MECHANISMS OF RNA-TARGETING CRISPR SYSTEMS AND THEIR APPLICATIONS FOR RNA EDITING

by

Joseph Edward Nichols

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

of

Master of Science

in

Microbiology & Immunology

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2022

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my family. Without their continual love and support, I would have struggled to accomplish anything through my graduate training. I am eternally grateful to my parents, Anne and Jim, and my brothers Charlie and Riley, for the role they have played in supporting my pursuits in science and beyond.

Through my graduate training, I have been supported by the incredible post-doctoral research fellows, Dr. Andrew Santiago-Frangos, Dr. Anna Nemudraia, and Dr. Artem Nemudryi as well as Dr. Royce Wilkinson. These are four of the most intelligent and hardworking people I have ever had the pleasure to know, and they serve as continual role models to me. Their guidance has been critical to my development as a scientist, and they have taught me innumerous skills in laboratory research, in critical reasoning, and how to present the story behind scientific findings to an audience.

The environment of the Wiedenheft has also played a major role in supporting me through my studies. Much of this atmosphere is thanks to the supportive and inquisitive culture across the team of other graduate students, research technicians, and undergraduate students. I have relished the opportunity to ponder the mysteries of microbiology with my lab mates both at the bench and over a Hyalite Pizza at Colombo's pizzeria.

Finally, I am deeply grateful to my mentor, Dr. Blake Wiedenheft. His passionate curiosity has completely changed how I think about scientific questions and go about answering them. He has pushed me to ask fundamental questions about biology and pursue projects that can have wide reaching and profound impacts. The skills I have learned by working under him will prove to be invaluable as I build my career in biomedical science as a physician-scientist.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES CONTINUED

ABSTRACT

Genetic modification studies are central to understanding gene function and are the bedrock of molecular biology. The development of novel, CRISPR-based technologies for genome engineering in the last decade has revolutionized nearly every field of biology by simplifying the process of editing DNA genomes. In contrast, there are currently no comparable tools for editing RNA. Our goal is to develop facile CRISPR-based RNA editing methods that will transform our understanding of RNA metabolism, viruses and the repair pathways that govern RNA biology.

I didn't initially come to MSU intending to study SARS-CoV-2, but the growing importance of this topic, combined with unanticipated intersections with my interest in CRISPRs, ultimately lead to several projects in this area. While participating in genomic surveillance, we identified a naturally occurring deletion within ORF7a, a viral accessory protein. We determined that this deletion results in the loss of function of ORF7a, limiting the virus' ability to evade host interferon responses, and reduced viral fitness.

My focus then moved to Type-III CRISPR systems. While CRISPR has become synonymous with genome engineering, these systems naturally evolved in prokaryotes as an adaptive immune system against bacteriophages. Type-III CRISPR systems are unique, as they are one of two groups of CRISPR systems to target RNA rather than DNA. To develop type III systems for editing RNA, we designed and purified a series of type III complexes and showed that these systems function as programable nucleases. We then adapted a method for targeted RNA repair *in vitro* following cleavage and demonstrate that this approach results in edited RNA.

In addition to cleaving the RNA target, target recognition by type III CRISPR systems also activates a polymerase domain that generates signaling molecules that activate ancillary CRISPR nucleases. Working with several members of the team, I set out to determine substrate preferences for each ancillary nuclease in *Thermus thermophilus*. We expected that activating these immune components would result in dramatic changes in bacterial growth kinetics. However, my experiments failed to identify a reliable phenotype, suggesting that this expression system is not a faithful representation of Type-III immunity.

Correspondence: j.e.nichols@outlook.com

CHAPTER ONE

PREFACE AND OUTLINE

Viruses are the most abundant biologic agents on earth, which have dramatically shaped the evolutionary history of all life on earth. Viruses straddle the border of living and non-living, as they rely upon the cellular machinery of host cells to build progeny viruses, nevertheless, they constantly evolve to better exploit their host cells. In response, host cells have evolved immune strategies to defend themselves from viruses. In humans, immunity to viruses is driven by the innate and adaptive arms of the immune system which use highly specialized cell types to kill invading microbes and infected cells. Like humans, prokaryotes also get infected with viruses and must defend themselves.

Bacteria and archaea have evolved multiple mechanisms to restrict invading foreign nucleic acids, including systems based on genome acquired Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs). CRISPR systems, which are present in ~40% of bacteria and 90% of archaea, are sequence specific antiviral defense mechanisms that protect host cells from infections[1,2]. The role of CRISPR systems in prokaryotic immune defense was first discovered in 2005[3], but the remarkable potential of CRISPR systems for genome engineering was not realized until the landmark 2012 paper from the labs of Jennifer Doudna and Emmanuel Charpentier, *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*[4]. This was the first instance in which a CRISPR effector, Cas9, was reprogrammed by a single guide RNA (sgRNA) to catalyze a double stranded cut within a specified DNA target. This paper initiated a revolution in the field of genome engineering. Previously, there were no

tools that could specifically cleave a target DNA sequence and easily be reprogrammed by changing a guide RNA. Previous tools for genome engineering, namely Zinc Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs), could target specific DNA sequences, but reprogramming these nucleases is a cumbersome, time-consuming process, which requires protein engineering of new nucleases for each new DNA target. In contrast, CRISPR nucleases use guide RNAs to direct cleavage of target nucleic acid, and the nuclease protein remains constant. This feature of CRISPR-Cas systems makes reprogramming remarkably easy and has facilitated their use as tools that have enabled genome engineering in organisms from all domains of life, including humans.

In contrast to this CRISPR-fueled DNA genome engineering revolution, there has been no such CRISPR revolution in the field of RNA engineering. As such, there are few tools to study the role of RNA in cellular processes, RNA repair mechanisms, or the genetic basis of RNA virus biology. RNA biology is not only a foundational component of all cellular life, but also an important factor behind human disease. Dysregulation of RNA processing underlies diseases including Hutchinson-Gilford progeria syndrome[5], spinal muscular atrophy[6], genetic retinopathies[7], and dysregulation of RNA metabolism has been tied to several neurodegenerative diseases[8]. RNA viruses are the etiologic agent of many human diseases, including AIDS, Ebola Hemorrhagic Fever, and COVID-19. To study the genetics of these viruses and understand mechanisms of their pathogenesis, we require tools to perform modifications within these genomes. Currently, the only tools available to make such modifications depend completely on reverse genetics systems. These systems are cumbersome, error prone, and are not easily applied to rapidly evolving viruses. In the case of SARS-CoV-2, the genome sequence was made public on January 10, 2020, but the first published method for a reverse genetics system was published

over a year later[9]. Further, this reverse genetic system is based on the earliest genome sequences of SARS-CoV-2 and does not reflect currently circulating viral variants.

A focus of my research as a master's student has been to address this need for new tools for RNA editing. To this end, we have adapted effector proteins from Type III CRISPR systems to develop a novel technology for RNA genome engineering. This technology has the potential to revolutionize how RNA is studied in both eukaryotic and prokaryotic systems and accelerate the speed at which RNA viruses can be studied.

The impetus for developing such RNA editing technology arose, in part, from the first project I was involved with in the Wiedenheft lab, *SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression*[10]. In this project, we identified a viral variant of SARS-CoV-2 circulating in Bozeman, MT, which carried a large mutation in an accessory protein, ORF7a. We measured a fitness defect in this viral variant, presumably due to the mutation of ORF7a. We were unable to directly tie this mutation to the observed fitness defect, though, because there are no tools for facile genetic manipulation of the RNA genome. As such, I began focusing on developing strategies for editing the SARS-CoV-2 RNA genome using a Type-III CRISPR system.

While focusing on these RNA editing projects, I also began developing a project centered on the basic science behind Type-III CRISPR immunity in bacteria. Type-III CRISPR systems are unique among CRISPR systems for two major reasons: first, they are one of only two systems that target sequence specific RNA, and second, upon recognition of target RNA, they initiate a signaling event which activates ancillary effectors (e.g., nonspecific nucleases and proteases) within the cell, which puts the cell into an antiviral state. The aim of my project was to identify the substrate preferences of different ancillary nucleases and determine the impact of

this immune activation upon bacterial growth kinetics to shed light on the immune strategies of bacteria harboring these Type-III CRISPR systems.

Chapter Two- When I came to the Wiedenheft lab, I joined a group studying a mutated variant of SARS-CoV-2 circulating in Bozeman, Montana. We identified a viral variant with a truncation mutation in the viral accessory protein, ORF7a, which resulted in a frameshift mutation, and initiated an early stop codon. *In vitro* viral challenge studies demonstrated that the virus with this mutation had impaired growth kinetics, compared to the wildtype virus. We went on to determine that this truncated ORF7a lost its function, which normally antagonizes the type-I interferon response in human cells[11]. We hypothesize that this loss in function is what primarily drove the fitness defect we observed, but because we were studying naturally occurring viral variants, we were unable to perform genetic modifications in these viral genomes to create isogenic controls and directly tie this impaired viral fitness to the ORF7a mutation. This constraint demonstrated the need for new tools for genome engineering within RNA genomes, which is the impetus for the project in chapter two.

Chapter Three- This chapter outlines the development of an *in vitro* RNA genome engineering approach. This method utilizes purified Csm complex from the bacterium *Streptococcus thermophilus* (SthCsm). Csm is a Type-III CRISPR effector which utilizes a guide RNA to identify and cleave target RNA. We demonstrate that SthCsm can be reprogrammed with a guide RNA to cleave an RNA sequence of interest. We leverage this RNA targeting activity to specifically cleave a target site within the SARS-CoV-2 RNA genome. We then direct the repair of this RNA *in vitro* by splint ligation. We then used long-read deep sequencing to demonstrate that this approach results in an edited target site with a specific 18 nucleotide deletion. Finally, to recover a virion with an edited genome, we planned to electroporate this

edited RNA genome into mammalian cells. Unfortunately, this electroporation method has yet to be successful, and has not resulted in replication of any virus.

Chapter Four- Working with several members of our lab, I developed a project that focused on the basic science behind Type-III CRISPR immune strategies. The most unique feature of Type-III immunity is its signaling pathway, in which the surveillance complex identifies target RNA, then produces signaling molecules that go on to activate nonspecific ancillary CRISPR nucleases within the cell. The goal of this project was to determine the substrate preferences of ancillary Cas effectors from a Type-III CRISPR system to better understand how activation of such ancillary nucleases may lead to growth arrest of a bacterium experiencing a phage challenge.

References Cited

1. Wilkinson, R. A., Martin, C., Nemudryi, A. A. & Wiedenheft, B. CRISPR RNA-guided autonomous delivery of Cas9. *Nat Struct Mol Biol* **26**, 14–24 (2019).
2. Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180087 (2019).
3. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**, 174–182 (2005).
4. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
5. Pollex, R. & Hegele, R. Hutchinson–Gilford progeria syndrome. *Clinical Genetics* **66**, 375–381 (2004).
6. Lorson, C. L., Hahnen, E., Androphy, E. J. & Wirth, B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences* **96**, 6307–6311 (1999).
7. Poulos, M. G., Batra, R., Charizanis, K. & Swanson, M. S. Developments in RNA Splicing and Disease. *Cold Spring Harb Perspect Biol* **3**, a000778 (2011).
8. Liu, E. Y., Cali, C. P. & Lee, E. B. RNA metabolism in neurodegenerative disease. *Disease Models & Mechanisms* **10**, 509–518 (2017).
9. Xie, X. *et al.* Engineering SARS-CoV-2 using a reverse genetic system. *Nature Protocols* **16**, 1761–1784 (2021).

10. Nemudryi, A. *et al.* SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Rep* **35**, 109197 (2021).

11. Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports* **33**, 108234 (2020).

