. coli* B185 is not only triggered by phage T7's Ocr protein, but also by Ptr1 and Ptr2 proteins from phage T5. Despite these proteins being small and negatively charged, they lack obvious sequence or structural similarities. Determining how PARIS detects such diverse proteins and whether different systems have distinct trigger sensitivities presents another intriguing avenue for future research.

Our work adds to a growing list of defense systems that trigger translational arrest upon infection. The PrrC and Retron I A (i.e., Ptu*AB* proteins) systems target tRNA degradation after detecting foreign proteins, while Cas13a targets tRNA degradation after detection of foreign RNA29–31. However, translational arrest isn't limited to tRNA degradation, Mogila *et al* recently identified a CRISPR-Cas system that cleaves mRNA associated with the ribosome which arrests translation32, and the Lit (Late Inhibition of T4) protease degrades the translation elongation factor EF-Tu after detection of the major capsid protein of phage T433.

We show that some phages circumvent PARIS-induced toxicity by expressing non-cleavable tRNA variants. Many tailed phages, such as those from *Demerecviridae* (T5-like phages), *Straboviridae* (T4-like phages) or *Ackermannviridae* families encode large arrays of tRNAs previously thought to be associated with translation optimization due to codon bias in the host and viral genome34. However, estimation of the abundance of host and phage tRNAs during infection, compared to the corresponding codon frequencies in produced mRNAs does not directly support this hypothesis34. A recent computational analysis suggests that phages may have evolved these variants in response to bacterial tRNA-targeting toxins, as evidenced by distinct anti-codon loop structures aimed at evading host nucleases35. This hypothesis is further supported by a parallel work demonstrating that phage T5 tRNATyr provides protection against the PtuAB toxin of the Eco7 (Ec78) retron31.

Interestingly, this strategy of encoding tRNA clusters is also observed in some eukaryotic viruses, like those from *Herpesviridae, Mimiviridae* and *Phycodnaviridae* families36, possibly serving a similar function in evading host immunity. Strikingly, an anti-codon nuclease was also recently found to protect humans against Pox viruses, demonstrating that the inhibition of translation through the inactivation of specific tRNAs is an anti-viral strategy in each domain of life37.

Methods

Bacterial Strains, Phages, and Plasmids

Phages, bacterial strains and plasmids used in the study are listed in Supplementary Table 2. Primers are listed in Supplementary Table 3. Infection with T5 phage was carried in LB supplemented with 1mM CaCl2. Unless otherwise indicated, 0.2% L-arabinose, 0.1mM IPTG, and 200 ng/mL aTc were used for the induction.

For structural studies, the coding sequences for AriA and AriB from E. coli B1854 were cloned under control of a T7 promoter into a pRSF-Duet vector with a C-terminal strep-tag II on AriB. For co-expression of PARIS with T7 Ocr trigger, the TOPRIM nuclease of AriB was inactivated (E26A) and the strep-tag II was removed by site-directed mutagenesis. The coding sequence for T7 Ocr protein was cloned into a pET-Duet vector with a C-terminal strep-tag II. For in vitro binding assays, the coding sequence for T7 Ocr protein was cloned into a pET-Duet vector with a N-terminal His6-TwinStrep-SUMO tag. All in vivo experiments were performed with pFR85, encoding PARIS from E. coli B185 with its native promoter and an inducible Ptet promoter upstream. For in vivo protein pull-downs, C-terminal AriB strep-tag II was introduced to the pFR85. Mutants of PARIS were constructed from pFR85 using the Gibson method38 or Q5 site-directed mutagenesis kit (NEB). The PARIS triggers Ptr1 (T5 ORF094) and Ptr2 (T5 ORF103) from T5 phage as well as T5 genome fragments, T5 tRNALys and E. coli tRNALys were cloned on the pBAD18 vector under control of araBAD promoter using the Gibson method. 20 bp upstream and downstream of tRNA genes were preserved during cloning. T5 nucleotide positions are provided according to the genome assembly AY543070.1. All constructions were verified using Sanger sequencing.

Expression and Purification of the PARIS Complex

AriA and AriB-strep were co-expressed from a pRSF-Duet plasmid in BL21-AI cells (Invitrogen). Cultures grown at 37°C to 0.4-0.5 OD600 were induced with 0.1 mM IPTG and 0.1% L-arabinose. Following induction, the cultures were incubated overnight at 16°C. Cells were pelleted (3,000 × g, 10 min, at 4°C), resuspended in lysis buffer (25 mM Tris, pH=8.5, 150 mM NaCl, 1 mM EDTA) and lysed by sonication. Cellular debris was removed by centrifugation (10,000 × g, 25 min, at 4°C) and AriAB complex was purified by affinity chromatography on 5 mL streptactin resin (IBA) and eluted with 2.5 mM desthiobiotin (IBA) in lysis buffer. The eluted protein was concentrated (100k MWCO PES Spin-X UF concentrator, Corning) and further purified by size exclusion chromatography (Sup200 column, Cytiva) in lysis buffer with 2% glycerol. Fractions of interest were combined and concentrated to 5μM (100k MWCO PES concentrator, Pierce) and used immediately for vitrification on cryo-EM grids.

In vivo Protein Pull-Down

To determine protein-of-interest interacting partners in vivo, AriB-Strep expressed from pFR85 or Ocr-Strep expressed from pBAD were used as bait, and pull-downs were performed in E. coli BW25113. 3 L of cell culture was grown in LB media at 37°C in Thomson flasks with aerated lids (1.5 L per flask). Expression of proteins was induced with 200 ng/mL aTc or 0.2% L-arabinose at OD600 ~ 0.1 and cells were harvested by centrifugation after reaching OD600 ~ 0.8-1. Cells were processed as described above, and Strep-tagged proteins were purified on two stacked 5 mL StrepTrap HP (Cytiva) columns. Identity of protein bands after SDS-PAGE was determined by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. ~1/3 of the Coomassie-stained band was cleaved from the gel and samples were

prepared with Trypsin Gold (Promega) in accordance with manufacturer's instructions. Mass spectra were obtained using the rapifleX system (Bruker).

For production of activated AriB, AriB-Strep from pFR85 and non-tagged Ocr from pBAD were expressed in separate cultures and cells were grown as described. Cell cultures were mixed, and cell walls were disrupted by sonication on ice (60% power, 10'' pulse, 20'' pause, 30-60 minutes) on Qsonica sonicator with 6 mm sonotrode, followed by StrepTrap HP purification of activated AriB-Strep. The molecular weight of protein complexes was determined after size exclusion chromatography performed on Superdex 200 Increase 10/300 column (GE Healthcare) calibrated with High Molecular Weight calibration kit (GE Healthcare).

Glutaraldehyde Cross-Linking Assay

To determine the oligomeric state of activated AriB in solution, we performed glutaraldehyde cross-linking assay. Concentrated AriB aliquots were mixed with glutaraldehyde (253857.1611; PanReac) to a final concentration of 0.025%, 0.125%, or 0.25% vol/vol and incubated for 30 min at room temperature. The reaction was quenched by the addition of one sample volume of 1 M Tris-HCl at pH 8.0 and a half sample volume of 5x concentrated SDS sample buffer. Samples were heated for 5 min at 60°C and separated by SDS-PAGE.

In vitro Protein Pull-Down

The T7 Ocr protein with an N-terminal His6-TwinStrep-SUMO was expressed from a pET-Duet plasmid in BL21(DE3) cells. Cultures grown at 37°C to 0.4-0.5 OD600 were induced with 0.5 mM IPTG. Following induction, the cultures were incubated overnight at 16°C. Cells were pelleted (3,000 × g, 10 min, at 4°C), resuspended in lysis buffer (25 mM Tris, pH=8.5, 150 mM NaCl, 1mM EDTA) and lysed by sonication. Cellular debris was removed by centrifugation

(10,000 × g, 25 min, at 4°C) and HSS-Ocr was purified by affinity chromatography on 5 mL streptactin resin (IBA) and eluted with 2.5 mM desthiobiotin (IBA) in lysis buffer. The eluted protein was incubated with SUMO protease at 4°C overnight before washing over 5 mL Ni-NTA resin (25 mM Tris pH=7.5, 300 mM NaCl, 1 mM TCEP, 20 mM Imidazole) to remove SUMO-cleaved tag. Flow through protein was concentrated (10k MWCO PES Spin-X UF concentrator, Corning) and further purified by size exclusion chromatography (Sup6 10 300 column, Cytiva) in lysis buffer with 2% glycerol. Fractions of interest were combined and concentrated (10k MWCO PES Spin-X UF concentrator, Corning). Cleaved Ocr (3.8 uM) was incubated with WT PARIS (11.3 μM) at 37°C for 15 minutes before analysis via size exclusion chromatography (Sup6 10 300 column, Cytiva) in lysis buffer with 2% glycerol. Fractions of interest were combined, concentrated (10k MWCO PES Spin-X UF concentrator, Corning), and analyzed via SDS-PAGE (15% acrylamide, 150V for ~1 hour).

Cryo-EM Sample Preparation and Data Collection

All samples were subjected to vitrification on 1.2/1.3 Carbon QuantifoilTM grids which were glow discharged using an easiGlowTM (Pelco) glow discharge station (glow 45s hold 15s). After applying 3 μL of 5μM PARIS, grids were subjected to double-sided blotting using a Vitrobot Mk IV with a force of 5 for 4 seconds at 4° and 100% relative humidity. After plunge-freezing in liquid ethane, grids were clipped and stored in liquid nitrogen for subsequent analysis.

For the PARIS complex in the inactive state obtained from AriB pulldowns, cryo-EM grids were screened, and all data were collected using a Talos Arctica microscope (ThermoFisher) operating at 200kV equipped with a K3 direct electron detector (Gatan).

SerialEM39 software version 3.8.6 was used to automate data collection. For final data collection, a total of 7,340 movies were collected in super resolution mode with a pixel size of 0.552Å/pixel and a total dose of 56.42 e/Å2 distributed over 50 subframes.

For in vitro Ocr pulldown data collection, samples were prepared for cryo-EM analysis as described above. After screening grids, 10,340 movies were collected using a Talos Arctica microscope (Thermofisher) operating at 200kV equipped with a K3 direct electron detector (Gatan). SerialEM software version 3.8.6 was used to automate data collection. Movies were recorded in super resolution mode with a pixel size of 0.552Å/pixel and a total dose of 56.31 electrons/Å2 distributed over 50 subframes.