CHAPTER TWO

SARS-COV-2 GENOMIC SURVEILANCE IDENTIFIES NATURALLY OCCURING TRUNCATION OF ORF7A THAT LIMITS IMMUNE SUPPRESSION

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Artem Nemudryi

Contributions: designed the research, performed the research, prepared manuscript figures, wrote the manuscript, wrote the manuscript

Co-Author: Anna Nemudraia

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Tanner Wiegand

Contributions: designed the research, prepared manuscript figures, performed the research, wrote the manuscript

Co-Author: Joseph Nichols

Contributions: designed the research, performed the research

Co-Author: Deann T. Snyder

Contributions: designed the research, performed the research

Co-Author: Jodi F. Hedges

Contributions: designed the research, performed the research

Co-Author: Calvin Cicha

Contributions: performed the research

Co-Author: Helen Lee

Contributions: performed the research

Co-Author: Karl K. Vanderwood

Contributions: performed the research

Co-Author: Diane Bimczok

Contributions: acquired samples

Co-Author: Mark A. Jutila

Contributions: wrote the manuscript

Co-Author: Blake Wiedenheft

Contributions: acquired samples, designed the research, wrote the manuscript

Manuscript Information

Artem Nemudryi, Anna Nemudraia, Tanner Wiegand, Joseph Nichols, Deann T. Snyder, Jodi F. Hedges, Calvin Cicha, Helen Lee, Karl K. Vanderwood, Diane Bimczok, Mark A. Jutila, and Blake Wiedenheft

Cell Reports

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

Cell Press

Volume 35

Issue 9

1 June 2021

DOI: 10.1016/j.celrep.2021.109197

SARS-COV-2 GENOMIC SURVEILANCE IDENTIFIES NATURALLY OCCURING TRUNCATION OF ORF7A THAT LIMITS IMMUNE SUPPRESSION

Artem Nemudryi[1,3,*], Anna Nemudraia[1,*], Tanner Wiegand[1,†],Joseph Nichols[1,†], Deann T. Snyder[1], Jodi F. Hedges[1], Calvin Cicha[1], Helen Lee[1], Karl K. Vanderwood[2], Diane Bimczok[1], Mark A. Jutila[1] and Blake Wiedenheft[1,4,5,‡]

[1]Department of Microbiology and Immunology,

Montana State University, Bozeman, MT 59717, USA

[2]Gallatin City-County Health Department, Bozeman, MT, 59715, USA

[3]Twitter: @artemnemudryi

[4]Twitter: @WiedenheftLab

[5]Lead contact

[*]These authors contributed equally.

[†]These authors contributed equally.

[‡]Correspondence: bwiedenheft@gmail.com (B.W.)

Abstract

Over 950,000 whole genome sequences of SARS-CoV-2 have been determined for viruses isolated from around the world. These sequences have been critical for understanding the spread and evolution of SARS-CoV-2. Using global phylogenomics, we show that mutations frequently occur in the C-terminal end of ORF7a. We have isolated one of these mutant viruses from a patient sample and used viral challenge experiments to link this isolate (ORF7a$^{\Delta 115}$) to a growth defect. ORF7a has been implicated in immune modulation, and we show that the C-terminal truncation negates anti-immune activities of the protein, which results in elevated type I interferon response to the viral infection. Collectively, this work indicates that ORF7a mutations occur frequently and that these changes affect viral mechanisms responsible for suppressing the immune response.

Introduction

The spillover of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) into the human population has resulted in a global pandemic of Coronavirus Disease 2019 (COVID-19) with more than 2.8 million deaths worldwide (https://coronavirus.jhu.edu/) (Andersen et al., 2020). While much of the research on this virus is focused on the Spike protein, recent reports demonstrate that accessory proteins of SARS-CoV-2 might be involved in COVID-19 pathogenesis by modulating antiviral host responses (Young et al., 2020; Zhang et al., 2020). SARS-CoV-2 uses multiple strategies to evade host immunity. Accessory proteins ORF3b, ORF6 and ORF7a antagonize various steps of type I interferon (IFN-I) production and signaling, while the proposed function of ORF8 is downregulating antigen presentation (Konno et al., 2020; Lei et al., 2020; Miorin et al., 2020; Tan et al., 2020; Xia et al., 2020). To occlude signal transmission from IFN receptors, ORF7a subverts phosphorylation of STAT2, suppressing transcriptional activation of antiviral IFN-stimulated genes (ISGs) that can otherwise restrict viral replication (Martin-Sancho et al., 2020; Xia et al., 2020). While the proposed function for ORF7a is intracellular, antibodies against ORF7a are elevated in the serum of COVID-19 patients (Hachim et al., 2020). Work that is currently under review indicates that recombinant ORF7a protein interacts with $CD14^{+}$ monocytes and triggers expression of pro-inflammatory cytokines, including interleukin-6 (IL-6) and tumor necrosis factor alpha (TNF-α) (Zhou et al., 2020b). However, it is unclear if these interactions happen in COVID-19 patients. Together, these data suggest that ORF7a plays a dual role in SARS-CoV-2 infection by modulating both the IFN and inflammatory responses.

Here, we show that C-terminal mutations in ORF7a occur frequently in samples isolated from patients around the globe. These mutations are not derived from a single lineage and they

do not persist over time. Using samples collected from infected patients, we have isolated a virus containing a deletion mutation in ORF7a that truncates the C-terminal half of the protein. *In vitro* viral challenge experiments demonstrate that this mutation results in a replication defect and obviates viral suppression of the immune response. Collectively, these data suggests that ORF7a truncations are defective in suppressing the host immune response, which may explain why these mutations quickly disappear in the immunocompetent population.

Results

SARS-CoV-2 genomic surveillance identifies truncations of ORF7a

Genome sequencing has been used to track the rise and spread of new SARS-CoV-2 lineages over the course of the pandemic (https://www.gisaid.org/) (Bedford et al., 2020; Korber et al., 2020). As part of this effort, we sequenced SARS-CoV-2 genomes isolated from patients in Bozeman, Montana (**Figure 2.1A, Table S2.1**). In total, we determined 55 whole genome sequences of SARS-CoV-2 viruses isolated from patients between April and July of 2020 using an amplicon-based Nanopore protocol from the ARTIC network (https://artic.network/) (Quick et al., 2017; Tyson et al., 2020). To place the local outbreak in the context of the global SARS-CoV-2 evolutionary trends, we aligned 55 whole genome sequences isolated from patients in Bozeman to 4,235 genomes subsampled (10 genomes per month per country) from 181,003 SARS-CoV-2 genomes downloaded from the GISAID. The subsampling protocol removes redundancy and bias introduced by uneven distribution of the global SARS-CoV-2 sequencing effort. The resulting alignment was used to build a phylogenetic tree (**Figure 2.1B**). Of the 55 SARS-CoV-2 genomes determined from patients in Bozeman, only one was derived from the WA1 lineage (lineage A in **Figure 2.1B**) that was introduced to Washington state from Wuhan, China in January 2020 (Bedford et al., 2020; Holshue et al., 2020). The remaining 54 genomes associate with clade B.1 and its offshoots, which are characterized by the D614G mutation in the spike and have prevailed globally since the spring of 2020 (Rambaut et al., 2020a)(**Figure 2.1B**). The genetic variability in SARS-CoV-2 circulating in Bozeman (April to July 2020) represents fourteen independent viral lineages, reflecting multiple introductions of the virus to the community from many different sources (**Figure S2.1A**).

During the annotation of these genomes, we noticed a reoccurring (7 out of 55 genomes) 115 nt deletion in the gene encoding accessory protein ORF7a (27,549 - 27,644 nt). The ORF7a$^{\Delta 115}$ mutation was found in patient swabs collected over a period of ~1.5 months (**Table S2.1**). RT-PCR and Sanger sequencing verified that all these seven ORF7a$^{\Delta 115}$ variants are bona fide mutations and not a sequencing artifact (**Figure S2.1B, C**). In addition to the ORF7a$^{\Delta 115}$ deletion, these genomes also share 10 single nucleotide variants (SNVs) compared to the Wuhan-Hu-1 reference genome. Seven of the 10 SNVs are found frequently worldwide and are a signature of the B1.1 lineage of SARS-CoV-2 (Rambaut et al., 2020a). The remaining 3 mutations are rare, do not co-occur in any other genomes on GISAID, and lead to amino acid changes in ORF3a (Q38P, L95F) and N (R195I) proteins. Interestingly, one of the genomes in the ORF7a$^{\Delta 115}$ cluster has two additional SNVs that likely were acquired during circulation in the community. This virus was sampled from a patient 41 days after sampling the first ORF7a$^{\Delta 115}$ variant, which agrees well with estimated SARS-CoV-2 mutation rates (2 nucleotides/month) (van Dorp et al., 2020).

To determine if the ORF7a$^{\Delta 115}$ genotype is unique to Bozeman we downloaded an alignment of 180,971 SARS-CoV-2 genomes from GISAID, extracted ORF7a sequences and determined their mutational profiles. In total, we identified 845 unique ORF7a gene variants that are different from the Wuhan-Hu-1 reference sequence. Next, we looked for mutations with major effect on the ORF7a amino acid sequence. We identified 189 unique ORF7a variants (825 total sequences) with frameshifts, deletions or missense mutations causing premature stop codons (**Figure 2.1C**). To understand the evolutionary history of these mutations, we integrated these genome sequence into our phylogenetic analysis. Detected ORF7a variants appear to have emerged independently on every continent and were not confined to any single lineage (**Figure**

2.1B; Table S1.1). Next, we aligned translated ORF7a sequences to look for patterns. Most of the identified ORF7a mutations (126 out of 189) truncate the C terminus but preserve the N-terminal half of the protein (**Figure 2.1C and Figure S2.1B**). A portion (36.5%) of these truncated ORF7a variants appeared in two or more patient samples (116 max). While some of these ORF7a mutants appear to have arisen on multiple independent occasions, others come from genomes that form monophyletic clades, suggesting that viruses with ORF7a truncations are capable of transmission within a host population (**Figure S2.1B**).

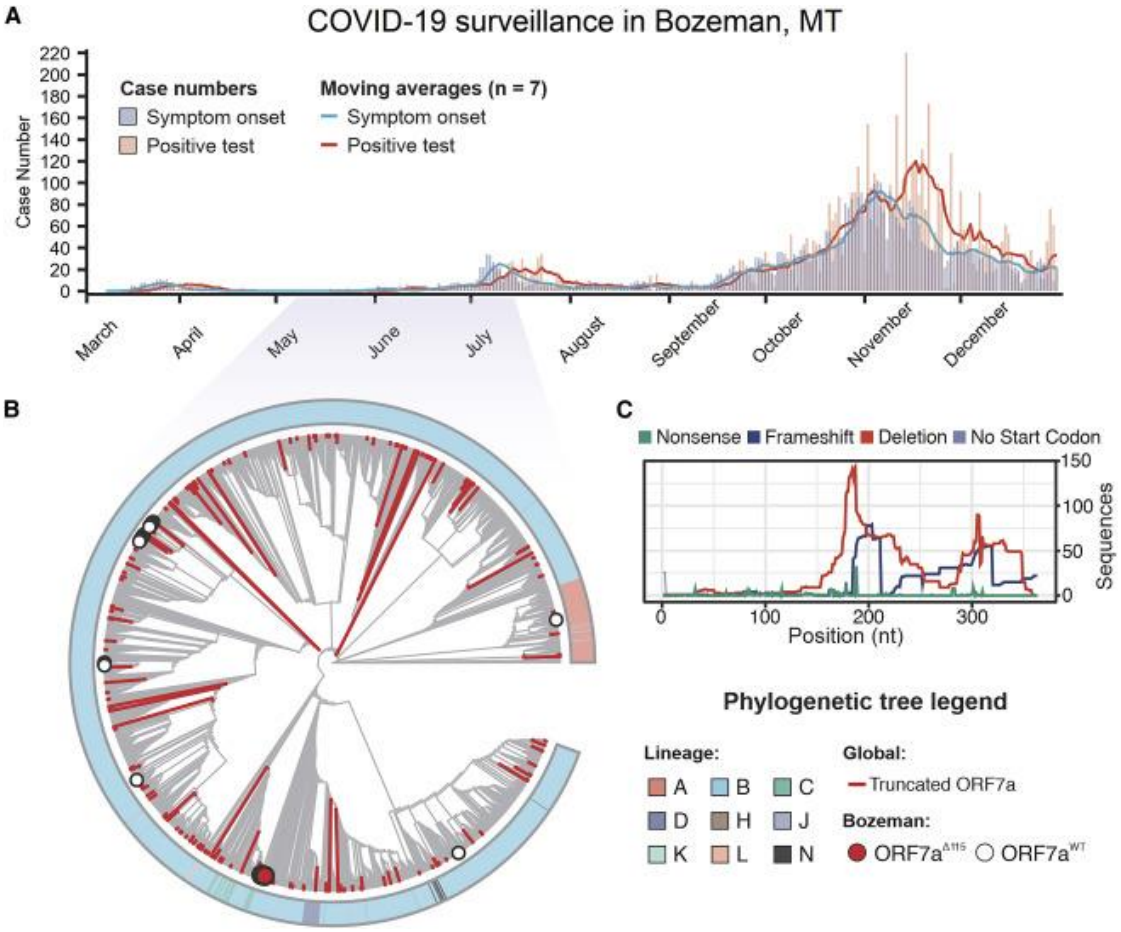


Figure 2.1. SARS-CoV-2 genomic surveillance identifies global reoccurrence of ORF7a truncations (A) Symptom onset (purple) and PCR-based SARS-CoV-2 test results (coral) for patients in Bozeman, Montana, are shown with vertical bars. Seven-day moving averages, shown with lines, were used to indicate epidemiological trends. (B) Phylogenetic analysis of SARS-CoV-2 genomes sampled in Bozeman and globally. The tree was constructed from an alignment of 55 Bozeman samples and 4,871 genomes subsampled from GISAID. Subsampling was performed using Augur utility (https://nextstrain.org) by selecting 10 genomes per country per month since the start of the pandemic. Outer ring shows SARS-CoV-2 lineages assigned to genome sequences (Rambaut et al., 2020a). Major lineages include A (pink) that is associated with initial outbreak in China and B (blue) that emerged later in Europe. Minor lineages (i.e., C-N) are offshoots of lineage B. Red branches identify truncated ORF7a variants (n = 205) detected in the global data and merged into the alignment. The red dot highlights 7 of the 55 ORF7a variants that were isolated in Bozeman between April and July (2020). White dots highlight 48 viral genomes isolated in Bozeman that have wild-type ORF7a sequences. (C) Distribution of different mutations that occur along the ORF7a coding sequence.

ORF7a truncation has a loss-of-function effect

ORF7a is a type I transmembrane protein with an N-terminal immunoglobulin-like (Ig-like) ectodomain, stalk, transmembrane (TM) domain and a short cytosolic tail (**Figure 2.2A**). The two β sheets of the ectodomain are held together by two disulfide bonds (i.e., Cys23-Cys58 and Cys35-Cys67) (**Figure 2.2B**). The cytosolic tail of ORF7a contains a dilysine (KRKTE) ER retrieval signal (ERRS) that mediates protein trafficking to the ER-Golgi intermediate compartment (ERGIC) (Nelson et al., 2005).

The Δ115 nt mutation in ORF7a introduces a premature stop codon that eliminates β5, β6 and β7, two of the cystines that form disulfides (Cys58 and Cys67), the TM and the cytosolic tail (**Figures 2.2A and B**). This truncation is expected to destabilize the protein structure and significantly impact protein function and trafficking in the host cell. To determine how loss of TM and sorting signal affects protein localization, we cloned and expressed Flag-tagged wildtype and truncated ORF7a proteins in HEK 293T-hACE2 cells. The wildtype ORF7a accumulated in the perinuclear region of the cell, which is consistent with previously reported ERGIC localization (**Figure 2.2D**) (Martin-Sancho et al., 2020; Nelson et al., 2005). In contrast, the truncated ORF7a is distributed throughout the cytoplasm and does not associate with specific subcellular compartments, which is consistent with the loss of the TM domain and the ERRS signals required for protein targeting (**Figure 2.2E**).

Extent of the truncation and the defect in intracellular targeting suggest possible loss of ORF7a function. Along with several SARS-CoV-2 accessory proteins, ORF7a has been implicated in suppression of host IFN response to the infection (Xia et al., 2020). To test the effect of the truncated ORF7a protein on IFN signaling, we generated reporter HEK 293T cells with integrated luciferase reporter gene controlled by the Interferon Stimulated Response

Element (ISRE). In these cells, activation of interferon response drives reporter expression, which is quantified using a luciferase assay. To test immune suppression, we overexpressed the wildtype and the deletion mutant of ORF7a (ORF7a$^{\Delta 115}$) from plasmids and then stimulated cells with IFNα2b. The ORF7a WT protein suppressed IFN-response by 34% (p-value = 0.0074) compared to cells transfected with control plasmid, which is consistent with recent results by Xia et al (Xia et al., 2020). In contrast, the ORF7a$^{\Delta 115}$ protein failed to suppress IFN response and showed no significant difference to a control (mCherry) plasmid (**Figure 2.2F**). These results indicate that the Δ115 mutation in ORF7a has a loss-of-function effect that negates antagonism of IFN signaling.

ORF7a mutation has no collateral effect on ORF7b

The ORF7a gene overlaps with the downstream ORF7b gene (**Figure S2.2**). To determine if the Δ115 mutation in ORF7a impacts ORF7b, we first examined whether it eliminates a transcription-regulatory sequence (TRS) that is required for subgenomic RNA (sgRNA) synthesis. None of the TRSs identified via direct RNA sequencing of SARS-CoV-2 overlap with the Δ115 mutation site (Kim et al., 2020), suggesting that ORF7b transcription is not affected (**Figure S2.2**). To confirm this prediction, we used RT-PCR with primers that detect ORF7a and ORF7b sgRNAs (**Figure 2.2G**). In this assay, we infected 293T-hACE2 cells (MOI = 0.05) with ORF7a^{WT} or ORF7a$^{\Delta 115}$ viral strains and extracted total RNA from cells at 24 hours post infection (hpi). Both viruses produced specific RT-PCR products corresponding to ORF7a and ORF7b sgRNAs. Finally, to verify that the ORF7a$^{\Delta 115}$ variant does not impact ORF7b translation, we probed cell lysates 24 hpi with an anti-ORF7b antibody (**Figure 2.2H**). We detected ORF7b protein in both viral strains with band intensities that suggest similar expression

levels. Collectively, these data indicate that the ORF7a$^{\Delta 115}$ mutation has no collateral effect on the adjacent gene.

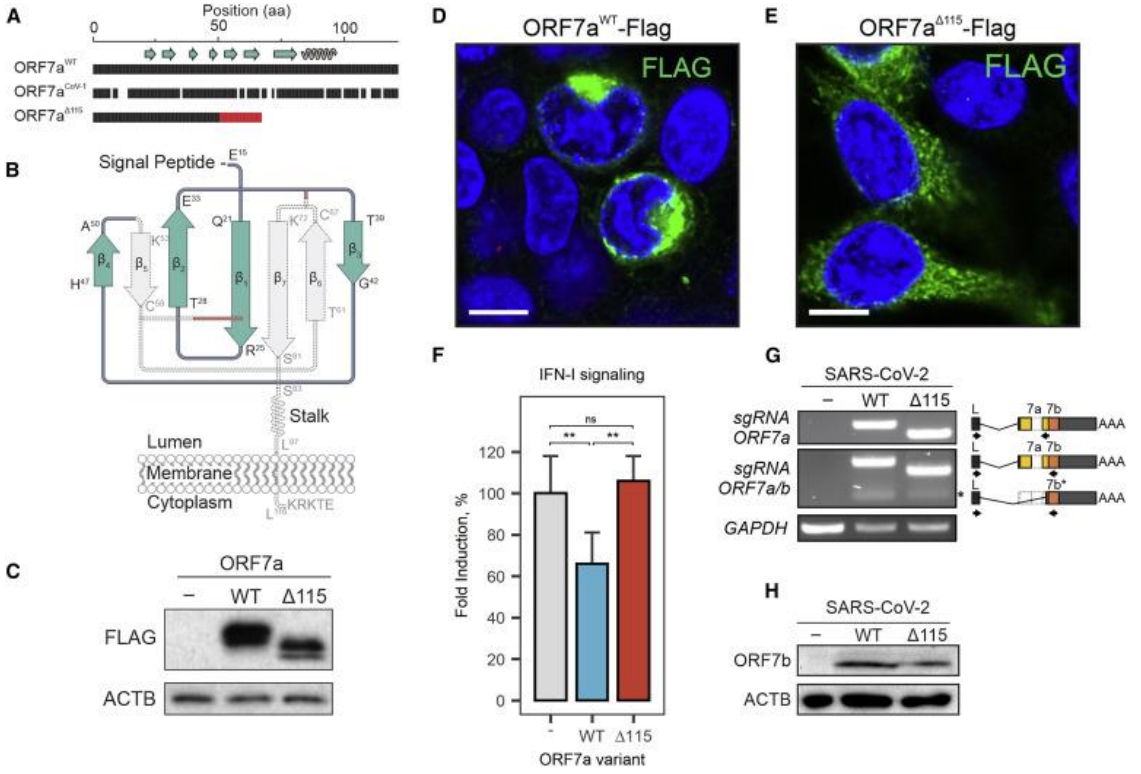


Figure 2.2. ORF7a truncation results in loss of function (A) Amino acid (aa) sequence alignment of SARS-CoV-2 ORF7aWT, ORF7a$^{\Delta 115}$, and SARS-CoV-1 ORF7a. Gaps show non-matching positions; red shows 17-aa sequence resulting from a frameshift in the ORF7a mutant. Beta strands (arrows) and alpha helices (coil) are shown above the alignment. (B) Diagram of SARS-CoV-2 ORF7a Ig-like fold. Disulfide bonds that stabilize the β sandwich structure are shown with red lines. The portion of the protein eliminated by the deletion is shown in gray. (C) C-terminally FLAG-tagged ORF7aWT and ORF7a$^{\Delta 115}$ were cloned and overexpressed in HEK293T-hACE2. Protein expression was confirmed with western blot using anti-FLAG antibody. β-Actin (ACTB) was used as a loading control. (D and E) FLAG-tagged ORF7aWT (D) or ORF7a$^{\Delta 115}$ (E) expressed in HEK293T-hACE2 cells. Immunostaining was performed using an anti-FLAG antibody (green). Cell nuclei were stained with Hoechst 33342 (blue). White scale bar is 10 μm. (F) HEK293T cells with integrated ISRE-luciferase reporter were transfected with pLV-mCherry ("-"), ORF7aWT, or ORF7a$^{\Delta 115}$ plasmids. Transfected cells were treated with 5 ng/mL human recombinant IFN-α2b for 24 h, and induction of type I IFN signaling was measured with luciferase assay. Fold induction versus non-treated control was determined and normalized to mCherry control. Means (n = 6) were compared with one-way ANOVA ($p = 0.00162$). Pairwise comparisons were performed using post hoc Tukey's test. Data are shown as mean ± SD. $**p < 0.01$; nsp > 0.05. (G) ORF7a and ORF7b sgRNAs were identified by RT-PCR. Lower molecular weight band corresponding to ORF7b sgRNA is indicated with asterisk (∗). Diagram on the right shows primer (arrows) positions. Specificity of PCR products was confirmed with Sanger sequencing. GAPDH was used as a control for cDNA synthesis. (H) Lysates from SARS-CoV-2-infected cells were probed with antibodies raised against ORF7b protein. ACTB was used as a loading control.

ORF7a$^{\Delta 115}$ isolate displays a replication defect

To investigate phenotype of SARS-CoV-2 Δ115 variant, we infected 293T-hACE2 and Vero E6 cells with ORF7a^{WT} and ORF7a$^{\Delta 115}$ viruses at an MOI 0.05 and measured viral RNA replication using RT-qPCR (**Figure 2.3A and B**). In this assay, we used viral strains sharing the same haplotype (C241T, C3037T, C14408T, A23403G) that defines the SARS-CoV-2 B.1 lineage and its derivatives (**Figure S2.3A**) (Rambaut et al., 2020a). Over the course of 120 hours, ORF7aWT viral RNA level increased ~960-fold in supernatants from Vero E6 cells (Figure 3A) and ~270-fold in supernatants from 293T-hACE2 cells (**Figure 2.3B**). The ORF7a$^{\Delta 115}$ virus displayed limited replication that resulted in only 18.6-fold and 7-fold RNA level increase at 120 hpi in VeroE6 and 293T-hACE2, respectively. This growth defect reaches statistical significance starting at 6 hpi in Vero E6 cells (p-value = 0.0135) and 24 hpi in 293T-hACE2 cells (p-value = 6.2×10-5).

To determine if this phenotype results from a defect in viral egress only, or if viral RNA replication in the host is also affected, we examined early steps in the infection. We extracted total RNA from the infected 293T-hACE2 cells at 0.5, 6 and 24 hpi, and measured SARS-CoV-2 RNA with RT-qPCR. Compared to wildtype, ORF7a$^{\Delta 115}$ RNA level was reduced 2.1-fold inside the host at 6 hpi (p-value = 0.002) and 4.3-fold at 24 hpi (p-value = 0.02) (**Figure 2.3C**). This decrease in viral RNA in the host cell differs from the corresponding decrease in the supernatant RNA at 24 hpi (4.3-fold vs 16.8-fold, respectively; **Figure 2.3B, C**). This disparity suggests that observed growth defect in SARS-CoV-2 Δ115 variant might result from defects in genomic RNA replication and transcription, as well as in viral egress.