Cryo-EM Data Processing of Inactive PARIS

All data were processed in cryoSPARC v4.41. After patch motion correction and CTF estimation, micrographs with CTF fits worse that 8 Å were discarded, yielding 5,998 micrographs for particle picking. After processing a subset of 200 micrographs, the AriA6:AriB3 assembly of the PARIS complex was apparent. Using this low-resolution initial volume, Alphafold2 predicted structures of AriA and AriB were fit into the map and used to generate a 20 Å lowpass filtered map using the Molmap command in ChimeraX. This volume was imported to cryoSPARC and used to generate templates for particle picking. A total of 4,078,384 particles were extracted and after initial 2-D classification, 1,643,515 particles were kept for further processing. These particles were sorted using a three-class ab initio reconstruction which revealed two junk classes and a volume corresponding to the assembled PARIS complex that contained 940,782 particles. After a second round of 2-D classification, 934,763 particles remained were selected and subjected to another round of ab initio reconstruction and

heterogeneous refinement. This produced a consensus volume containing 532,010 particles that contains conformational heterogeneity for two AriB subunits, while 1 asymmetric unit of the complex is well aligned. Masked local refinement was used to determine a 3.2 Å-resolution map of one asymmetric unit of the complex. By subjecting the consensus refinement with 532,010 particles was also subjected to multiple rounds of masked 3-D classification to isolate particles in the 'cis' and 'trans' arrangements. The fully assembled PARIS complex demonstrates flexibility based on 3DVA, which limited the attainable resolution of the fully assembled complex. The density map for both the local refinement and the full complex are available at EMDB-42719,43103, and 43104 and raw micrographs are available at EMPIAR-11832.

Cryo-EM Data Processing of Ocr-Pulldown

All data were processed in cryoSPARC v4.41. Of the 10,340 collected micrographs, 9,399 were selected for further processing based on the CTF fit criteria described above. Using the blob picker with a particle diameter of 200 Å, particles were extracted and subjected to 2-D classification. In the 2-D classes, there were two distinct macromolecular complexes present, one of which was interpreted to be the AriA scaffold of the PARIS complex, while the other was interpreted to correspond to RNA polymerase (RNAP).

From seven selected classes corresponding to the AriA hexamer, 667,479 particles were selected for downstream refinement. From a 2-class ab initio reconstruction of these particles, 392,989 were kept. Iterative sorting by multi-class ab initio reconstruction and heterogenous refinements yielded a volume containing 62,732 particles. AriA subunits can be fit into this density, but the poor resolution makes this structure indistinguishable to the AriA scaffold identified in the intact PARIS complex. No density is observed for Ocr, which suggests multiple

binding sites may be present on the PARIS complex, or that Ocr is attached to AriA in a flexible state. Density map available at EMDB-43105 and raw micrographs are available at EMPIAR-11833.

From 6 selected classes corresponding to RNAP, 548,741 particles were subjected to a 2-class ab initio reconstruction. After removing a junk class, 371,555 particles remained. Further sorting using iterative ab initio reconstructions and heterogenous refinements isolated a stack of 190,177 particles. Non uniform refinement of these particles reveals a density map that closely resembles the structure of RNAP but suffers from severe anisotropy, preventing identification of bound Ocr.

Model Building and Refinement

To determine the structure of the asymmetric unit of the PARIS complex, density halfmaps were obtained from cryoSPARC local refinements before being provided as inputs to DeepEMhancer for map sharpening. Alphafold predicted structures of AriA and AriB were then docked into the map density and map to model fit was improved using ISOLDE's molecular dynamics simulation environment in ChimeraX. After initial equilibration, the PDB files were saved and used as inputs for realspace refinement using PHENIX. After initial refinement, problematic areas were manually inspected and modified using COOT and sidechains were removed from regions with poorer than 4 Å-resolution. After addressing clashes and non-rotameric sidechain orientations, iterative refinements were carried out in Phenix until Clashscores, Ramachandran and sidechain outliers failed to improve. The final model and their corresponding density map was submitted to the PDB for deposition (PDB ID: 8UX9, EMDB-42719,)

To generate the biological assemblies for the PARIS complex atomic model, halfmaps from C3 symmetric, and nonuniform refinements of particles corresponding to the ‘cis’ and ‘trans’ orientations were obtained from cryoSPARC and sharpened in DeepEMhancer. After fitting the three copies of the asymmetric unit into the ‘cis’ and ‘trans’ density maps using the ChimeraX fitmap command, the find NCS tool in Phenix was then used to calculate the appropriate symmetry operators for the assembled forms of the complex. These symmetry operators were applied to the Asymmetric unit to generate biological assembly 1 and 2 using the Apply NCS tool in Phenix. Maps for the ‘cis’ and ‘trans’ orientations of the PARIS complex are available at EMDB-43104 and EMDB-43105, respectively.

ATPase Assays

ATPase activity was measured using 6uM Ocr, 100 nM PARIS and 128 μM ATP supplemented with 10 nM [α-32 P]-ATP (PerkinElmer). The reaction buffer includes 20 mM Tris-HCl pH 8.0, 150 mM Sodium Chloride, 1 mM DTT, and 5 mM magnesium chloride. The reactions incubated at 37ºC and 10 uL aliquots were quenched with Phenol at 1, 2, 4, 8,16, and 32 minutes. For the ‘trigger’ only reactions, 100 nM Ocr was mixed with 128 μM ATP supplemented with 10 nM [α- 32 P]-ATP (PerkinElmer) in the reaction buffer. Reactions were incubated for 32 minutes at 37ºC and quenched with phenol. The ATP and ADP makers were created using T4 polynucleotide kinase (NEB). T4 PNK mixed with 128 μM ATP supplemented with 10 nM [α-32 P]-ATP, and 1 mM DNA oligo in T4 polynucleotide kinase buffer (NEB). This mixture was then incubated at 37ºC for 1 hour. All reaction products were phenol-chloroform extracted and resolved on silica TLC plates (Millipore). Reaction products were spotted 2.5 cm above the bottom of the TLC plate. The plate was placed in a TLC developing chamber filled to

~1.5 cm with developing solvent (0.2 M ammonium bicarbonate pH 9.3, 70% ethanol, and 30% water) and covered with aluminum foil for 4 hrs at room temperature. The TLC plate was exposed to a phosphor screen, imaged with Typhoon phosphor imager and quantified with the ImageQuant software package (Cytiva).

Efficiency of Plaquing Assay

The activity of PARIS mutants was measured in comparison to the wild-type system by performing efficiency of plaquing (EOP) assays with phage T7. E. coli K-12 MG1655 carrying each of the AriA or AriB mutants, the control plasmid pFR66 (sfGFP), or the wild-type PARIS system (pFR85) were grown overnight in LB + kanamycin 50 μg/ml. Bacterial lawns were prepared by mixing 100 μl of a stationary culture with 5 ml of LB + 0.5% agar, and the mixture was poured on Petri dish of LB + kanamycin 50 μg/ml. Ten-fold serial dilutions of high titer stock of T7 phage were spotted on each plate and incubated 5h at 37°C. EOPs with T5 and T5-like phages were performed with E. coli BW25113 as described above, except that plates were incubated at 37°C overnight, and top agar was supplemented with 1mM CaCl2. Induction of tRNA genes or phage T5 fragments was achieved by supplementing top agar with indicated amount of L-arabinose. Plaque assays were performed in at least two independent replicates.

Liquid Culture Phage Infection

To monitor dynamics of T5Mos phage infection of the PARIS+ (pFR85) and PARIS- (pFR66) culture in liquid, we used EnSpire Multimode Plate Reader (Perkin Elmer). Overnight bacterial cultures were diluted 100-fold in LB medium with appropriate antibiotics and grown at 37°C in 10 mL of LB supplemented with 1mM CaCl2. At OD600 = 0.6, 200 μl aliquots were transferred to 96-well plates and infected at indicated Multiplicity of Infection (MOI). Optical

density was monitored for 10 hours. All experiments were performed in three biological replicates.

Toxicity Assay

PARIS toxicity in the presence of T7 Ocr or Ptr1 and Ptr2 was measured in a spot-test assay. PARIS+ (pFR85) and PARIS- (pFR66) cultures carrying pBAD vector encoding indicated trigger were grown overnight in 10 ml LB at 37°C. Stationary cultures were diluted to OD=0.6 and plated on LB or minimal media (M9 with 5% v/v LB) agar plates supplemented with 0.2% L-arabinose by serial 10-fold dilution. Control plates without induction contained 0.2% glucose to prevent leakage of araBAD promoter.

Fluorescence Microscopy

PARIS+ (pFR85) and PARIS- (pFR66) carrying pBAD Ocr were diluted 1:100 and cultivated in LB supplemented with appropriate antibiotics at 37°C. When OD600 reached 0.4, an aliquot was mixed with DAPI (Invitrogen) at 1 mg/mL final concentration and propidium iodide (PI, Invitrogen) at 1 μg/mL final concentration and incubated at room temperature in the dark for 5 minutes. LB + 1.5% agarose slabs (~0.2 mm thick) were prepared on a 75 x 25 mm microscopy slide (Fisher Scientific). Agarose slabs contained staining dyes at the same concentration and optionally were supplemented with 0.2% L-arabinose, to induce Ocr expression. ~1 μl of stained cells was placed on agarose slab and after 1-min drying the slab was covered by a small 22 × 22 mm coverslip (Fisher Scientific). Imaging was performed on a Nikon Eclipse Ti-E inverted epifluorescence microscope. Each field of view was imaged in transmitted light channel (200 ms exposure) and in DAPI (filterset Semrock DAPI-50LP-A, 200 ms exposure) and in PI (filterset TxRed-4040C, 200 ms exposure) channels.

Extraction of Total DNA

Total DNA was extracted at different time points (15, 30, 60 and 120 min) with or without DAPG to induce the expression of Ocr in the presence of PARIS. DNA was extracted using the Wizard® genomic DNA purification kit (Promega) according to the manufacturer's instructions and treated with RNase at 0.1 mg/ml final. Each sample of total DNA was loaded on 1% agarose 1X TAE gel.