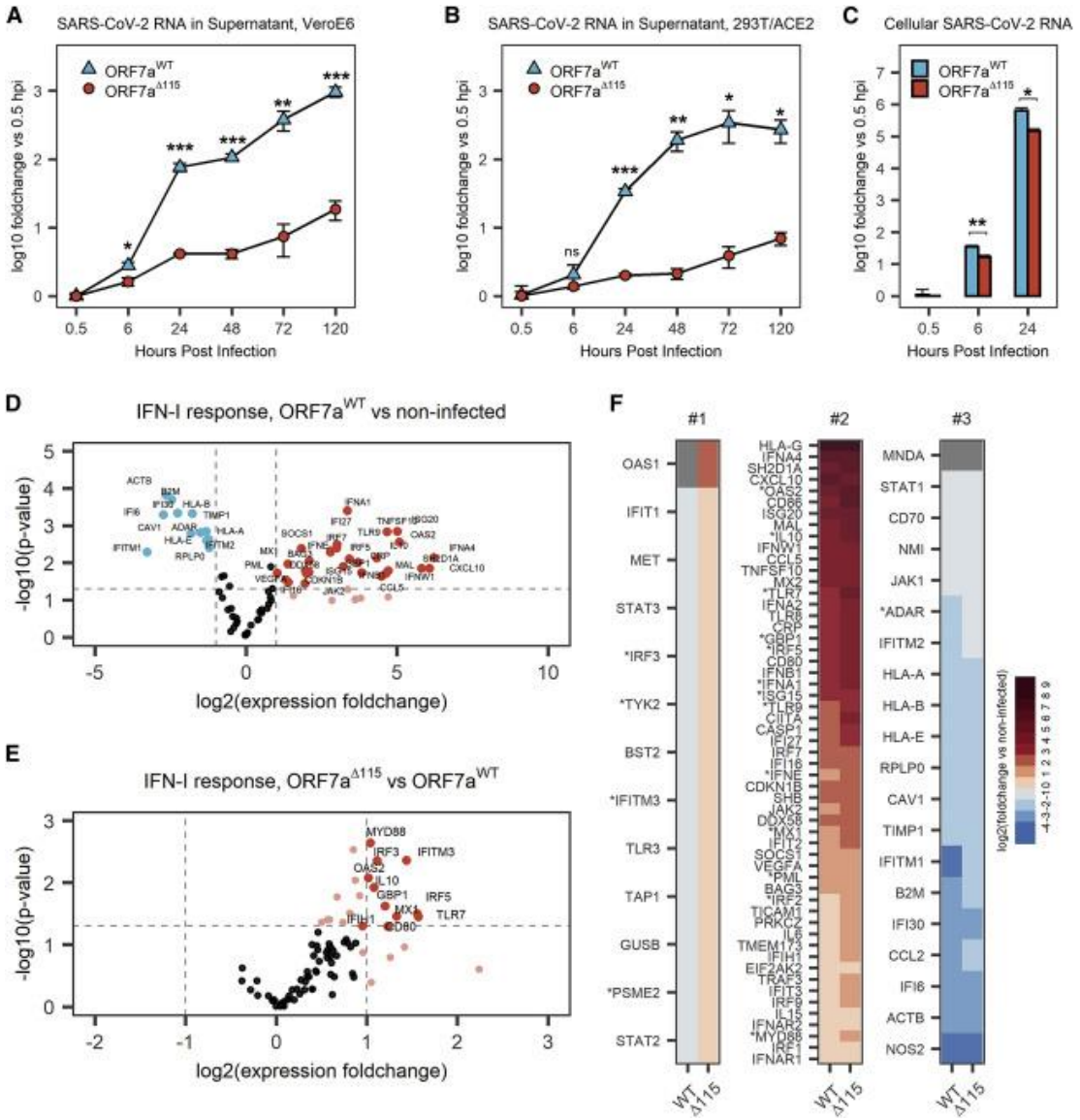


Figure 2.3. The ORF7a$^{\Delta 115}$ mutation results in replication and immune suppression defects (A and B) VeroE6 (A) or HEK293T-hACE2 (B) cells were infected with ORF7a^{WT} or ORF7a$^{\Delta 115}$ SARS-CoV-2 strain at MOI = 0.05. Viral RNA was measured in supernatant using qRT-PCR at different time points after infection. Measured RNA levels were normalized (using ΔCt method) to RNA levels at 0.5 hpi. (C) Total RNA was extracted from infected HEK293T-hACE2 cells, and viral RNA was measured. The ΔΔCt method was used to normalize viral RNA levels to host RNA (ACTB) and 0.5 hpi time point. Data in (A)–(C) are presented as mean ± SD of three biological replicates. Significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, or ns$p > 0.05$ (no significant difference). (D) Volcano plot showing IFN-I response in 293T-hACE2 cells infected with ORF7aWT SARS-CoV-2 strain at MOI = 0.05. Expression of IFN-I response genes was studied 24 hpi using qRT-PCR array targeting 91 human transcripts (88 targets and 3 references). Experiment was performed in three biological replicates. Dashed lines show regulation (≥2-fold) and statistical significance ($p < 0.05$) thresholds. Each dot represents mean (n = 3) normalized expression of a single gene relative to non-infected host. Genes that passed the threshold are

labeled. (E) Volcano plot showing IFN-I response in ORF7a$^{\Delta 115}$ versus ORF7a^{WT} infection. (F) Data shown in (D) and Figure S3A were plotted as a heatmap. Genes were classified into three groups. Group #1 genes are oppositely regulated between the two viral strains. Group #2 genes are upregulated by both ORF7a^{WT} and ORF7a$^{\Delta 115}$. Group #3 genes are downregulated in both. Genes that have statistically significant difference in expression between two viral strains are marked with asterisk. Gray shows genes with no detectable expression over 40 cycles of qRT-PCR.

ORF7a$^{\Delta 115}$ virus induces elevated Type I interferon response in vitro

Suppressed or delayed IFN responses characterize SARS-CoV-2 infections (Lei et al., 2020). The C-terminal truncation of ORF7a limits suppression of the type I IFN signaling (**Figure 2.2F**). Therefore, we hypothesized that ORF7a$^{\Delta 115}$ virus isolated from the community is deficient in antagonizing IFN response.

To test this hypothesis, we measured expression of type I IFN response genes upon infection with ORF7a^{WT} and ORF7a$^{\Delta 115}$ strains using an RT-qPCR array targeting 91 human transcripts (**Table S2.2**). The array included 3 reference genes that we used to normalize expression of 88 target genes. In agreement with recent transcriptomic studies, we detected a type I IFN-mediated antiviral response after infecting HEK 293T-hACE2 with SARS-CoV-2 isolates (**Figure 2.3D, Figure S2.3B**) (Banerjee et al., 2020; Blanco-Melo et al., 2020). Out of 88 type I IFN response genes included in the RT-qPCR array, 42 were differentially expressed (≥2-fold change, p-value <0.05) in ORF7a^{WT} infected cells and 55 in ORF7a$^{\Delta 115}$ infection, as compared to the non-infected control (**Figure 2.3D, Figure S2.3B**). Both viruses stimulated expression of multiple type-I interferon genes, antiviral ISGs, interleukins and proinflammatory cytokines (**Table S2.2**).

To understand the effect of ORF7a$^{\Delta 115}$ on the host type I IFN response, we compared expression of genes targeted with RT-qPCR array between the two viruses. In general, the ORF7a$^{\Delta 115}$ variant induces elevated type I IFN response to the infection (**Figure S2.3C, D**). The median expression fold change of 88 genes targeted in the RT-qPCR array was 1.72 in ORF7a^{WT} and 2.51 in ORF7a$^{\Delta 115}$ virus (p-value = 1.4E-12; **Figure S2.3C**). Analysis of genes differentially expressed between the two viral infections results in an asymmetric volcano plot (**Figure 2.3E**). A subset of ISGs had expression levels that are significantly different (p-value < 0.05) between

the two viral infections (**Figure 2.3E, F, Table S2.2**). This subset includes sensors (*TLR7),* signal transducers (*MYD88, OAS2),* transcriptional regulators (*IRF3*, *IRF5*) and restriction factors (*GBP1*, *IFITM3, MX1*) known to combat infections by RNA viruses (Schoggins, 2019). One of these differentially expressed ISGs (*IFITM3*) restricts SARS-CoV-2 entry (Shi et al., 2020; Zang et al., 2020). Additionally, polymorphisms in *IFITM3, MX1* and *TLR7* are linked to the severity of COVID-19 (Andolfo et al., 2020; Jin et al., 2020a; Pati et al., 2021). Upregulation of ISGs indicates that the ORF7a$^{\Delta 115}$ virus is inept at antagonizing the IFN-I response to the infection, which is consistent with results from our ISRE reporter assay (**Figure 2.2F**).

Discussion

In this study, we use genomic surveillance and phylogenetics to identify ORF7a variants that have emerged globally throughout the SARS-CoV-2 pandemic. Local occurrence of ORF7a mutations has been reported by others (Addetia et al., 2020; Holland et al., 2020; Rosenthal et al., 2020). However, the effect of these mutations on viral fitness and host immune responses has not been investigated. Here, we isolated a virus with a deletion mutation in ORF7a (ORF7a$^{\Delta 115}$) and show that this genotype results in an *in vitro* growth defect. This growth defect is associated with elevated IFN response to SARS-CoV-2.

Interferon systems are a frontline of host defense that signals infection and elicits an antiviral state (McNab et al., 2015). Pretreatment with type I, type II and type III IFNs restricts SARS-CoV-2 infection and replication in cell culture (Mantlo et al., 2020; Miorin et al., 2020). SARS-CoV-2 deploys multiple proteins (e.g., nsp6, nsp13, ORF6, ORF7a, ORF7b and ORF9b) to shut down IFN signaling and dampen innate immune responses (Jiang et al., 2020; Lei et al., 2020; Xia et al., 2020). In particular, when overexpressed, ORF7a subverts STAT2

phosphorylation, blocking IFN-dependent transcriptional activation of antiviral ISGs (Xia et al., 2020). Both type I and III IFNs signal through the JAK-STAT pathway and overlap substantially in the transcriptional responses they induce (Kotenko and Durbin, 2017). Therefore, ORF7a likely suppresses both IFN types, however its role in suppressing the type III IFN response has yet to be verified (Xia et al., 2020). Using ISRE reporter system we show that C-terminal truncation negates the ability of ORF7a to suppress IFN signaling (**Figure 2.2**). The ORF7a$^{\Delta 115}$ variant shows limited replication in HEK 293T-hACE2 as well as in Vero E6, the latter of which do not express type I IFN genes (Emeny and Morgan, 1979; Osada et al., 2014). While type I IFN genes are deleted in Vero E6 cells, the type III IFNs are intact and are activated upon infection with RNA viruses, which results in ISG upregulation (Stoltz and Klingström, 2010; Wang et al., 2020). ISGs exert anti-viral activities that combat the infection (Schoggins, 2019). We hypothesize that the replication defect of the ORF7a$^{\Delta 115}$ strain in Vero E6 and 293T-hACE2 cells results from an IFN-dependent response that the virus fails to suppress.

Several groups have replaced ORF7a with a reporter gene (Hou et al., 2020; Thi Nhu Thao et al., 2020; Xie et al., 2020). Thao *et al.* show that complete deletion of ORF7a reduces replication of the synthetic virus, which agrees well with growth defect in the ORF7a$^{\Delta 115}$ strain. However, two other studies have not observed this effect. We believe that our results with a naturally occurring ORF7a deletion provide valuable insight to this discussion. Finally, it should be noted that different mutations (i.e., complete deletion vs partial) might have a different outcome for the viral phenotype.

Cytokine profiling and RNA-seq of clinical samples reveal suppressed or delayed IFN responses in COVID-19 patients (Arunachalam et al., 2020; Hadjadj et al., 2020; Lucas et al., 2020). Dysregulated immune responses coupled with exacerbated inflammatory cytokines drive

pathogenesis of the disease and correlate with its severity (Jin et al., 2020b; Tay et al., 2020; Zhou et al., 2020a). Here, we detected enhanced IFN signaling in cells infected with a naturally occurring ORF7a$^{\Delta 115}$ variant of SARS-CoV-2.

Given the importance of the IFN response in COVID-19, we anticipate that ORF7a mutations might affect SARS-CoV-2 pathogenicity. Clinical comparisons will be required to examine this potential effect on COVID-19 features.

Deletions in ORF7a, 7b and ORF8 have previously been reported (Addetia et al., 2020; Holland et al., 2020; Rosenthal et al., 2020; Su et al., 2020). Similar deletions emerged in MERS- and SARS-CoV, which were linked to viral attenuation (Chinese, 2004; Muth et al., 2018). A 382-nt deletion in ORF8 is associated with milder SARS-CoV-2 infection in patients (Young et al., 2020). While this mutation affects clinical features of COVID-19, it has no evident effect on viral replication or host transcriptional responses in human cell culture (Pekosz et al., 2020).

Several studies illustrate how persistent SARS-CoV-2 infections in immunocompromised hosts associate with accelerated viral evolution (Avanzato et al., 2020; Choi et al., 2020). In these hosts there is less selective pressure on viral anti-immune proteins, which might permit loss-of-function mutations in accessory genes. The ORF7a$^{\Delta 115}$ variant that we isolated, was present only in 7 genomes over a course of a ~1.5 month in local community and then disappeared. We hypothesize that deletions in accessory genes, such as ORF7a$^{\Delta 115}$, might emerge in immunocompromised patients but do not persist in the immunocompetent population. Similarly, a deletion mutation in ORF8 (i.e., Δ382), which was associated with milder COVID-19 disease, disappeared soon after it was discovered (Young et al., 2020). However, founder effect or mutations that co-occur with other mutation that increase transmissibility (e.g. spike

mutations) can result in fixation of an otherwise attenuating mutation. A new more transmissible lineage of SARS-CoV-2 (i.e., B.1.1.7) has recently emerged (Davies et al., 2021; Rambaut et al., 2020b). Aside from changes in the Spike protein, this lineage also includes a premature stop codon in the accessory ORF8 that truncates most of the protein (Volz et al., 2021). This lineage might illustrate the second scenario for fixing mutations in accessory genes.