TUNEL Assay

TUNEL assays was performed to estimate in vivo accumulation of dsDNA breaks, according to methods described previously40. PARIS+ (pFR85) culture carrying pBAD Ocr vector was grown to OD600=0.3, followed by Ocr induction with 0.2% L-arabinose for 2 hours. As a positive control, cells were treated with 0.1% H2O2 to induce accumulation of DNA breaks. 2 ml of cells were harvested by centrifugation and treated according to the standard protocol (TUNEL Assay Kit – FITC, ab66108, Abcam). Measurement of FITC fluorescence was performed with CytoFLEX cytometer (Beckman Coulter) in the FITC channel. 100,000 events were collected for each sample in three biological replicates. Data was analyzed and visualized in FlowJo v10.

Hi-C Procedure and Sequencing

E. coli MG1655 carrying plasmids pFR85 (Ptet ariAB) and pFD250 (PPhlF ocr), or the pFR66 (sfGFP) + pFD245 (PhlF sfGFP) control plasmids were diluted 1:100 from an overnight culture in LB + kanamycin 50 μg/ml + chloramphenicol 20 μg/ml and grown until OD600 = 0.3. DAPG 50 μM was then added, followed by incubation for 15 and 30min. To compare the effect of PARIS activation with that of treatment with chloramphenicol, E. coli MG1655 harboring

pFR66 were diluted 1:100 from an overnight culture in LB + kanamycin 50 μg/ml until OD600 = 0.3, followed by addition of chloramphenicol 20 μg/ml and incubation for 15 and 30min. Cell fixation was performed with 4% formaldehyde (Sigma-Aldrich, Cat. F8775) as described in Cockram et al. (2021)41. Quenching of formaldehyde with 300 mM glycine was performed at 4°C for 20 min. Hi-C experiments were performed with the Arima kit. Samples were sonicated using Covaris (DNA 300bp).

Preparation of the samples for paired-end sequencing was performed using Invitrogen TM Colibri TM PS DNA Library Prep Kit for Illumina according to manufacturer instructions. The detailed protocol is available in Cockram et al. (2021)41. Reads were aligned with bowtie2 v2.4.4 and Hi-C contact maps were generated using hicstuff v3.0.3 (https://github.com/koszullab/hicstuff) with default parameters and using HpaII enzyme to digest. Contacts were filtered as described in42, and PCR duplicates (defined as paired reads mapping at exactly the same position) were discarded. Matrices were binned 4kb. Balanced normalizations were performed using ICE algorithm43. For all comparative analyses, matrices were downsampled to the same number of contacts. The Hi-C signal was computed as the contacts between adjacent 5kb bins as described in Lioy et al. (2018)44. In order to compare this signal with other genomics tracks, we binned it at the desired resolution.

Metabolic Labeling

Metabolic Labeling with 3H-methionine, 3H-uridine and 3H-thymidine Metabolic labelling experiments were performed as described previously45, with minor modifications. E. coli BW25113 PARIS+ (pFR85) strain was transformed with pBAD plasmid carrying Ocr trigger (for L-arabinose-inducible expression). Transformed cells were initially plated on LB plates

supplemented with 100 μg/mL ampicillin, 25 μg/mL kanamycin, and 0.2% glucose. Using individual E. coli colonies for inoculation, 2 mL liquid cultures were prepared in defined Neidhardt MOPS minimal media (supplemented with 100 μg/mL ampicillin, 25 μg/mL kanamycin, 0.1% casamino acids, and 1% glucose, and grown overnight at 37°C with shaking. Subsequently, experimental 20 mL cultures were prepared in 125 mL conical flasks in MOPS medium, supplemented with 0.5% glycerol, 100 μg/mL ampicillin, 25 μg/mL kanamycin, as well as a set of 19 amino acids (lacking methionine), each at a final concentration of 25 μg/mL. These cultures were inoculated at OD600 of 0.05 and grown at 37°C with shaking until the OD600 reached 0.3. At this point, 1 mL aliquots (designated as the pre-induction zero time-point) were transferred to 1.5 mL Eppendorf tubes containing 10 μL of pre-incubated at 37°C respective radioisotope (3H methionine [0.77 μCi, Perkin Elmer], 3H uridine [0.1 μCi, Perkin Elmer] or 3H thymidine [0.32 μCi, Perkin Elmer]). Concurrently, Ocr expression in the remaining 19 mL culture was induced by adding L-arabinose to a final concentration of 0.2%. Throughout the Ocr induction time course, 1 mL aliquots were taken from the 20 mL culture and transferred to 1.5 mL Eppendorf tubes containing 10 μL of the appropriate radioisotope. Radioisotope incorporation was halted after 8 minutes of incubation at 37°C by adding 200 μL of ice-cold 50% trichloroacetic acid (TCA) to the 1 mL cultures. Additionally, 1 mL aliquots were sampled for OD600 measurements during induction time course. The resultant 1.2 mL TCA-stopped culture samples were loaded onto GF/C filters (Whatman) prewashed with 5% TCA and unincorporated label was removed by washing the filter twice with 5 mL of ice-cold TCA followed by a 5 mL wash with 95% EtOH (twice). The filters were placed in scintillation vials, dried for two hours at room temperature, followed by the addition of5mLEcoLite™-scintillation cocktail (MP

Biomedicals). After shaking for 15 minutes radioactivity was quantified using automatic TDCR Liquid Scintillation Counter (HIDEX). Isotope incorporation was quantified by normalizing radioactivity counts (CPM) to OD600, with the pre-induction zero time-point serving as the reference (set to 100%). All experiments were performed as biological triplicates using three independent liquid starter cultures inoculated with different colonies.

In vitro Translation

In vitro translation was monitored by the production of luciferase signal in a PURExpress in vitro protein synthesis kit (NEB), using firefly luciferase mRNA as an input. pT7-Fluc, encoding firefly luciferase (fluc) gene under control of T7 promoter, was used as a template for T7 in vitro transcription with MEGAscript kit (Thermo Scientific). mRNA was treated with DNase I (Thermo Scientific) and purified with Monarch PCR & DNA cleanup kit (NEB). PURExpress in vitro translation reaction was assembled in RNase-free tubes according to manufacturer's instructions, with minor modifications. Reaction mixtures were adjusted to a total volume of 5 µl (2 µl solution A; 1,5 µl solution B; 0,3 µl RiboLock (40 U/µl, Thermo Scientific); 0,2 µl D-Luciferin (10 mM, Sigma)), supplemented with 0.5 µl of activated AriB or AriB (E26A) (60 µg/ml final concentration) or the same volume of buffer A (50 mM NaCl; 1mM EDTA; 100 mM Tris-HCl; 5mM β-ME; pH 8) as a control. Reaction was started by the addition of 0.5 µl of purified fluc mRNA (500 ng), immediately placed in 384-well white plate and covered with optically clear film. Accumulation of luminescent signal was measured by EnSpire Multimode Plate Reader (PerkinElmer) for 3h at 37°C. AriB concentration was determined with Qubit Protein Broad Range Assay Kit (Invitrogen).

Isolation and sequencing of T5 escaper mutants

To select PARIS escape mutants, phage T5Mos was continuously incubated with PARIS+ (pFR85) culture for 3 days. In short, after reaching an OD600 = 0.6, cells were infected with T5Mos at MOI~0.1, and the culture was incubated overnight at 37°C. Phage was collected the next day and 1 ml of lysate was used to initiate the next round of infection with fresh PARIS+ (pFR85) culture. Escaper mutants were re-purified from the single plaques obtained on PARIS+ (pFR85) culture and produced on PARIS-(pFR66) culture. 8 mL of high-titer lysate (~1010 pfu/ml) was PEG-precipitated, and phage genomic DNA was purified as described previously46. DNA libraries were prepared according to a standard protocol and sequenced on MiniSeq platform (Illumina) with paired-end 150 cycles (75 + 75). Genome assemblies were performed with SPAdes implemented in Unicycler47. To identify mutations, genomes of T5Mos escapers were aligned with the sequenced genomes of T5Mos and T5wt from initial stocks. T5 strains reported by Glukhov et. al.28 were sequenced on DNBSeq-G400 platform (BGI) to validate boundaries of deletions.

Extraction of Total RNA

Total RNA was isolated from E. coli strain MG1655 harboring pFR85 (Ptet ariAB) or a control plasmid pFR66 (Ptet sfGFP). Cells also carried pFD250 (PPhlF ocr) with Ocr under the control of an inducible DAPG PPhlF promoter (Fig 5F and Sup Fig 7B), as well as plasmid pFD287 carrying the tRNALys of phage T5 under the control of ParaBAD promoter inducible by arabinose (Fig 5G).

Cells were diluted 1:100 in LB with the appropriate antibiotics from overnight cultures and grown with agitation at 37°C. When cultures reached an OD600 of 0.2, aTc was added to a final concentration of 0.5 μg/ml to induce the expression of the PARIS system. For the experiment

shown in Fig 5G, arabinose was added to a final concentration of 0,2% to induce the T5-tRNALys from plasmid pFD287. Growth continued until 0.4 and the Ocr trigger was induced with DAPG added to a final concentration of 50 µM, followed by incubation for an additional 15 and 30mn post induction.

Cells were centrifuged at 4,000 g for 10 min, resuspended with 0.2 ml of lysozyme buffer (20 mM Tris-HCl, pH8.0, 2 mM EDTA, 1% Triton X-100 and 20 mg/ml lysozyme) and lysed for 30 min at 37°C. One milliliter of TRIzol reagent (Zymo Research) was added to the samples and incubated for 5 min at room temperature (RT). Lysates were extracted with 0.2 ml of cold chloroform and centrifuged at 12,000 g for 15 min at 4°C. The aqueous phase of the sample was precipitated with 0.5 ml of 100% cold isopropanol, incubated 10 min at RT and centrifuged at 12,000 g for 10 min at 4°C. RNA pellets were washed with 1 ml of 75% ethanol, dried at RT and dissolved in 50 µl of nuclease free water. Total RNA concentration was measured with a nanodrop spectrophotometer and then the samples were resolved in a 7 M Urea 10% acrylamide gel before visualization with SYBR-Gold.

Northern Blot

RNAs were run for 1 hr at 180V in a 7M Urea 10% acrylamide gel and then transferred to a nylon membrane (Invitrogen) using XCell SureLock Mini-Cell system with the XCell II Blot Module (Invitrogen). The membrane was cross-linked 5 min with UV. After prehybridization with 30 ml of solution containing 6 X SSC, 0.5% SDS and 0.05 % casein for 1h at 45°C, the membrane was hybridized with 20 µM FAM or IR dye-labeled oligonucleotide probes (Sup Table 2) in 10 ml of prehybridization solution at 45°C overnight. Finally, the membrane was

washed twice with 2 X SSC, 0,1% SDS for 10 min at RT and twice with 0.2 X SSC, 0.1% SDS for 10 min at 60°C and images captured using a ChemiDoc MP Imaging system (BIO-RAD).