Limitations of Study

The viral replication defect and cellular immune response assay presented here were performed using HEK 293T-hACE2 and Vero E6 cells that are not the natural targets of SARS-CoV-2. However, both cell lines are permissive for SARS-CoV-2 infection and are used extensively to study efficacy of antiviral compounds (Riva et al., 2020), virion assembly (Klein et al., 2020), SARS-CoV-2 entry (Hoffmann et al., 2020), replication (Thi Nhu Thao et al., 2020) and anti-immune activities of SARS-CoV-2 proteins (Xia et al., 2020). While we acknowledge that our conclusions are limited to Vero E6 and 293T-hACE2 cells, we anticipate that the results presented here will provide important context for clinical comparisons, similar to those recently published for ORF8 (Young et al., 2020).

Studies of naturally occurring SARS-CoV-2 variants have limitations including unavailability of isogenic controls. Here, we compared two viral isolates, ORF7a^{WT} and ORF7a$^{\Delta 115}$, that in addition to Δ115 differ at several other nucleotide positions (**Figure S2.3A**). While ISRE reporter assay unambiguously links the immunosuppression defect to the truncation of ORF7a, it is not possible to rule out a role for these mutations on the viral lifecycle.

Acknowledgements

We are grateful to members of Bozeman Health that provided de-identified patient samples. Specifically, C. Nero, D. Smoot, W. Wallace, C. Faurot-Daniels, V. Lawrence and M. Blauvelt. We are also grateful to M. Flenniken, K. Daughenbaugh, and other members of the COVID task force at MSU for assistance establishing the COVID testing center. We thank the Lefcort lab for generous use of their fluorescent confocal microscope and Dr. Pincus for providing the HEK 293T-hACE2 cells. Research in the Wiedenheft lab is supported by the National Institutes of Health (1R35GM134867), the Montana State University Agricultural Experimental Station, the MJ Murdock Charitable Trust, the Gianforte Foundation and the MSU Office of the Vice President for Research. We thank the GISAID's EpiFlu™ Database and contributing laboratories (**Table S2.4**). The phylogenetic analysis in this paper would not have been possible without their willingness to share data. Graphical abstract was created with BioRender.com.

Declaration of interests

B.W. is the founder of SurGene LLC, and VIRIS Detection Systems Inc. B.W., A. Nemudryi, and A. Nemudraia are inventors on patents related to CRISPR-Cas systems and applications thereof.

References

1 Addetia, A. *et al.* Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol* **129**, 104523 (2020).

2 Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat Med* **26**, 450–452 (2020).

3 Andolfo, I. *et al.* Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19. *iScience* **24**, 102322 (2021).

4 Arunachalam, P. S. *et al.* Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).

5 Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901-1912.e9 (2020).

6 Banerjee, A. *et al.* Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *iScience* **24**, 102477 (2021).

7 Bedford, T. *et al.* Cryptic transmission of SARS-CoV-2 in Washington state. *Science* **370**, 571–575 (2020).

8 Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).

9 Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *New England Journal of Medicine* **383**, 2291–2293 (2020).

10 Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).

11 Emeny, J. M. & Morgan, M. J. Y. 1979. Regulation of the Interferon System: Evidence that Vero Cells have a Genetic Defect in Interferon Production. *Journal of General Virology* **43**, 247–252.

12 Gamage, A. M. *et al.* Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLOS Pathogens* **16**, e1009130 (2020).

13 Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology* **20**, 8 (2019).

14 Hachim, A. *et al.* ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. *Nat Immunol* **21**, 1293–1301 (2020).

15 Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).

16 Harcourt, J. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg Infect Dis* **26**, 1266–1273 (2020).

17 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

18 Holland, L. A. *et al.* An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *Journal of Virology* (2020) doi:10.1128/JVI.00711-20.

19 Holshue, M. L. *et al.* First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine* **382**, 929–936 (2020).

20 Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **182**, 429-446.e14 (2020).

21 Jiang, H. *et al.* SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol* **17**, 998–1000 (2020).

22 Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921.e10 (2020).

23 Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885 (2020).

24 Konno, Y. *et al.* SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Reports* **32**, 108185 (2020).

25 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).

26 Kotenko, S. V. & Durbin, J. E. Contribution of type III interferons to antiviral immunity: location, location, location. *Journal of Biological Chemistry* **292**, 7295–7303 (2017).

27 Lei, X. *et al.* Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun* **11**, 3810 (2020).

28 Lemieux, J. E. *et al.* Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* **371**, eabe3261 (2021).

29 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

30 Loveday, E. K. *et al.* Effect of Inactivation Methods on SARS-CoV-2 Virion Protein and Structure. *Viruses* **13**, 562 (2021).

31 Lucas, C. *et al.* Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).

32 Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S. & Huang, C. Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Research* **179**, 104811 (2020).

33 Martin-Sancho, L. *et al.* Functional landscape of SARS-CoV-2 cellular restriction. *Molecular Cell* **81**, 2656-2668.e8 (2021).

34 McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat Rev Immunol* **15**, 87–103 (2015).

35 Miorin, L. *et al.* SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *PNAS* **117**, 28344–28354 (2020).

36 Muth, D. *et al.* Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep* **8**, 15177 (2018).

37 Nelson, C. A., Pekosz, A., Lee, C. A., Diamond, M. S. & Fremont, D. H. Structure and Intracellular Targeting of the SARS-Coronavirus Orf7a Accessory Protein. *Structure* **13**, 75–85 (2005).

38 Osada, N. *et al.* The Genome Landscape of the African Green Monkey Kidney-Derived Vero Cell Line. *DNA Research* **21**, 673–683 (2014).

39 Pati, A., Padhi, S., Suvankar, S. & Panda, A. K. Minor Allele of Interferon-Induced Transmembrane Protein 3 Polymorphism (rs12252) Is Covered Against Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Mortality: A Worldwide Epidemiological Investigation. *The Journal of Infectious Diseases* **223**, 175–178 (2021).

40 Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261–1276 (2017).

41 Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).

42 Riva, L. *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113–119 (2020).

43 Rosenthal, S. H. *et al.* Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. 2020.12.10.418855 (2020) doi:10.1101/2020.12.10.418855.

44 Schoggins, J. W. Interferon-Stimulated Genes: What Do They All Do? *Annual Review of Virology* **6**, 567–584 (2019).

45 Stoltz, M. & Klingström, J. Alpha/Beta Interferon (IFN-α/β)-Independent Induction of IFN-λ1 (Interleukin-29) in Response to Hantaan Virus Infection. *Journal of Virology* **84**, 9140–9148 (2010).

46 Su, Y. C. F. *et al.* Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio* (2020) doi:10.1128/mBio.01610-20.

47 Tan, Y., Schneider, T., Leong, M., Aravind, L. & Zhang, D. Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related Viruses. *mBio* (2020) doi:10.1128/mBio.00760-20.

48 Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* **20**, 363–374 (2020).

49 THE CHINESE SARS MOLECULAR EPIDEMIOLOGY CONSORTIUM. Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China. *Science* **303**, 1666–1669 (2004).

50 Thi Nhu Thao, T. *et al.* Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).

51 Tyson, J. R. *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* 2020.09.04.283077 (2020) doi:10.1101/2020.09.04.283077.

52 van Dorp, L. *et al.* No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* **11**, 5986 (2020).

53 Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. 2020.12.30.20249034 (2021) doi:10.1101/2020.12.30.20249034.

54 Wang, C. *et al.* The Coronavirus PEDV Evades Type III Interferon Response Through the miR-30c-5p/SOCS1 Axis. *Frontiers in Microbiology* **11**, (2020).

55 Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

56 Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports* **33**, 108234 (2020).

57 Xie, X. *et al.* An Infectious cDNA Clone of SARS-CoV-2. *Cell Host & Microbe* **27**, 841-848.e3 (2020).

58 Young, B. E. *et al.* Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet* **396**, 603–611 (2020).

59 Zang, R. *et al.* Cholesterol 25-hydroxylase suppresses SARS-CoV-2 replication by blocking membrane fusion. *PNAS* **117**, 32105–32113 (2020).

60 Zhang, B. *et al.* Clinical characteristics of 82 cases of death from COVID-19. *PLOS ONE* **15**, e0235458 (2020).

61 Zhang, Y. *et al.* The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. *PNAS* **118**, (2021).

62 Zhang, Y. *et al.* Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C Associated With Disease Severity in Coronavirus Disease 2019. *The Journal of Infectious Diseases* **222**, 34–37 (2020).

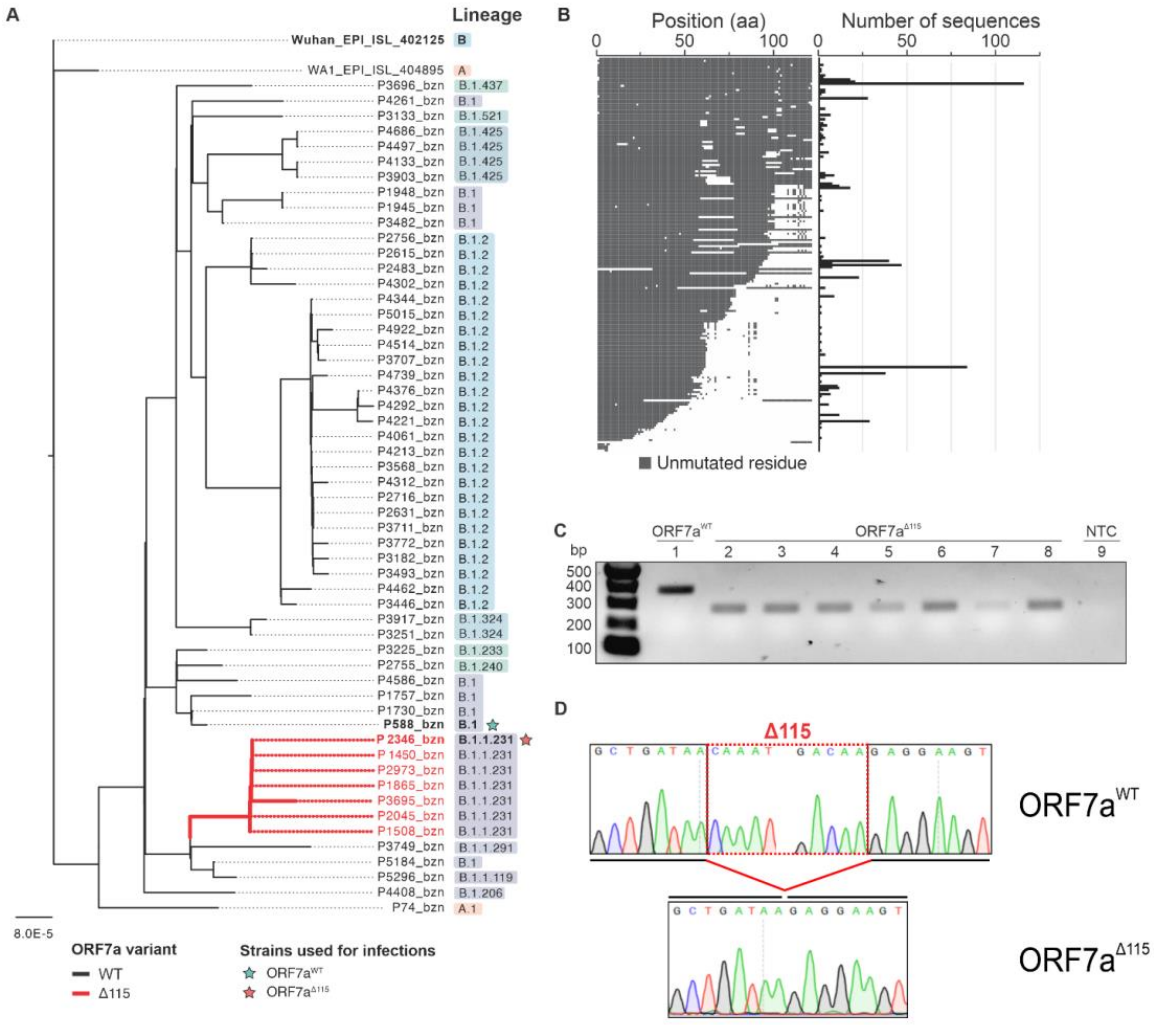
63 Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **395**, 1054–1062 (2020).