Defense System Detection

Defense system detection was performed using DefenseFinder v1.1.148 and DefenseFinder models v1.2.3 on all RefSeq complete genomes from July 2022.
Domain annotation Defense system domain annotation was done using HHpred49 webserver with the PFAM50 database and standard setting. For PARIS, trigger binding domain and Coiled-coil domain were annotated using the structure information and ESMfold51 structure prediction for AriAB. AriA and AriB tree
PARIS systems detected by DefenseFinder v1.1.1 on RefSeq complete genomes from July 2022 were used to build phylogenetic trees. Only hits to AriA and AriB with a coverage of more than 75% were kept to avoid pseudogenes. Hits were then clustered using Mmseqs252 v13.45111 on AriA and AriAB with an identity threshold of 80% and coverage threshold of 90%. A single representative of each cluster was then used to build the AriA tree, and the cognate AriB was used to build the AriB tree. The AriA tree was built using the AAA15 ATPase domain only and the AriB tree using the DUF4435 domain only. Domains were detected by HMMsearch (from HMMER 3.3.2) on the E. coli B185 PARIS system and a first alignment using Mafft v7.50553 (default parameters) was used to extract the corresponding sequences in other PARIS variants. The final alignment used to make the tree was done using MUSCLE v.5.1.linux6454 using the model super5 and trim using clipkit v1.3.0 with option seq-gap. The DUF4435 tree using M5 Ribonuclease as outgroup was done using selected representative hits from different clades and

aligned using muscle v5.1.linux64 with model super5. All trees were built using Iqtree v2.0.655 using ModelFinder and Ultrafast Bootstrap 1000.

Defense System ATPase AAA15/21 Tree

ATPase AAA15/21 defense system containing detection was done using hmmsearch from HMMER 3.3.2 and Pfam Hidden Markov Model 33.1 on defense systems previously detected. Random selection of 20 proteins from the detection results with a coverage of more than 75% were selected in addition to experimentally validated homologs to build the tree. The ATPase alignment was done using Mafft v7.505 and the tree was built using Iqtree v2.0.6 using ModelFinder and Ultrafast Bootstrap 1000.

Acknowledgements

Thanks to members of the B.W. laboratory for feedback and discussions, Coltran Hophan-Nichols and the cyber security team at Montana State Univeristy for computational support, Dr. Heather Callaway for helpful suggestions for data processing, Dr. Martin Lawrence and Colin Gauvin for maintenance and operation of the Cryo-EM Core Facility at Montana State University. The cryo-EM Facility at Montana State University is supported by NSF 1828765 and the M.J. Murdock Charitable Trust. Research in the Wiedenheft lab is supported by the National Institutes of Health (R35GM134867), the M.J. Murdock Charitable Trust, a young investigator award from Amgen, and the Montana State University Agricultural Experimental Station (USDA NIFA). N.B. is support by an F31 from the NIH (GM153146). A.S-F. is supported by NIH (K99GM147842) and the Burroughs Wellcome Fund (G-1021106.01). A.B.G. is supported by Montana State University's Undergraduate Scholars Program, and by the NIH NIGMS IDeA

program (P20GM103474). Skolkovo Institute of Science and Technology team was supported by the Ministry of Science and Higher Education grant (075-10-2021-114) and RSF grant (22-14-00004). We thank Anatoly Glukhov for sharing a collection of T5 deletion variants, Tinashe Prince Maviza for the help with *in vitro* translation assays, and Alina Demkina for the help with BGI sequencing. Illumina sequencing has been performed at the Skoltech Genomics Core Facility. D.B and F.D are supported by the European Research Council [677823]; European Research Council [101044479]; Agence Nationale de la Recherche [ANR-10-LABX-62-IBEID]. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. Funders had no role in the conceptualization, designing, data collection, analysis, decision to publish, or preparation of the manuscript.

Manuscript References

1. Wein, T. & Sorek, R. Bacterial origins of human cell-autonomous innate immune mechanisms. Nat Rev Immunol 22, 629–638 (2022).

2. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems ▿. J Bacteriol 193, 6039–6056 (2011).

3. Rocha, E. P. C. & Bikard, D. Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? PLOS Biology 20, e3001514 (2022).

4. Rousset, F. et al. Phages and their satellites encode hotspots of antiviral systems. Cell Host & Microbe 30, 740-753.e5 (2022).

5. Studier, F. W. Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. Journal of Molecular Biology 94, 283–295 (1975).

6. Isaev, A. et al. Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. Nucleic Acids Res 48, 5397–5406 (2020).

7. Schiltz, C. J., Lee, A., Partlow, E. A., Hosford, C. J. & Chappie, J. S. Structural characterization of Class 2 OLD family nucleases supports a two-metal catalysis mechanism for cleavage. Nucleic Acids Research 47, 9448–9463 (2019).

8. Schiltz, C. J., Adams, M. C. & Chappie, J. S. The full-length structure of Thermus scotoductus OLD defines the ATP hydrolysis properties and catalytic mechanism of Class 1 OLD family nucleases. Nucleic Acids Research 48, 2762–2776 (2020).

9. Bregegere, F. Bacteriophage P2-λ interference: Inhibition of protein synthesis involves transfer RNA inactivation. Journal of Molecular Biology 90, 459–467 (1974).

10. Wilkinson, M. et al. Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. eLife 5, e22963.

11. Haggård-Ljungquist, E., Barreiro, V., Calendar, R., Kurnit, D. M. & Cheng, H. The P2 phage old gene: sequence, transcription and translational control. Gene 85, 25–33 (1989).

12. Dot, E. W., Thomason, L. C. & Chappie, J. S. Everything OLD is new again: How structural, functional, and bioinformatic advances have redefined a neglected nuclease family. Mol Microbiol 120, 122–140 (2023).

13. Antine, S. P. et al. Structural basis of Gabija anti-phage defence and viral immune evasion. Nature 1–3 (2023) doi:10.1038/s41586-023-06855-2.

14. Cheng, R. et al. A nucleotide-sensing endonuclease from the Gabija bacterial defense system. Nucleic Acids Res 49, 5216–5229 (2021).

15. Mestre, M. R., González-Delgado, A., Gutiérrez-Rus, L. I., Martínez-Abarca, F. & Toro, N. Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. Nucleic Acids Research 48, 12632–12647 (2020).

16. Ernits, K. et al. The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin–antitoxin and related phage defense systems. Proceedings of the National Academy of Sciences 120, e2305393120 (2023).

17. Bryant, P., Pozzati, G. & Elofsson, A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun 13, 1265 (2022).

18. Hopfner, K. Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. Current Opinion in Structural Biology 13, 249–255 (2003).

19. Krishnan, A., Burroughs, A. M., Iyer, L. M. & Aravind, L. Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems. Nucleic Acids Research 48, 10045–10075 (2020).

20. Gut, F. et al. Structural mechanism of endonucleolytic processing of blocked DNA ends and hairpins by Mre11-Rad50. Mol Cell 82, 3513-3522.e6 (2022).

21. Seifert, F. U., Lammens, K., Stoehr, G., Kessler, B. & Hopfner, K. Structural mechanism of ATP-dependent DNA binding and DNA end bridging by eukaryotic Rad50. EMBO J 35, 759–772 (2016).

22. Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. Nat Struct Mol Biol 11, 918–926 (2004).

23. Deep, A. et al. The SMC-family Wadjet complex protects bacteria from plasmid transformation by recognition and cleavage of closed-circular DNA. Molecular Cell 82, 4145-4159.e7 (2022).

24. Weiß, M. et al. The MksG nuclease is the executing part of the bacterial plasmid defense system MksBEFG. Nucleic Acids Research 51, 3288–3306 (2023).

25. Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Research 26, 4205–4213 (1998).

26. Yirmiya, E. et al. Phages overcome bacterial immunity via diverse anti-defence proteins. Nature 1–3 (2023) doi:10.1038/s41586-023-06869-w.

27. Wang, J. et al. Complete genome sequence of bacteriophage T5. Virology 332, 45–65 (2005).

28. Glukhov, A. S., Krutilina, A. I., Kaliman, A. V., Shlyapnikov, M. G. & Ksenzenko, V. N. Bacteriophage T5 Mutants Carrying Deletions in tRNA Gene Region. Mol Biol 52, 1–6 (2018).

29. Jain, I. et al. tRNA anticodon cleavage by target-activated CRISPR-Cas13a effector. 2021.11.10.468108 Preprint at https://doi.org/10.1101/2021.11.10.468108 (2021).

30. Jiang, Y., Meidler, R., Amitsur, M. & Kaufmann, G. Specific interaction between anticodon nuclease and the tRNA(Lys) wobble base. J Mol Biol 305, 377–388 (2001).

31. Azam, A. H. et al. Viruses encode tRNA and anti-retron to evade bacterial immunity. 2023.03.15.532788 Preprint at https://doi.org/10.1101/2023.03.15.532788 (2023).

32. Mogila, I. et al. Ribosomal stalk-captured CARF-RelE ribonuclease inhibits translation following CRISPR signaling. Science 382, 1036–1041 (2023).

33. Bingham, R., Ekunwe, S. I. N., Falk, S., Snyder, L. & Kleanthous, C. The Major Head Protein of Bacteriophage T4 Binds Specifically to Elongation Factor Tu*. Journal of Biological Chemistry 275, 23219–23226 (2000)

34. Yang, J. Y. et al. Degradation of host translational machinery drives tRNA acquisition in viruses. Cell Systems 12, 771-779.e5 (2021).

35. Van den Berg, D. F., van der Steen, B. A., Costa, A. R. & Brouns, S. J. Phage tRNAs evade tRNA-targeting host defenses through anticodon loop mutations. eLife 12, e85183 (2023).

36. Albers, S. & Czech, A. Exploiting tRNAs to Boost Virulence. Life 6, 4 (2016).

37. Zhang, F. et al. Human SAMD9 is a poxvirus-activatable anticodon nuclease inhibiting codon-specific protein synthesis. Science Advances 9, eadh8502 (2023).

38. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345 (2009).

39. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology 152, 36–51 (2005).

40. Rohwer, F. & Azam, F. Detection of DNA Damage in Prokaryotes by Terminal Deoxyribonucleotide Transferase-Mediated dUTP Nick End Labeling. Applied and Environmental Microbiology 66, 1001–1006 (2000).