64 Zhou, Z. *et al.* Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience* **24**, 102187 (2021).

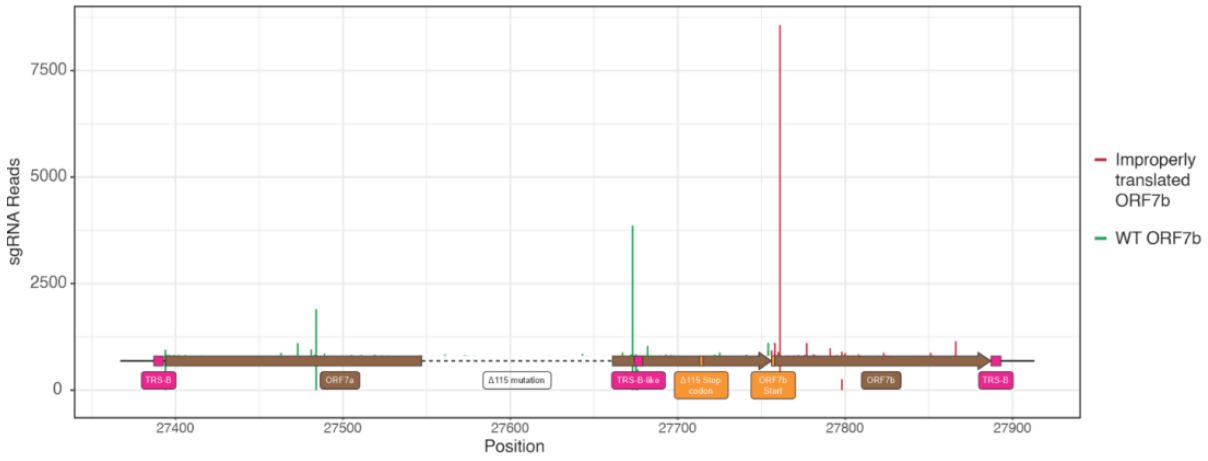
65 Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological* https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).

66 Opposing activities of IFITM proteins in SARS-CoV-2 infection. *The EMBO Journal* **40**, e106501 (2021).

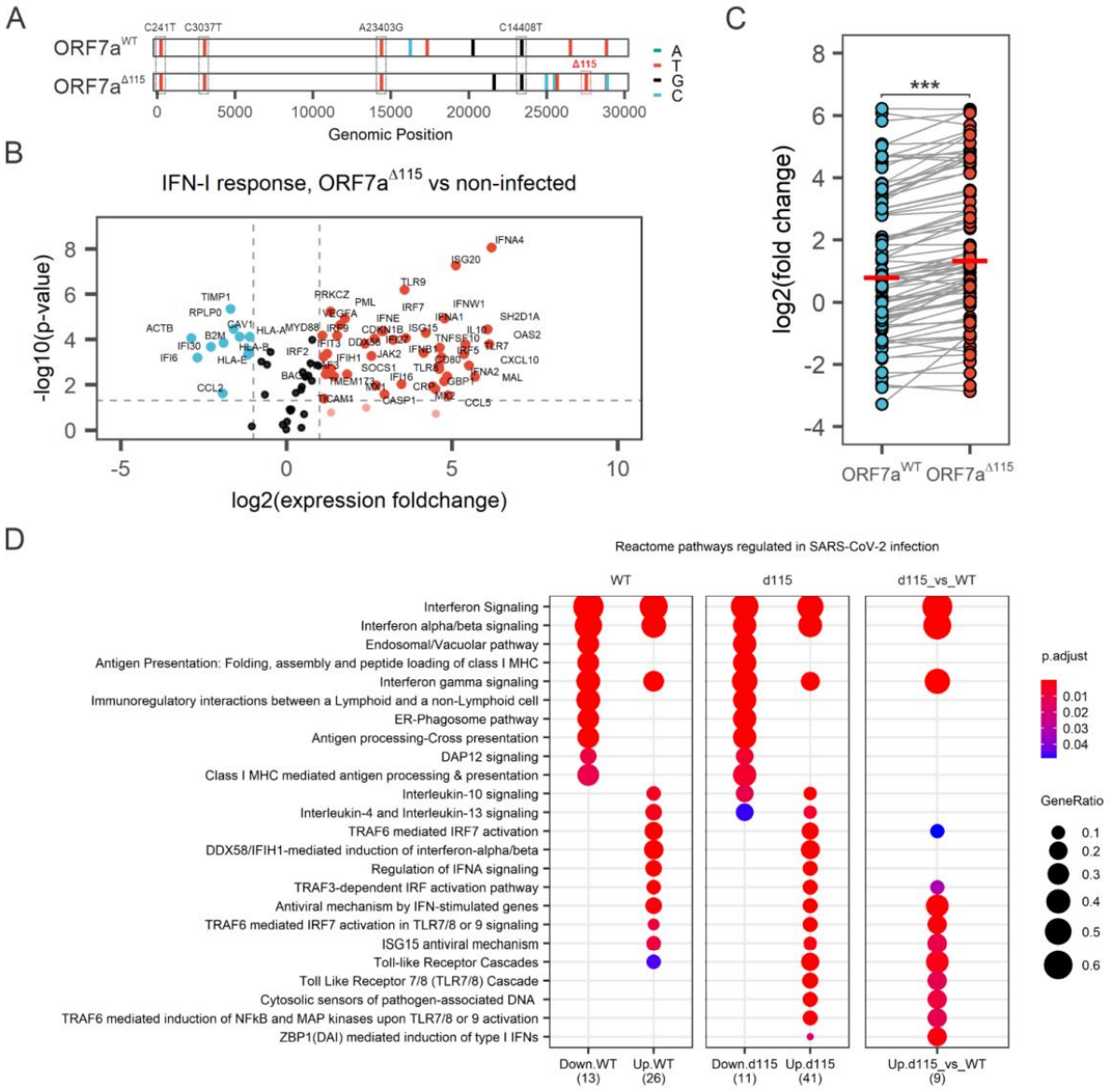
Supplemental Figures



Supplemental Figure S2.1, Phylogenetic and mutational analysis of ORF7a gene sequences. Related to main Figure 2.1. A) A maximum-likelihood phylogenetic tree indicates that ORF7a$^{\Delta 115}$ strains form a monophyletic clade (red branches). Viral isolate P588 (teal star) was chosen as a "wildtype" control in our experiments. This strain shares a recent common ancestor with the predominant ORF7a$^{\Delta 115}$ clade. SARS-CoV-2 lineages were assigned using pangolin utility (cov-lineages.org, (Rambaut, Andrew et al., 2020)). B) Alignment of unique ORF7a protein sequences with five or more mutations as compared to the SARS-CoV-2 reference strain (Wuhan-Hu-1). Conserved residues are depicted in gray, while mutations and deletions are shown in white. The frequency of each genotype is shown on the right. C) RTPCR with primers targeting regions that flank the ORF7a Δ115 deletion sampled in Bozeman, MT. D) All PCR products in (C) were Sanger sequenced. Representative sequencing chromatograms are shown. Deleted region is shown with dotted red box.



Supplemental Figure S2.2, Mapping TRS-B sites required for ORF7a and ORF7b discontinuous sgRNA transcription. Related to main Figure 2.2F-H. Direct RNA sequencing data published by Kim et al was used to map TRS-B sites in SARS-CoV-2 ORF7a and ORF7b genes (Kim et al., 2020). Vertical lines indicate number of reads spanning the junction between TRS-L and "body" of the gene. Dashed line indicate region deleted in ORF7aΔ115 strain. The deletion does not remove any of the mapped TRS-B sites suggesting that it has not affected ORF7b sgRNA transcription.



Supplemental Figure S2.3, Related to Main Figure 3. IFN-I response and gene set enrichment analysis. A) Genome sequences of ORF7a^{WT} and ORF7a$^{\Delta115}$ are compared to Wuhan-Hu-1 reference sequence. Colored lines mark differences to reference genome. Mutations that define haplotype of B.1 SARS-CoV-2 lineage are boxed. There are 11 nucleotide positions that differ between WT and Δ115 and are outside ORF7a. Seven of these SNVs change amino acid. In ORF7a^{WT}: S194L in N protein and L1286F in ORF1ab. In ORF7a$^{\Delta115}$: R195I, R203K, G204K in N protein; Q38P and L95I in ORF3a. B) Volcano plot depicting IFN-I response in HEK 293T-hACE2 cells infected with ORF7a$^{\Delta115}$ SARS-CoV-2 strain at MOI = 0.05. RNA was extracted from infected cells 24 hpi and reverse transcribed. Expression of IFN-I response genes was studied using RT-qPCR array targeting 91 human transcripts (88 targets and 3 references). Experiment was performed in 3 biological replicates. Dashed lines show regulation (≥2-fold) and statistical significance thresholds (p-value < 0.05). Each dot represents mean (n = 3) normalized

expression of a single gene relative to non-infected host. Genes that passed the threshold are labeled. C) Plot showing distribution of log2(expression fold change) values in ORF7a^{WT} and ORF7a$^{\Delta 115}$ infection vs non-infected control. Each dot represents relative normalized expression for one transcript targeted with the RT-qPCR array. Gray lines connect values for the same transcript in the infection with ORF7a^{WT} and ORF7a$^{\Delta 115}$ variants. Horizontal red bar shows median value. Median ISG activation between two infections was compared using Wilcoxon signed-rank test (***, p-value < 0.001) D) List of IFN-I response genes that were significantly regulated (p-value < 0.05, ≥ 2-fold change in expression) upon infection with ORF7a$^{\Delta 115}$ and ORF7a^{WT} (vs non-infected control) were analyzed for enrichment of associated Reactome Pathways (Jassal et al., 2019; Zhou, Z. et al., 2020). Gene set enrichment analysis (GSEA) was performed using clusterProfiler package in RStudio (Yu and He, 2016; Yu et al., 2012). Down.WT – genes downregulated during infection with ORF7a^{WT}; Up.WT – genes upregulated during infection with ORF7a^{WT}; Down.d115 – Reactome pathways enriched in genes downregulated during infection with ORF7a$^{\Delta 115}$; Up.d115 – genes upregulated during infection with ORF7a$^{\Delta 115}$. Up.d115_vs_WT – Reactome pathways enriched in genes upregulated during infection with ORF7a$^{\Delta 115}$ vs ORF7a^{WT} (Figure 1.3E).

Materials and Methods

Cell Culture

HEK293T-hACE2 cells (BEI, NR-52511) were generously provided by Dr. Seth Pincus (Montana State University) and maintained at 37°C and 5% CO2 in Dulbecco's Modified Eagle Medium (DMEM, GIBCO, cat. #12100-061) supplemented with 10% fetal bovine serum (FBS, ATLAS Biologicals, Lot. #F31E18D1), sodium bicarbonate (3.7 g/L), 50 I.U./mL penicillin and 50 µg/mL streptomycin. Vero E6 cells (ATCC® cat. #CRL-1586) were maintained at 37°C and 5% CO2 in Eagle's Modified Eagle Medium (EMEM, ATCC®, cat. #30-2003) supplemented with 10% fetal bovine serum (FBS, ATLAS Biologicals, Lot. #F31E18D1) and 50 I.U./mL penicillin and 50 µg/mL streptomycin. All cell lines tested negative for mycoplasma.

Plasmids

Human-codon optimized dsDNA gene fragments (gBlocks) encoding for SARS-CoV-2 C-terminally FLAG-tagged ORF7aWT and ORF7aΔ115 proteins were synthesized by Integrated DNA Technologies (IDT). Sequence of ORF7a from Wuhan-Hu-1 reference (GenBank MN908947.3) was used for ORF7aWT gBlock. The gene fragments were cloned into the pEGFP-N1 (Clontech cat# 6085-1) backbone to replace the eGFP gene. pLV-mCherry was a gift from Pantelis Tsoulfas (Addgene plasmid # 36084; http://www.addgene.org/36084/; RRID:Addgene_36084).

Antibodies

Western blotting: Rabbit anti-ACTB (Cat: AC026, 1:20,000) antibodies were from ABclonal, Mouse anti-FLAG antibodies (Cat: MA1-91878-1MG, 1:1000) were from ThermoFisher Scientific, sheep anti-ORF7b antibodies (1:120) were from MRC I PPU, College

of Life Sciences, University of Dundee, Scotland; Goat anti-mouse IgG peroxidase-conjugated (Cat: 115-035-003, 1:10,000), Goat anti-rabbit IgG peroxidase-conjugated (Cat: 111-035-003, 1:10,000) and Donkey Anti-Sheep IgG peroxidase-conjugated (Cat: 713-035-147, 1:10,000) antibodies were from Jackson ImmunoResearch. Immunocytochemistry: Mouse anti-FLAG antibodies (Cat: MA1-91878-1MG, 1:200), Goat anti-mouse-AlexaFluor-488 (Cat: A11001, 1:2,000) antibodies were from Invitrogen.