41. Cockram, C., Thierry, A. & Koszul, R. Generation of gene-level resolution chromosome contact maps in bacteria and archaea. STAR Protoc 2, 100512 (2021).

42. Cournac, A., Marie-Nelly, H., Marbouty, M., Koszul, R. & Mozziconacci, J. Normalization of a chromosomal contact map. BMC Genomics 13, 436 (2012).

43. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 9, 999–1003 (2012).

44. Lioy, V. S. et al. Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. Cell 172, 771-783.e18 (2018).

45. Kurata, T. et al. A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains. Proceedings of the National Academy of Sciences 119, e2102212119 (2022).

46. Sambrook, J. Molecular cloning : a laboratory manual. (Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 2001)

47. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology 13, e1005595 (2017).

48. Tesson, F. et al. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13, 2561 (2022).

49. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33, W244–W248 (2005).
50. Mistry, J. et al. Pfam: The protein families database in 2021. Nucleic Acids Research 49, D412–D419 (2021).

51. Lin, Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science 379, 1123–1130 (2023).

52. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics 35, 2856–2858 (2019).

53. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066 (2002).

54. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797 (2004).

55. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32, 268–274 (2015).

CHAPTER SIX

CONCLUSION

The selective pressure that phage predation exacts on prokaryotes has led to the evolution of diverse defense systems[1,2]. Interrogating these systems not only enhances our understanding of nature but also drives biotechnological innovation. One of the most notable prokaryotic defense systems, CRISPR-Cas, has revolutionized molecular biology as we know it, by allowing for precise, programmable DNA cleavage[3]. This activity has been used at the bench to interrogate gene function, and at the bedside for the treatment of genetic disease. In addition to their roles in DNA engineering, CRISPR Cas systems have been repurposed as molecular diagnostics, offering a promising alternative to current technologies[4–8].

While CRISPR is the most notable phage defense system in biotechnology it will not be the last. The study of the primordial conflict between host and parasite has led to the discovery of dozens of novel defense systems, some of which have already been repurposed as diagnostics and programmable nucleases. In this thesis, I have outlined my work on adapting type III CRISPRs for use in molecular diagnostics and RNA editing in response to the SARS-CoV-2 pandemic. I also showcased my contributions on determining the mechanism by which the immune system, PARIS, recognizes viral antigens and prevents virion replication.

Current Landscape of CRISPR Diagnostics

The SARS-CoV-2 pandemic showcased the need for rapid and sensitive diagnostics. Current technologies were inadequate to diagnose infection promptly to inhibit community spread. In response to this technological deficit, the Wiedenheft lab repurposed a type III

CRISPR system from *Thermus thermophilus* (TtCsm) as a rapid molecular diagnostic. While the first-generation diagnostic was able to detect SARS-CoV-2 genomic RNA in patient samples, it required the use of a nucleic acid preamplification step, which diluted the impact of the technology[5] . In response to this pitfall, we developed a method termed 'Capture and Concentrate' which utilizes magnetic beads conjugated to TtCsm, to concentrate target RNA from large sample volumes. In combination with this technology, we identified and incorporated a new ancillary effector, AaCan2, which doesn't degrade the signaling molecule produced by TtCsm. Both innovations led to an exponential increase in sensitivity, allowing us to directly detect 10^4 RNA target copies per uL in complex mixtures. While this increase in sensitivity is significant, we ultimately want to be able to detect down to 10^2 RNA copies per uL. To inform future development, I developed an *in silico* model of the diagnostic to identify key targets for diagnostic improvement. Modelling indicated that there is not a single step or process in the diagnostic that contributes unequally to overall sensitivity, allowing us to choose steps that are most amenable to future improvements.

In vitro RNA Editing

While type III CRISPRs are unique for their antiviral signaling cascades, they also contain specific single-stranded RNAse activity[9]. In 2023 we reported that we can repurpose this activity for *in vitro* editing of RNA viruses. We used the type III CRISPR from *Streptococcus thermophilus* (SthCsm) as an RNA guided RNA nuclease to make targeted cuts in a recombinant sindbis virus. Using SthCsm in combination with a DNA splint and an RNA ligase we made targeted deletions and introduced new segments of RNA within the sindbis genome. This technology is the first of its kind and allows for the facile manipulation of RNA.

Mechanistic Understanding of Bacterial Antigen Sensing

While the role of antibodies and lymphocyte receptors in antigen sensing is well understood, antigen recognition in prokaryotes is poorly defined. PARIS, is a prokaryotic immune system that senses viral proteins and attenuates viral infection by inducing cell suicide. Interestingly, PARIS is often encoded on integrated prophages, further supporting the 'guns for hire' paradigm first suggested by Koonin in 2015[10]. In our work, which was recently submitted to *Nature,* we used a combination of Cryo-EM, biochemistry and microbiology to probe the requirements for antigen recognition by PARIS and determine how an abortive infection phenotype is elicited.

PARIS is a two-component immune system comprised of the TOPRIM nuclease AriB and the ABC ATPase AriA. Using gel filtration chromatography and cryo-electron microscopy, we determined that six molecules of AriA and three molecules AriB assemble into a 435 kDa supramolecular complex. Interestingly, AriB is not a functional nuclease when in complex, but AriA is still an active ATPase when bound to AriB.

Moreover, we demonstrated that ATPase activities are necessary for antigen recognition, through immunoprecipitation assays with AriA mutants and viral 'triggers'. Finally, we showed that PARIS kills the host upon recognition of a viral trigger, by releasing an active AriB which will cleave the host lysine tRNA. However, some viruses have been able to subvert PARIS-mediated immunity by encoding their own, non-cleavable, lysine tRNAs. Collectively, our work demonstrates that prokaryotes contain antigen-sensing systems that rely on ATPase activities for antigen recognition.

Outlook

While CRISPR Cas systems have been commonly repurposed to treat genetic diseases, they offer a flexible platform for molecular diagnostics and RNA editing technologies. The importance of both applications was proven in 2020 with the emergence of the SARS-CoV-2. While this was the first pandemic to cause global medical, social, and economic hardships in the 21st century, it will likely not be the last. To better prepare for future pandemics we must continue to develop the technologies that were created in response to the 2020 pandemic.

Type III CRISPR Cas systems are the best-suited class of CRISPRs for direct RNA detection and RNA editing as target binding does not elicit a non-specific RNAse activity like other RNA targeting systems[8,11]. Additionally, type V and VI CRISPRs do not contain the necessary activities to generate a measurable signal in a rapid timeframe for diagnostics[12,13]. Further work to improve the diagnostic may involve identifying the biochemical activities of novel type III effectors. Makarova et al have already identified putative type III CRISPR effectors comprised of a signal recognition domain, an RNAse and putative polymerase which may serve as a signal amplification mechanism in nature[14]. Incorporation of these effectors may provide a significant increase in diagnostic sensitivity.

Type III CRISPRs are also poised to revolutionize genetic disease treatment through the editing of erroneous transcripts associated with disease states. The Wiedenheft lab is already working on an *in vivo* editing approach to edit mammalian transcripts with SthCsm. If RNA editing techniques follow the same timeline as CRISPR based DNA editing technologies, we should expect to see this technology translated from the bench to the bedside within the next decade.

References

1. Keen, E. C. A century of phage research: bacteriophages and the shaping of modern biology. BioEssays News Rev. Mol. Cell. Dev. Biol. 37, 6–9 (2015).

2. Millman, A. et al. An expanded arsenal of immune systems that protect bacteria from phages. Cell Host Microbe 30, 1556-1569.e5 (2022).

3. Jinek, M. et al. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816–821 (2012).

4. Nemudraia, A. et al. Sequence-specific capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. Nat. Commun. 13, 7762 (2022).

5. Santiago-Frangos, A. et al. Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. Cell Rep. Med. 2, 100319 (2021).

6. Steens, J. A. et al. SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. Nat. Commun. 12, 5033 (2021).

7. Broughton, J. P. et al. CRISPR–Cas12-based detection of SARS-CoV-2. Nat. Biotechnol. 38, 870–874 (2020).

8. Abudayyeh, O. O. et al. RNA targeting with CRISPR–Cas13. Nature 550, 280–284 (2017).

9. Niewoehner, O. et al. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. Nature 548, 543–548 (2017).

10. Koonin, E. V., Dolja, V. V. & Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479, 2–25 (2015).

11. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 360, 436–439 (2018).

12. Huyke, D. A. et al. Enzyme Kinetics and Detector Sensitivity Determine Limits of Detection of Amplification-Free CRISPR-Cas12 and CRISPR-Cas13 Diagnostics. Anal. Chem. 94, 9826–9834 (2022).

13. Ramachandran, A. & Santiago, J. G. CRISPR Enzyme Kinetics for Molecular Diagnostics. Anal. Chem. 93, 7456–7464 (2021).

14. Makarova, K. S. et al. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. Nucleic Acids Res. 48, 8828–8847 (2020).

APPENDICES

APPENDIX A

IN SILICO MODELLING OF TYPE III CRISPR DIAGNOSTIC

To improve type III CRISPR diagnostic sensitivity, we hypothesized that an *in silico* model would allow us to identify steps in the diagnostic that significantly contribute to sensitivity. To test this hypothesis, I built a mathematical model that used a system of differential equations to relate concentrations of cleaved reporter RNA to initial target RNA concentrations(Fig A1). I then populated the model with rate constants for each of the kinetic 'checkpoints' of the diagnostic. While we were able to use some constants from the literature and data collected in the lab, we had to estimate the on and off rates of target, cOA, and reporter binding.

To determine the effect that each kinetic 'checkpoint' had on diagnostic sensitivity, I first defined theoretical limit of detection. To do so, I used a definition from the Santiago group, which proposes that the cleavage of 1% of total reporter RNA in the reaction is an acceptable metric for LoD. To determine the effect that each kinetic checkpoint had on LoD, I recorded what concentrations of target RNA were needed to cleave 1% of reporter RNA, while varying a specific kinetic parameter over several orders of magnitude. Importantly, while I varied one parameter, I kept all other kinetic parameters constant. I then iterated through every kinetic step in the diagnostic.

Surprisingly, my results demonstrated that each step in the diagnostic contributes equally to overall sensitivity (Fig A2A-E). In other words, there is not a singular rate-limiting step. When I increased the binding affinity or the catalytic rate of any diagnostic process there was a proportional increase in sensitivity. While surprising, this result allows us to target steps that are most amenable to improvement.