Human clinical sample collection and preparation

Clinical samples were obtained with local IRB approval (protocol #DB033020) and informed consent from patients undergoing testing for SARS-CoV-2 at Bozeman Health Deaconess Hospital. Patients tested for SARS-CoV-2 included both in-patients and out-patients. The latter included individuals who developed symptoms and sought medical care and asymptomatic (at the time of testing) individuals who were exposed to known COVID-19 case and therefore were tested. Nasopharyngeal swabs from patients that tested positive for SARS-CoV-2 were collected in viral transport media. RNA was extracted from all patient samples using QIAamp Viral RNA Mini Kit (QIAGEN) a biosafety level 3 (BSL3) laboratory. All samples were heat-inactivated before removing from BSL3.

Production and titration of coronavirus stocks

The nasopharyngeal swabs in the viral transport media were used to generate viral stocks as previously described (Harcourt, *et. al;* 2020). Briefly, 100 ul of the viral transport media was two-fold serially diluted and applied to Vero E6 cells. When infected cells showed extensive cytopathic effect, the media was collected and utilized to generate a greater volume second passage in Vero E6 cells.

When CPE was apparent, supernatants were centrifuged 1,000 RCF for 5 minutes to remove cellular debris. Virus titer in clarified supernatants was determined with plaque assay in Vero E6 cells as described. For plaque assay, Vero E6 cells were incubated with viral inoculum at limiting dilutions. Inoculated cells were overlayed with either 0.75% methylcellulose, DMEM supplemented with 2% FBS and 1% pen-strep and incubated for 4 days. Cells were fixed and stained with 0.5% methylene blue in 70% ethanol. Plaques were counted and the overall titer was calculated. Viral stock's identity was confirmed via whole genome sequencing on Oxford Nanopore. All SARS-CoV-2 experiments were performed in a BSL3 laboratory.

Symptom onset data and clinical test results

Suspect cases of COVID-19 were tested in a CLIA lab and instructed to self-quarantine until notified of the RT-qPCR test results. All laboratory confirmed positive cases of COVID-19 were contacted via telephone by local public health nurses to complete contact tracing. During this interview, the nurses collected recorded symptoms, symptom onset date, travel history, contact with other known laboratory confirmed cases, close contacts and activities on the two days before symptom onset up until notification of a positive test. Data collection was conducted as part of a public health response. The study was reviewed by the Montana State University Institutional Review Board (IRB) For the Protection of Human Subjects (FWA 00000165) and was exempt from IRB oversight in accordance with Code of Federal regulations, Part 46, section 101. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived.

Quantitative Reverse Transcription PCR (qRT-PCR)

qRT-PCR was performed using CDC primers (N1 and N2) and probes from the 2019-nCoV RUO Kit (IDT# 10006713). SARS-CoV-2 RNA was quantified using one-step qRT-PCR in ABI 7500 Fast Real-Time PCR System according to CDC protocol (https://www.fda.gov/media/134922/download). In brief, 20 μL reactions 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 template. Nuclease-free water was used as negative template control (NTC). Amplification was performed using following program: 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. Run data was analyzed in SDS software v1.4 (Applied Biosystems). The NTC showed no amplification throughout the 40 cycles of qPCR.

RT-PCR and SARS-CoV-2 Genome Sequencing

For sequencing 10 mL of RNA from SARS-CoV-2 positive patient sample was reverse transcribed using SuperScript IV (Thermo Fisher Scientific) according to the supplier's protocol. The protocol developed by ARTIC Network was used to generate and sequence amplicon library that covers whole SARS-CoV-2 genome on Oxford Nanopore using ligation sequencing kit (SQK-LSK109) (https://artic. network/ncov-2019). Briefly, two multiplex PCR reactions were performed with primer pools from ARTIC nCoV-2019 V3 Panel (IDT; Table S3) using Q5 High-Fidelity DNA Polymerase (New England Biolabs). Reactions were performed with the following thermocycling conditions: 98°C for 2min, 30 cycles of 98°C for 15 s and 65°C for 5 min. Two resulting amplicon pools for each patient sample were combined and used for library preparation. After end repair (NEB E7546) samples were barcoded using Native Barcoding Expansion Kits EXP-NBD104 and EXP-NBD114 from Oxford Nanopore. A total of 24

barcoded samples were pooled together and Nanopore adaptors were ligated. After clean-up 20 ng of multiplexed library DNA was loaded onto the MinION flowcell for sequencing. A total of 0.3 Gb of raw sequencing data was collected per patient sample. Nuclease-free water was used as a negative control for library preparation and sequencing. Non-specific amplification in negative control was additionally checked using an agarose gel. SARS-CoV-2 Isolate USA-WA1/2020 (BEI, NR-52281) was used as a positive control for genome sequencing.

SARS-CoV-2 Genome Assembly

MinKNOW software was used to basecall raw Nanopore reads in high-accuracy mode. ARTIC bioinformatic pipeline for COVID-19 was used to analyze reads (https://artic.network/ncov-2019). Pipeline included demultiplexing with guppy barcoder and generating consensus sequences with minimap2 and calling single nucleotide variants with nanopolish relative to Wuhan-Hu-1/2019 reference genome.60-62 20X coverage was used as a threshold both for consensus and variant calling. Nucleotide positions with less than 20X coverage were masked with Ns in the final consensuses. Consensus sequences were uploaded to GISAID (https://www.gisaid.org/), accession IDs are available in the online supplementary material associated with this manuscript (https://doi.org/10.1016/j.celrep.2021.109197).

RT-PCR and Sanger Sequencing

To verify the ORF7aΔ115 mutation we used RT-PCR with primers that flank the deletion region. Reactions were performed with Q5 High-Fidelity DNA Polymerase in 25 μL volume (New England Biolabs) using following program: 98°C for 2min, 35 cycles of 98°C for 15 s and 65°C for 5 min, 35 cycles. PCR products were analyzed on 1% agarose gels stained with SYBR Safe (Thermo Fisher Scientific). The remaining volume of the reaction was purified using DNA

Clean & Concentrator kit (Zymo Research) and sent to Psomagen for Sanger sequencing. Each PCR product was sequenced with both forward and reverse primers used for PCR. To detect sgRNAs of ORF7a and ORF7b primers annealing to Leader region of SARS-CoV-2 and inside the ORF were used. PCR products were cut out from 1% agarose gel and purified with Zymoclean Gel DNA Recovery Kit. Purified products were sequenced with forward and reverse primers. Sequence was verified by aligning to SARS-CoV-2 genome in UGENE software (Unipro).

Phylogenetic and ORF7a Mutational Analysis

An alignment of 181,003 SARS-CoV-2 genomes was downloaded from the GISAID database at 9:11 AM on 2020-11-11, and sequences without corresponding metadata entries were removed. The nucleotide positions encoding ORF7a in the reference sequence (Wuhan-Hu-1; EP_ISL_402125) were extracted for the remaining 180,971 sequences, and ORF7a sequences with 100% sequence identity were clustered using CD-HIT (with settings: -c 1 -aL 1). One representative was randomly selected from each of the 846 ORF7a sequence clusters and mutations in each nucleotide and translated sequence were determined using the Biostrings package in R. Graphs of these mutations were rendered with the ggplot2 package. To construct a phylogenetic tree of global and Bozeman SARS-CoV-2 sequences, up to ten sequences from each country in the world were selected for each month of the pandemic using the Filter utility in the Augur pipeline (–group-by country year month –sequences-per-group 10).2,63 This resulted in an alignment of 4,235 GISAID sequences. Then, the Nextclade online tool was used to determine the quality of 73 SARS-CoV-2 genomes (% genome covered) sequenced from Bozeman patients. 55 Bozeman sequences that were determined to be of “Good” quality were merged with the alignment of GISAID sequences, as well as an alignment of 669 representative

SARS-CoV-2 genomes that had non-synonymous mutations in ORF7a. The bat coronavirus RaTG13 (MN996532.2) and the Wuhan- Hu-1 SARS-CoV-2 sequences were also merged, resulting in an alignment of 4,959 full genome sequences, and positions that are hypervariable or prone to sequencing artifacts were masked with a VCF file downloaded from EMBL at 12:45PM on 2020-05- 29. The alignment was trimmed of columns composed of 90% or more gaps using Trimal (-gt 0.9), and IqTree was used to construct a maximum likelihood phylogeny (-m GTR -ninit 2 -n 2 -me 0.05). Branch lengths and internal nodes were rescaled on the resulting tree using TreeTime, and tree was re-rooted to the RaTG13 sequence using the APE package before being visualized with the ggTree package in R. Clades were assigned to SARS-CoV-2 genomes using Nextclade and pangolin (https://cov-lineages.org/) utilities.

Western Blot

293T-hACE2 cells were transfected using Lipofectamine 3000 (Invitrogen, L3000-015) with plasmids encoding for ORF7a^{WT}, ORF7a$^{\Delta 115}$ or control pEGFP-N1 plasmid. To detect ORF7b expression cells were infected with ORF7a^{WT} or ORF7a$^{\Delta 115}$ SARS-CoV-2 strains at MOI = 0.05. After 24 hours, cells were washed two times with PBS and lysed in RIPA buffer (150 mM sodium chloride, 1% NP-40, 0.5% sodium deoxycholate, 0.2% SDS, 50 mM Tris, pH8.0) at 4°C for 30 min. The lysates were clarified by centrifugation (10’000 g, 20 min) and stored at −80°C. For western blot lysates were mixed with 6X Laemmli SDS-PAGE buffer and heated at 98°C for 5 min, resolved in 12% SDS-PAGE gel and transferred onto a PVDF (ORF7b) or nitrocellulose (ORF7a) membrane using Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, #1703930). Membranes were blocked and probed with indicated antibodies. The proteins were visualized using Pierce ECL Western Blotting Substrate (ThermoFisher Scientific, #32106) and exposed to X-Ray film (sc-201696, Santa Cruz Biotech).

Immunocytochemistry

HEK293T-hACE2 were transfected with plasmid DNA using Lipofectamine 3000 (Invitrogen REF: L3000-015) according to manufacturer's instructions. Cells were either transfected with expression vectors for SARS-CoV-2 ORF7aWT, SARS-CoV-2 ORF7aΔ115, or mock transfected. 24 hours after transfection, cells were detached and seeded at 50% confluency on glass coverslips (Fisherbrand, #12-545-80) pre-coated with poly-D-lysine (Cultrex, #3439-100-01) in a 24 well plate. The next day, media was removed, and cells were washed with PBS and fixed with ice-cold 100% methanol on ice for 10 minutes. Half of the methanol was replaced with PBS three times before the methanol solution was discarded and cells were washed three times in PBS (5 minutes each wash). Cells were blocked with 1% BSA (Fisher Scientific BP9703-100) in PBS + 0.1% Tween-20 (PBST) for 30 minutes at room temperature. After blocking, cells were stained with mouse monoclonal anti-Flag antibody (Invitrogen, MA1-91878) (1:200 in 1% BSA in PBST) overnight at 4°C. The next day, cells were washed 3X in PBS (5 min each) and stained with secondary antibody diluted 1:2000 in 1% BSA in PBST for 1 h at RT in the dark (Goat anti-mouse AlexaFluor488; Invitrogen, A11001). After 3x PBS washes nuclei were stained with Hoechst 33342 in PBS for 5 minutes at RT. Coverslips were then mounted with ProLong Gold antifade mounting media (Invitrogen, P36934) on Superfrost Plus microscope slides (#22-037-246). Immunostained cells were imaged using a Leica SP8 confocal microscope.

IFN Inhibition Assay

To generate ISRE reporter system, HEK293T cells were transduced with lentiviruses carrying pGreenFire1-ISRE (Cat. TR016PA-1, SBI) construct and selected with puromycin (1ug/ml) for 2 days. After selection cells (1x105 cells/well, 48-well plate) were transfected with

expression plasmids (250 ng) for SARS-CoV-2 ORF7a^{WT}, SARS-CoV-2 ORF7a$^{\Delta 115}$, or control plasmid pLV-mCherry. 16 hours post transfection, cells were treated for 24 h with 5 ng/ml of human IFN-α2b (Cat. rcyc-hifna2b, InvivoGen) according to manufacturer's recommendations. After treatment activation of IFN signaling was measured using Luciferase Assay System (Cat. E1500, Promega) on Cytation 5 (BioTek) plate reader. Assay was performed in six replicates for each condition.