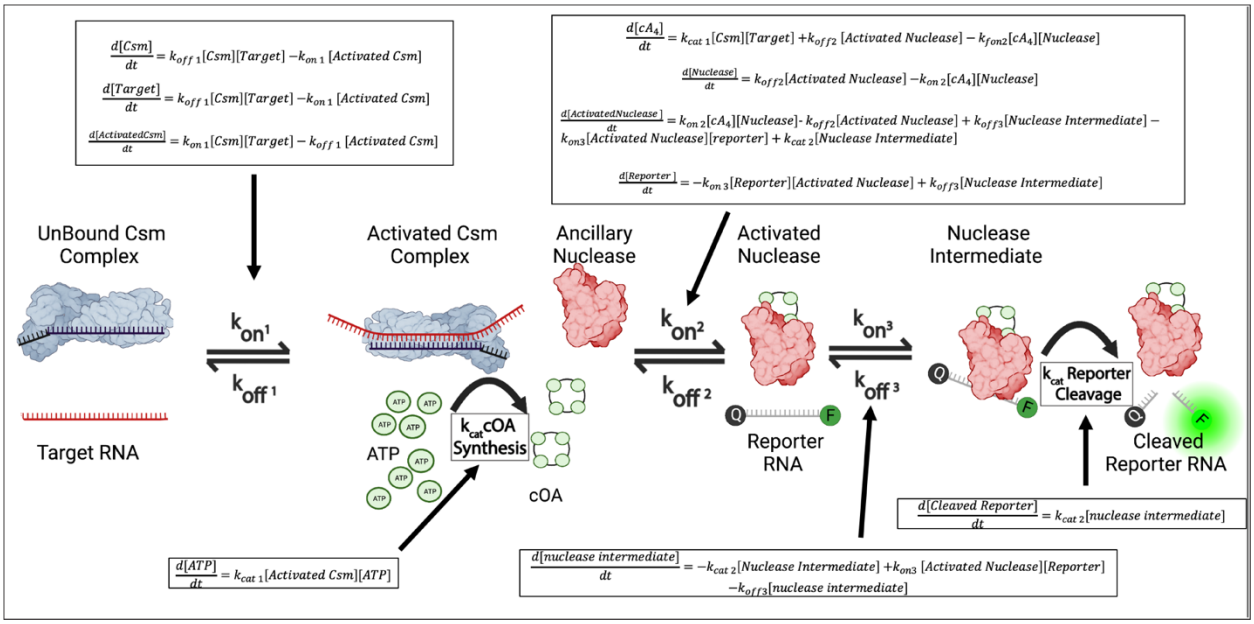


Figure A1: Mathematical model of type III CRISPR based diagnostic. Mathematical model of type III CRISPR diagnostic utilizes differential equations to relate concentrations of cleaved reporter RNA to concentrations of target RNA.

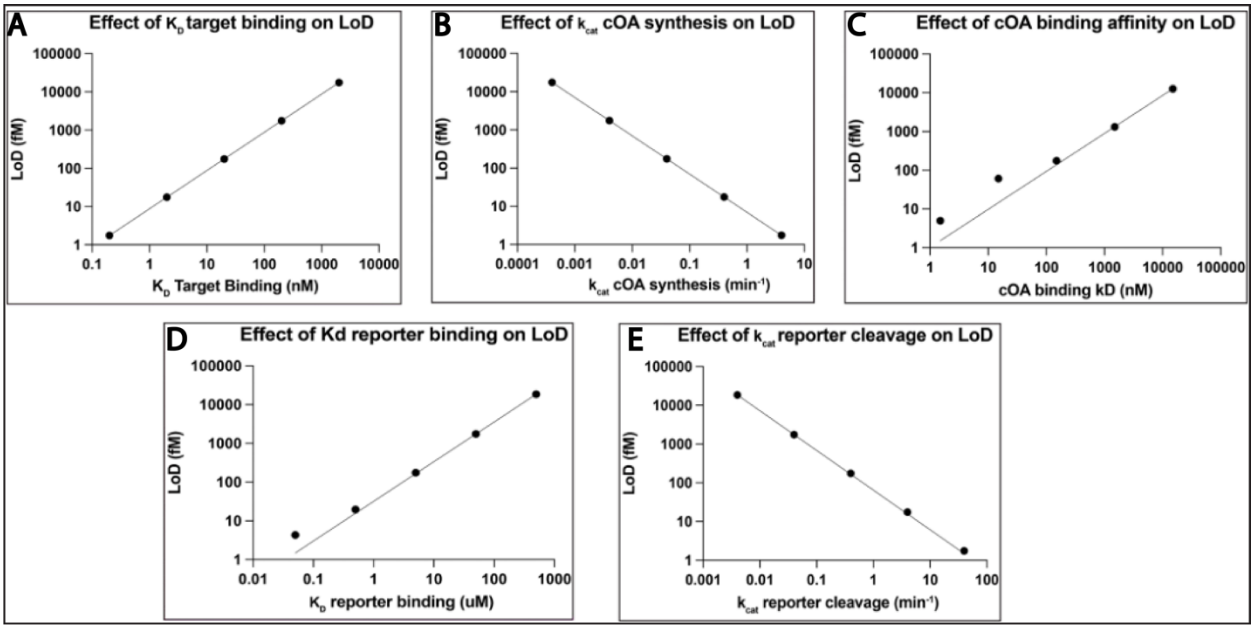


Figure A2: All steps of diagnostic contribute equally to theoretical LoD: (A-E) Plots displaying theoretical LoD as a function of respective kinetic parameter. For all parameters for, there is a linear relationship between binding affinity (A,C,D) or catalytic rate (B,E) and diagnostic sensitivity.

References

1. Athukoralage, J. S. & White, M. F. Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. *Rna* **27**, 855–867 (2021).

2. Ramachandran, A. & Santiago, J. G. CRISPR Enzyme Kinetics for Molecular Diagnostics. *Anal. Chem.* **93**, 7456–7464 (2021).

APPENDIX B

PARIS MEDIATED DEPLETION OF INTRACELLULAR ATP

We initially hypothesized that PARIS-mediated defense would involve a multifaceted approach, where the nuclease AriB would cleave host nucleic acids, while AriA would increase deplete cellular ATP. To test this hypothesis, we performed some back-of-the-envelope calculations using the AriA ATPase rates I determined in Chapter 2. I wanted to determine how long it would take a single paris complex deplete all intracellular ATP assuming a no regeneration of ATP pools.

To perform this calculation, I first assumed that there is a fixed pool of ATP at 9.6 mM in an E coli cell. Assuming that the average E coli cell has a volume of $1um^3$, we can calculate that 1nM of PARIS within the cell is equivalent to one complex. Then plugging in the ATPase rates I determined in chapter 2, I determined that in the absence of a viral trigger, it would take 1nM of PARIS 333 days to consume all intracellular ATP. This timeframe is even longer for ‘activated’ PARIS which would take 1300 days to consume all intracellular ATP.

Even if we assume that PARIS is at 1uM within the cell, it would still take three to 10 days to consume all ATP in the cell depending on the presence of viral antigen. Given that the average phage replication time ranges from 30-60 minutes, ATP hydrolysis is not a major contributor to the abortive infection phenotype.

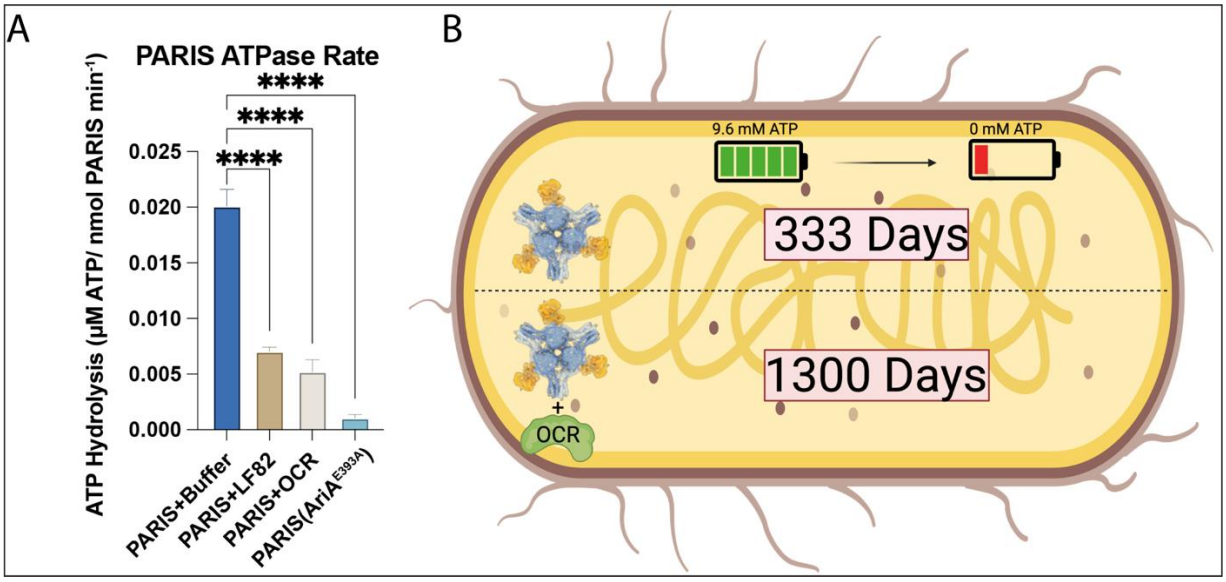


Figure 1: PARIS ATPase activities do no not deplete Intracellular ATP in response to viral infection. (A) The rate of PARIS-mediated ATP hydrolysis at 32 minutes. The presence of virally encoded triggers decreases PARIS ATPase activity. Pair-wise comparisons of ATP hydrolysis rates of each reaction group the PARIS-only group (n=3) were compared with a post hoc Dunnett's test. ****=$p<0.0001$. (B) Estimated time that 1 nM of PARIS would take to deplete all intracellular ATP in its most active state (top), and antigen bound state (bottom).

REFERENCES CITED

Abudayyeh, O. O. & Gootenberg, J. S. CRISPR diagnostics. *Science* **372**, 914–915 (2021).

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

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

Ackerman, C. M. *et al.* Massively multiplexed nucleic acid detection with Cas13. *Nature* **582**, 277–282 (2020).

Altae-Tran, H. *et al.* The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).

Arizti-Sanz, J. *et al.* Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2. *Nature Communications* **11**, (2020).

Athukoralage, J. S. & White, M. F. Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. *Rna* **27**, 855–867 (2021).

Azhar, M. *et al.* Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. *Biosensors and Bioelectronics* **183**, (2021).

Balderston, S. *et al.* Discrimination of single-point mutations in unamplified genomic DNA via Cas9 immobilized on a graphene field-effect transistor. *Nature Biomedical Engineering* **5**, 713–725 (2021).

Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).

Basu, A. S. Digital Assays Part I: Partitioning Statistics and Digital PCR. *SLAS Technology* **22**, 369–386 (2017).

Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology* **5**, 593–599 (2009).

Béraud, A. *et al.* Graphene field-effect transistors as bioanalytical sensors: Design, operation and performance. *Analyst* **146**, 403–428 (2021).

Broughton, J. P. *et al.* CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology* (2020) doi:10.1038/s41587-020-0513-4.

Broughton, J. P. *et al.* CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology* **38**, 870–874 (2020).

Bruch, R. *et al.* CRISPR/Cas13a-Powered Electrochemical Microfluidic Biosensor for Nucleic Acid Amplification-Free miRNA Diagnostics. *Advanced Materials* **31**, (2019).

Bruch, R. *et al.* CRISPR-powered electrochemical microfluidic multiplexed biosensor for target amplification-free miRNA diagnostics. *Biosensors and Bioelectronics* **177**, (2021).

Bustin, S. A. *et al.* The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* **55**, 611–622 (2009).

Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).

Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).

Chuai, G. *et al.* DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. *Genome Biology* **19**, 1–18 (2018).

Dai, Y. *et al.* Exploring the Trans-Cleavage Activity of CRISPR-Cas12a (cpf1) for the Development of a Universal Electrochemical Biosensor. *Angewandte Chemie - International Edition* **58**, 17399–17405 (2019).

Dao Thi, V. L. *et al.* A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Science translational medicine* **7075**, 1–20 (2020).

Daum, L. T. & Fischer, G. W. Ion torrent genomic sequencing. (2013).

de Puig, H. *et al.* Minimally instrumented SHERLOCK (miSHERLOCK) for CRISPR-based point-of-care diagnosis of SARS-CoV-2 and emerging variants. *Science Advances* **7**, 23–26 (2021).

Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

Deng, R., Zhang, K. & Li, J. Isothermal Amplification for MicroRNA Detection: From the Test Tube to the Cell. *Accounts of Chemical Research* **50**, 1059–1068 (2017).

East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270–273 (2016).

Ferrario, A., Scaramuzza, M., Pasqualotto, E., De Toni, A. & Paccagnella, A. Development of a Disposable Gold Electrodes-Based Sensor for Electrochemical Measurements of cDNA Hybridization. *Procedia Chemistry* **6**, 36–45 (2012).

Fozouni, P. *et al.* Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell* 1–11 (2020) doi:10.1016/j.cell.2020.12.001.

Garcia-Doval, C. *et al.* Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nature Communications* **11**, 1–9 (2020).

Gill, P. & Ghaemi, A. Nucleic Acid Isothermal Amplification Technologies—A Review. *Nucleosides, Nucleotides and Nucleic Acids* **27**, 224–243 (2008).

Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science* **360**, 439–444 (2018).

Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science* **360**, 439–444 (2018).

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

Grieshaber, D., MacKenzie, R., Vörös, J. & Reimhult, E. Electrochemical biosensors - Sensor principles and architectures. *Sensors* **8**, 1400–1458 (2008).

Grüschow, S., Adamson, C. S. & White, M. F. Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector. *Nucleic Acids Research* **49**, 13122–13134 (2021).

Hajian, R. *et al.* Detection of unamplified target genes via CRISPR–Cas9 immobilized on a graphene field-effect transistor. *Nature Biomedical Engineering* **3**, 427–437 (2019).

Hale, C. R. *et al.* Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Molecular Cell* **45**, 292–302 (2012).

Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945–956 (2009).

Hayat, A., Catanante, G. & Marty, J. L. Current trends in nanomaterial-based amperometric biosensors. *Sensors (Switzerland)* **14**, 23439–23461 (2014).

He, R. *et al.* Pyrococcus furiosus Argonaute-mediated nucleic acid detection. *Chemical Communications* **55**, 13219–13222 (2019).

Hegge, J. W. *et al.* DNA-guided DNA cleavage at moderate temperatures by Clostridium butyricum Argonaute. *Nucleic Acids Research* **47**, 5809–5821 (2019).

Hegge, J. W., Swarts, D. C. & Van Der Oost, J. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nature Reviews Microbiology* **16**, 5–11 (2018).

Higuchi, R., Dollinger, G., Walsh, P. S. & Griffith, R. Simultaneous Amplification and Detection of Specific DNA Sequences. *Biotechnology* **10**, 413–417 (1992).

Holland, P. M., Abramson, R. D., Watson, R. & Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5'—-3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences* **88**, 7276–7280 (1991).

Huang, M., Zhou, X., Wang, H. & Xing, D. Clustered Regularly Interspaced Short Palindromic Repeats/Cas9 Triggered Isothermal Amplification for Site-Specific Nucleic Acid Detection. *Analytical Chemistry* **90**, 2193–2200 (2018).

Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

Huyke, D. A. *et al.* Enzyme Kinetics and Detector Sensitivity Determine Limits of Detection of Amplification-Free CRISPR-Cas12 and CRISPR-Cas13 Diagnostics. *Analytical Chemistry* **94**, 9826–9834 (2022).

Jain, I. *et al.* tRNA anticodon cleavage by target-activated CRISPR-Cas13a effector. (2021) doi:https://doi.org/10.1101/2021.11.10.468108.

Jia, N., Jones, R., Yang, G., Ouerfelli, O. & Patel, D. J. CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Molecular Cell* **75**, 944-956.e6 (2019).

Jiao, C. *et al.* Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. *Science* **7106**, eabe7106 (2021).

Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

Jolly, S. M. *et al.* Thermus thermophilus Argonaute Functions in the Completion of DNA Replication. *Cell* **182**, 1545-1559.e18 (2020).

Joung, J. *et al.* Detection of SARS-CoV-2 with SHERLOCK One-Pot Testing. *New England Journal of Medicine* **382**, NEJMc2026172 (2020).

Jung, C. *et al.* Massively Parallel Biophysical Analysis of CRISPR- Cas Complexes on Next Generation Article Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* **170**, 35-47.e13 (2017).

Kaminski, M. M., Abudayyeh, O. O., Gootenberg, J. S., Zhang, F. & Collins, J. J. CRISPR-based diagnostics. *Nature Biomedical Engineering* **5**, 643–656 (2021).

Karvelis, T. *et al.* Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357**, 605–609 (2017).

Keen, E. C. A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* **37**, 6–9 (2015).

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O. & Zhang, F. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nature Protocols* **14**, 2986–3012 (2019).

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O. & Zhang, F. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nature Protocols* **14**, 2986–3012 (2019).

Kerppola, T. K. Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annual Review of Biophysics* **37**, 465–487 (2008).

Kim, H. K. *et al.* SpCas9 activity prediction by DeepSpCas9, a deep learning–based model with high generalization performance. *Science Advances* **5**, 1–10 (2019).

Kim, S., Ji, S. & Koh, H. R. Crispr as a diagnostic tool. *Biomolecules* **11**, (2021).

Knott, G. J. & Lapinaite, A. CRISPR Gets Its Origin Story. *The CRISPR Journal* **4**, 631–633 (2021).

Koczula, K. M. & Gallotta, A. Lateral flow assays. *Essays in Biochemistry* **60**, 111–120 (2016).

Koonin, E. V., Dolja, V. V. & Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **479**, 2–25 (2015).

Koopal, B. *et al.* Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. *Cell* 1–16 (2022) doi:10.1016/j.cell.2022.03.012.

Kostyusheva, A. *et al.* CRISPR-Cas systems for diagnosing infectious diseases. *Methods* (2021) doi:10.1016/j.ymeth.2021.04.007.

Kumar, M. *et al.* FnCas9 based CRISPR diagnostic for rapid and accurate detection of major SARS-CoV2 variants on a paper strip. *eLife* **10**, 1–25 (2021).

Kuzmenko, A. *et al.* DNA targeting and interference by a bacterial Argonaute nuclease. *Nature* **587**, 632–637 (2020).

Kuzmenko, A., Yudin, D., Ryazansky, S., Kulbachinskiy, A. & Aravin, A. A. Programmable DNA cleavage by Ago nucleases from mesophilic bacteria Clostridium butyricum and Limnothrix rosea. *Nucleic Acids Research* **47**, 5822–5836 (2019).

Lau, R. K. *et al.* Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Molecular Cell* **77**, 723-733.e6 (2020).

Le Ru, E. C., Meyer, M. & Etchegoin, P. G. Proof of single-molecule sensitivity in Surface Enhanced Raman Scattering (SERS) by means of a two-analyte technique. *Journal of Physical Chemistry B* **110**, 1944–1948 (2006).

Lee, D. H. & Schleif, R. F. In vivo DNA loops in araCBAD: size limits and helical repeat. *Proceedings of the National Academy of Sciences* **86**, 476–480 (1989).

Li, L. *et al.* HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. *ACS Synthetic Biology* **8**, 2228–2237 (2019).

Li, S.-Y. *et al.* CRISPR-Cas12a-assisted nucleic acid detection. *Cell Discovery* **4**, 20 (2018).

Lisitskaya, L., Aravin, A. A. & Kulbachinskiy, A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nature Communications* **9**, 1–12 (2018).

Liu, T. Y. *et al.* Accelerated RNA detection using tandem CRISPR nucleases. *Nature Chemical Biology* **17**, 982–988 (2021).

Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annual Review of Virology* **7**, 371–384 (2020).

Ludwig, K. U. *et al.* LAMP-Seq enables sensitive, multiplexed COVID-19 diagnostics using molecular barcoding. *Nature Biotechnology* **39**, 1556–1562 (2021).

Makarova, K. S. *et al.* Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Research* 1–20 (2020) doi:10.1093/nar/gkaa635.

Makarova, K. S. *et al.* Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Research* **48**, 8828–8847 (2020).

Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **18**, 67–83 (2020).

Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **18**, 67–83 (2020).

Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, V. E. 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 (2009).

Mammoth Biosciences, Inc. SARS-CoV-2 RNA DETECTR Reagent Kit. *FDA* (2020).

Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).

McMahon, S. A. *et al.* Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nature Communications* **11**, (2020).

Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host & Microbe* **30**, 1556-1569.e5 (2022).

Miyoshi, T., Ito, K., Murakami, R. & Uchiumi, T. Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nature Communications* **7**, (2016).

Mullis, K. B. The Polymerase Chain Reaction (Nobel Lecture). *Angewandte Chemie International Edition in English* **33**, 1209–1213 (1994).

Mullis, K. B. *et al.* PROCESS FOR AMPLIFYING, DETECTING, AND/OR-CLONING NUCLEIC ACID SEQUENCES. (1987).

Nemudraia, A. *et al.* Sequence-specific capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. *Research square* (2022) doi:10.21203/rs.3.rs-1466718/v1.

Nemudraia, A. *et al.* Sequence-specific capture and concentration of viral RNA by type III CRISPR system enhances diagnostic. *Nature Communications* **13**, 7762 (2022).

Nguyen, P. Q. *et al.* Wearable materials with embedded synthetic biology sensors for biomolecule detection. *Nature Biotechnology* (2021) doi:10.1038/s41587-021-00950-3.

Nie, S. & Emory, S. R. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering. *Science* **275**, 1102–1106 (1997).

Niewoehner, O. *et al.* Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).

Niewoehner, O. *et al.* Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).

Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic acids research* **28**, E63 (2000).

Nyren, P. METHOD OF SEQUENCING DNA BASED ON THE DETECTION OF THE RELEASE OF PYROPHOSPHATE AND ENZYMATIC NUCLEOTIDE DEGRADATION. (2001).

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial Argonaute Samples the Transcriptome to Identify Foreign DNA. *Molecular Cell* **51**, 594–605 (2013).

Pang, Y. *et al.* CRISPR-cas12a mediated SERS lateral flow assay for amplification-free detection of double-stranded DNA and single-base mutation. *Chemical Engineering Journal* **429**, 132109 (2022).

Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165**, 1255–1266 (2016).

Peto, L. *et al.* Diagnosis of SARS-CoV-2 infection with LamPORE, a high-throughput platform combining loop-mediated isothermal amplification and nanopore sequencing. *Journal of Clinical Microbiology* **59**, (2021).

Piepenburg, O., Williams, C. H., Stemple, D. L. & Armes, N. A. DNA detection using recombination proteins. *PLoS Biology* **4**, 1115–1121 (2006).

Ptasinska, A. *et al.* Diagnostic accuracy of loop-mediated isothermal amplification coupled to nanopore sequencing (LamPORE) for the detection of SARS-CoV-2 infection at scale in symptomatic and asymptomatic populations. *Clinical Microbiology and Infection* **27**, 1348.e1-1348.e7 (2021).

Quan, J. *et al.* FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nucleic acids research* **47**, e83 (2019).

Ramachandran, A. *et al.* Electric field-driven microfluidics for rapid CRISPR-based diagnostics and its application to detection of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 29518–29525 (2020).

Ramachandran, A. & Santiago, J. G. CRISPR Enzyme Kinetics for Molecular Diagnostics. *Analytical Chemistry* **93**, 7456–7464 (2021).

Ramachandran, A. & Santiago, J. G. CRISPR Enzyme Kinetics for Molecular Diagnostics. *Analytical Chemistry* **93**, 7456–7464 (2021).

Richards, F. M. ON THE ENZYMIC ACTIVITY OF SUBTILISIN-MODIFIED RIBONUCLEASE. *Proceedings of the National Academy of Sciences* **44**, 162–166 (1958).

Rissin, D. M. & Walt, D. R. Digital concentration readout of single enzyme molecules using femtoliter arrays and poisson statistics. *Nano Letters* **6**, 520–523 (2006).

Rogacs, A., Marshall, L. A. & Santiago, J. G. Purification of nucleic acids using isotachophoresis. *Journal of Chromatography A* **1335**, 105–120 (2014).

Ronaghi, M., Uhlén, M. & Nyrén, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).

Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. *Nature Microbiology* **4**, 656–662 (2019).

Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *eLife* **7**, e36734 (2018).

Rouillon, C. *et al.* SAVED by a toxin: Structure and function of the CRISPR Lon protease. *bioRxiv* (2021) doi:doi.org/10.1101/2021.12.06.471393.

Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host & Microbe* **30**, 740-753.e5 (2022).

Rousset, F. *et al.* A conserved family of immune effectors cleaves cellular ATP upon viral infection. *Cell* **186**, 3619-3631.e13 (2023).

Saiki, R. K. *et al.* Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).

Samai, P. *et al.* Co-transcriptional DNA and RNA cleavage during type III CRISPR-cas immunity. *Cell* **161**, 1164–1174 (2015).

Santiago-Frangos, A., Buyukyoruk, M., Wiegand, T., Krishna, P. & Wiedenheft, B. Distribution and phasing of sequence motifs that facilitate CRISPR adaptation. *Current Biology* 1–10 (2021) doi:10.1016/j.cub.2021.05.068.

Santiago-Frangos, A. *et al.* Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. *Cell Reports Medicine* **2**, 100319 (2021).

Santiago-Frangos, A. *et al.* Intrinsic signal amplification by type III CRISPR-Cas systems provides a sequence-specific SARS-CoV-2 diagnostic. *Cell Reports Medicine* **2**, 100319 (2021).

Sha, Y. *et al.* Cascade CRISPR/cas enables amplification-free microRNA sensing with fM-sensitivity and single-base-specificity. *Chemical Communications* **57**, 247–250 (2021).

Shen, J. *et al.* Sensitive detection of a bacterial pathogen using allosteric probe-initiated catalysis and CRISPR-Cas13a amplification reaction. *Nature Communications* **11**, 1–10 (2020).

Sherlock Biosciences, Inc. Sherlock CRISPR SARS-CoV-2 kit. *FDA* (2021).

Shi, K. *et al.* A CRISPR-Cas autocatalysis-driven feedback amplification network for supersensitive DNA diagnostics. *Science Advances* **7**, 1–10 (2021).

Shinoda, H. *et al.* Amplification-free RNA detection with CRISPR–Cas13. *Communications Biology* **4**, 476 (2021).

Sridhara, S., Goswami, H. N., Whyms, C., Dennis, J. H. & Li, H. Virus detection via programmable Type III-A CRISPR-Cas systems. *Nature Communications* **12**, 1–10 (2021).

Staals, R. H. J. *et al.* RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Molecular Cell* **56**, 518–530 (2014).

Steens, J. A. *et al.* SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. *Nature Communications* **12**, 1–12 (2021).

Steens, J. A. *et al.* SCOPE enables type III CRISPR-Cas diagnostics using flexible targeting and stringent CARF ribonuclease activation. *Nature Communications* **12**, 5033 (2021).

Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7702–7707 (2009).

Stüder, F., Petit, J. L., Engelen, S. & Mendoza-Parra, M. A. Real-time SARS-CoV-2 diagnostic and variants tracking over multiple candidates using nanopore DNA sequencing. *Scientific Reports* **11**, 1–11 (2021).

Swarts, D. C. *et al.* The evolutionary journey of Argonaute proteins. *Nature Structural & Molecular Biology* **21**, 743–753 (2014).

Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014).

Swarts, D. C. *et al.* Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Molecular Cell* **65**, 985-998.e6 (2017).

Tamulaitis, G. *et al.* Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell* **56**, 506–517 (2014).

Tanner, N. A., Zhang, Y. & Evans, T. C. Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes. *BioTechniques* **58**, 59–68 (2015).

Tian, T. *et al.* An Ultralocalized Cas13a Assay Enables Universal and Nucleic Acid Amplification-Free Single-Molecule RNA Diagnostics. *ACS Nano* **15**, 1167–1178 (2021).

Tomita, N., Mori, Y., Kanda, H. & Notomi, T. Loop-mediated isothermal amplification (LAMP) of gene sequences and simple visual detection of products. *Nature Protocols* **3**, 877–882 (2008).

Turner, A. How Many People Have Smartphones Worldwide (Apr 2022).

Wang, F. *et al.* PfAgo-based detection of SARS-CoV-2. *Biosensors and Bioelectronics* **177**, 112932 (2021).

Wang, K. *et al.* Facile synthesis of high-performance SiO2@Au core-shell nanoparticles with high SERS activity. *RSC Advances* **8**, 30825–30831 (2018).

Wang, T., Liu, O., Sun, H.-H., Yin, B.-C. & Ye, B.-C. Analytical Methods An RNA-Guided Cas9 Nickase-Based Method for Universal Isothermal DNAAmplification. doi:10.1002/ange.201901292.

Wang, T., Liu, Y., Sun, H. H., Yin, B. C. & Ye, B. C. An RNA-Guided Cas9 Nickase-Based Method for Universal Isothermal DNA Amplification. *Angewandte Chemie - International Edition* **58**, 5382–5386 (2019).

Xiao, R. *et al.* Portable and multiplexed lateral flow immunoassay reader based on SERS for highly sensitive point-of-care testing. *Biosensors and Bioelectronics* **168**, 112524 (2020).

Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nature Methods* **16**, 687–694 (2019).

Yirmiya, E. *et al.* Phages overcome bacterial immunity via diverse anti-defence proteins. *Nature* **625**, 352–359 (2024).

Yue, H. *et al.* Droplet Cas12a Assay Enables DNA Quantification from Unamplified Samples at the Single-Molecule Level. *Nano Letters* **21**, 4643–4653 (2021).

Zaremba, M. *et al.* Sir2-domain associated short prokaryotic Argonautes provide defence against invading mobile genetic elements through NAD+ depletion. *bioRxiv* (2021) doi:https://doi.org/10.1101/2021.12.14.472599.

Zeng, Z. *et al.* A short prokaryotic argonaute cooperates with membrane effector to confer antiviral defense. *bioRxiv* (2021) doi:https://doi.org/10.1101/2021.12.09.471704.

Zhang, D. *et al.* CRISPR/Cas12a-Mediated Interfacial Cleaving of Hairpin DNA Reporter for Electrochemical Nucleic Acid Sensing. *ACS Sensors* **5**, 557–562 (2020).

Zhang, Y. *et al.* Paired Design of dCas9 as a Systematic Platform for the Detection of Featured Nucleic Acid Sequences in Pathogenic Strains. *ACS Synthetic Biology* **6**, 211–216 (2017).

Zhou, W. *et al.* A CRISPR–Cas9-triggered strand displacement amplification method for ultrasensitive DNA detection. *Nature Communications* **9**, 1–11 (2018).