SARS-CoV-2 Replication Assays

HEK293T-hACE2 and Vero E6 cells were seeded in the 48-well plate and infected with ORF7aWT or ORF7aΔ115 SARS-CoV-2 strains at MOI = 0.05 in 50 µL of the cell culture media. Infections were performed in triplicates. Next, 0.5 h post infection (hpi) 500 µL of the fresh cell culture media was added to each well. The supernatants (140 ul) from the infected cells were harvested at the following time points 0.5, 6, 24, 48, 72 and 120 hpi. The RNAs from the supernatants were extracted using QIAamp Viral RNA Mini Kit (QIAGEN) and used as an input in qRT-PCR with CDC N1 and N2 primers and probes as described above. Relative quantification was used (ΔCt method) to calculate viral replication versus 0.5 hpi time point as 2-ΔCt. The infected cells were harvested 0.5, 6 and 24 hpi and washed with PBS. Total RNA from cells was extracted using RNeasy kit (QIAGEN, Cat No./ID: 74104) with on-column DNase digestion step (QIAGEN, 79254) according to the manufacturers protocol. Viral RNA was quantified using qRT-PCR CDC N1 primers and probe. Human ACTB endogenous control was quantified with TaqMan assay (ThermoFisher, 4333762T). ΔΔCt method was used to normalize viral RNA level to host RNA and 0.5 hpi. Viral replication was calculated for each replicate as 2-ΔΔCt.

Type I Interferon Response Assay

HEK293T-hACE2 were seeded in the 48-well plate and infected with ORF7aWT or ORF7aΔ115 SARS-CoV-2 strains at MOI = 0.05 in 50 ul of the media. After 30 min on infection the 500 μL of the cell culture media was added to each well. At 24 hpi the cells were washed once with PBS, detached using Trypsin and the total RNA was extracted using RNeasy kit (QIAGEN, Cat No./ID: 74104) with on-column DNase digestion (QIAGEN, Cat No./ID: 79254) according to the manufacturer's protocol. Non-infected cells were processed using the same protocol with mock infection. After extraction, 0,5-1 μg of total RNA was reverse transcribed Reverse using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) according to the supplier's protocol. The cDNA was diluted 10-fold in water and analyzed with Type I interferon response (SAB Target List) H96 qPCR array. Each reaction contained 2 μL of cDNA, 10 μL of 2X SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) and 8 μL of water. Reactions were performed on QuantStudio 3 Real-Time PCR System instrument with following thermocycling conditions: 95°C for 2min, 40 cycles of 95°C for 5 s and 60°C for 30 s. After amplification, melt curve analysis was performed to examine product specificity. Assay was performed in triplicates for non-infected cells and both viruses. Run data was analyzed in PrimePCR Analysis software (Bio-Rad). Expression of IFN-I response genes was normalized to geometric mean of 3 reference transcripts (TBP, GAPDH, HPRT1) and non-infected control (ΔΔCt method).

Quantification and Statistical Analysis

All statistical analyses were performed in RStudio v1.2.1335. Moving averages (n = 7) for symptom onset and positive COVID-19 tests data were calculated using geom_ma function from tidyquant package and plotted with ggplot2 in RStudio. qRT-PCR data is shown as mean of

three biological replicates (each with three technical replicates) ± standard deviation (SD). Data in ISRE reporter assay was analyzed with one-way ANOVA with Tukey's post hoc pairwise comparisons. Replication of viruses and expression of IFN-I response genes was compared using Student's t test. Medians of IFN-I response between two viruses were compared with Wilcoxon signed-rank test. Significance levels in figures: $*$ $p < 0.05$, $**$ $p < 0.01$, $***$ $p < 0.001$ or ns (no significant difference, $p > 0.05$).

CHAPTER THREE

PROGRAMMABLE *IN VITRO* EDITING OF THE SARS-COV-2 GENOME WITH A TYPE-III CRISPR-CAS COMPLEX

Contributions of Authors and Co-authors

Manuscript in Chapter 3

Author: Joseph Nichols

Contribution: Designed the research, performed the research, prepared manuscript figures, wrote the manuscript

Author: Artem Nemudryi

Contribution: Designed the research, performed the research, prepared manuscript figures, wrote manuscript

Author: Anna Nemudryi

Contribution: Designed the research, performed the research, prepared manuscript figures

Co-author: Tanner Wiegand

Contribution: Designed the research, performed the research

Co-author: Trevor Zahl

Contributions: Performed the research

Co-author: Andrew Santiago-Frangos

Contributions: Designed the research, performed the research

Co-author: Deann Snyder

Contributions: performed the research

Co-author: Mark Jutila

Contributions: Designed the research

Co-author: Jodi Hedges

Contributions: Designed the research, performed the research

Co-author: Blake Wiedenheft

Contributions: Designed the research

Manuscript Information

Joseph Nichols, Artem Nemudryi, Anna Nemudraia, Tanner Wiegand, Trevor Zahl, Andrew Santiago-Frangos, Deann Snyder, Mark Jutila, Jodi Hedges, and Blake Wiedenheft

The CRISPR Journal

Status of Manuscript:

__X_ Prepared for submission to a peer-reviewed journal

____ Officially submitted to a peer-reviewed journal

____ Accepted by a peer-reviewed journal

____ Published in a peer-reviewed journal

PROGRAMMABLE *IN VITRO* EDITING OF THE SARS-COV-2 GENOME WITH A TYPE-III CRISPR-CAS COMPLEX

Joseph Nichols*, Artem Nemudryi*, Anna Nemudraia*, Tanner Wiegand, Trevor Zahl, Andrew Santiago-Frangos, Deann Snyder, Mark Jutila, Jodi Hedges, and Blake Wiedenheft

Department of Microbiology and Cell Biology, Montana State University, 59717, USA

*These authors contributed equally.

Correspondence to: bwiedenheft@gmail.com (B.W.)

Abstract

Through the course of the last several decades, developments in molecular biology have generated a diverse toolbox for modifying nucleic acids. One of the most significant developments has been the repurposing of prokaryotic adaptive immune systems, known as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems. Currently, CRISPR systems enable edits to be made in DNA genomes, the most common method employs CRISPR associated protein 9 (Cas9) to make targeted double-stranded breaks in a DNA genome or direct the action of base editors to a specific nucleotide within a genome. There is yet to be an established method to use a CRISPR system to perform targeted edits in viral RNA genomes. We demonstrate a new approach to generate site specific edits in large RNA molecules using the type III-A CRISPR effector complex, Csm, from the bacterium *Streptococcus thermophilus* (SthCsm). Large RNA virus genomes are difficult to modify through traditional cloning methods, and normally require the sequence to be reverse transcribed into cDNA and split onto several DNA cloning vectors or artificial chromosomes. The approach outlined here avoids a DNA intermediate and directly edits the RNA genome before transfecting the edited genome into mammalian cells. In this way, a wild-type viral RNA genome can be directly edited without the introduction of restriction enzyme sites and without the possibility of polymerase-derived nucleotide polymorphisms. We demonstrate the use of this gene editing pipeline to make a targeted deletion in the ~30 kilobase SARS-CoV-2 RNA genome to knock out the accessory gene, ORF7a. This novel method of RNA editing has implications for the study of any RNA molecule, with specific utility for studying RNA viruses.

Introduction

The repurposing of CRISPR/Cas9 as a genome editing platform has revolutionized nearly every area of biologic research[1–3]. Cas9 is an endonuclease from the Type II-A CRISPR system which makes targeted double stranded brakes to DNA, guided by a guide RNA (gRNA) [1–3]. This enzyme has become a useful tool for DNA genome editing, as the double stranded breaks are repaired through imprecise cellular mechanisms, commonly resulting in indel frameshift mutations[2,3]. Despite these CRISPR-fueled breakthroughs in our ability to perform DNA genome engineering, there are currently few available tools for the engineering of RNA genomes. The standard approach for genetic engineering of RNA viruses makes use of a DNA intermediate, segmenting cDNA from the viral RNA into cloning vectors[4], or artificial chromosomes in yeast[5] or bacteria[6], then making subsequent edits through conventional DNA mutagenesis techniques (i.e. PCR mutagenesis, restriction enzyme cloning). These approaches are resource and time intensive, and susceptible to error due to multiple polymerase steps[4,6]. Additionally, these reverse genetics systems are not easily pivoted to studying rapidly evolving viral variants because the entire process must be redone for each new viral variant. With the rise of rapidly evolving variants of a virus during a pandemic, or novel emerging viruses, there is a need for tools to make facile edits to RNA genomes in order to gain an understanding of the genetic basis of virus biology[7,8].

CRISPR systems are adaptive prokaryotic immune systems which defend the cell from genetic parasites like plasmids, transposons, or bacteriophage[1]. While these immune systems are diverse, CRISPR immunity occurs in three universal steps[9,10]. The first is spacer acquisition, the process by which a cell under attack by a phage or plasmid integrates a sequence from the invader into the host chromosome at a CRISPR locus[9,10]. The second step, CRISPR RNA

biogenesis, generates the mature CRISPR RNA (crRNA), or guide RNA (gRNA), from transcribed RNA from the CRISPR locus. crRNA is then complexed with a CRISPR associated (Cas) protein or protein complex to form the effector complex[9,10].That effector complex drives the final step of CRISPR immunity, CRISPR interference. The effector complex surveils nucleic acid within the cell searching for sequence which is complementary to its guiding crRNA. When target nucleic acid matching the guide is identified, the effector complex begins cleaving, or degrading, the target nucleic acid[9,10]. This interference step is what has been commonly applied for genome engineering[1–3].

There is a wide diversity of Cas genes evolved by prokaryotes that are effectors of this adaptive immune system and these systems recognize a diverse set of target molecules (dsDNA, ssDNA, ssRNA)[11].One such effector complex that targets ssRNA is the Type III-A Csm complex[10–14]. The Csm complex is made up of five different proteins: Cas10, Csm2, Csm3, Csm4, and Csm5[10–14]. This complex has a guide RNA (gRNA) which recognizes a complementary single stranded RNA target[12–16]. After target sequence recognition, Csm catalyzes three simultaneous activities. Csm3 subunits initiate RNA cleavage at the target site, which knocks down gene expression of the target gene[12,15,16]. That RNA cleavage excises 18 or 24 nucleotides from the target region, depending on the number of Csm3 subunits in the effector complex and the efficiency of cleavage of each subunit[12,13]. Upon recognition and binding of the target RNA, the Cas10 HD nuclease domain becomes activated and cleaves nonspecific ssDNA to sheer DNA being actively transcribed or replicated[17,18], and the Cas10 Palm domain synthesizes cyclic oligoadenylates from ATP which go on to activate ancillary Type III-A Cas nucleases within the cell[19,20].

These activities, which are dependent upon specific RNA detection, have been used for a number of applications including knockdown of specific RNA transcripts within prokaryotic and eukaryotic cells, and *in vitro* detection of target RNA[18,21,22]. Here, we repurpose the type III-A CRISPR complex from *Streptococcus thermophilus* (SthCsm) to excise 18 nucleotides at a specific target site in the SARS-CoV-2 genome, and the ligation of the resulting cleavage products to produce an edited RNA genome. We predicted that the resultant edited genome could then be transfected into cells, as done previously with *in vitro* transcribed RNA genomes [4,5], but we have yet to succeed in this transfection. Once efficient transfection protocols are identified for large RNA genomes, this approach of RNA genome engineering will be much less time, resource, and energy intensive than previous approaches, allowing for RNA editing to be performed on the time scale of days, compared to the weeks or months needed for traditional approaches[4–6]. Further, the RNA editing strategy outlined here can be directly applied to any single-stranded RNA genome.

Results

Purified SthCsm efficiently cleaves target RNA

To test the utility of type III CRISPR-mediated RNA editing, we designed a synthetic CRISPR loci of *Streptococcus thermophilus* with four identical spacer sequences that are complementary to the SARS-CoV-2 N gene. These spacers were designed so that the resulting crRNA would be complementary to the SARS-CoV-2 N gene. This synthetic CRISPR array is coded on an expression vector that also includes a gene for the *Streptcoccus thermophilus* Cas6 protein, which processes nascent crRNA transcripts into mature crRNA molecules. Expression vectors coding for the protein subunits of the Csm complex were co-transfected along with the CRISPR plasmid into a protein expression strain of *Escherichia coli* and the resulting Csm complex was purified by affinity chromatography and size exclusion chromatography, as previously described[23]. The expression vectors used in this purification strategy are illulstrated in **Figure S3.1**. We found that the SthCsm complex assembled in two distinct sizes, associated with two distinctly sized CRISPR RNAs, which is consistent with a previous report[23] (**Figure 3.1B, C, D)**.

Incubation of target RNA with these SthCsm complexes resulted in the expected cleavage of target RNA at the site corresponding to the crRNA of the Csm complex, resulting in two distinct cleavage products, demonstrated by urea-PAGE in **Figure 3.1E**. The 5’ cleavage product is 1kb, and the 3’ product is 0.5kb. This cleavage was quantified by RT-qPCR using primers that flank the cleavage site (**Supplementary table 3.1**), such as the cleavage of RNA eliminates qPCR signal (**Figure 3.1F, G)**. The cleavage by Csm was about 98% efficient (**Figure 3.1F**).

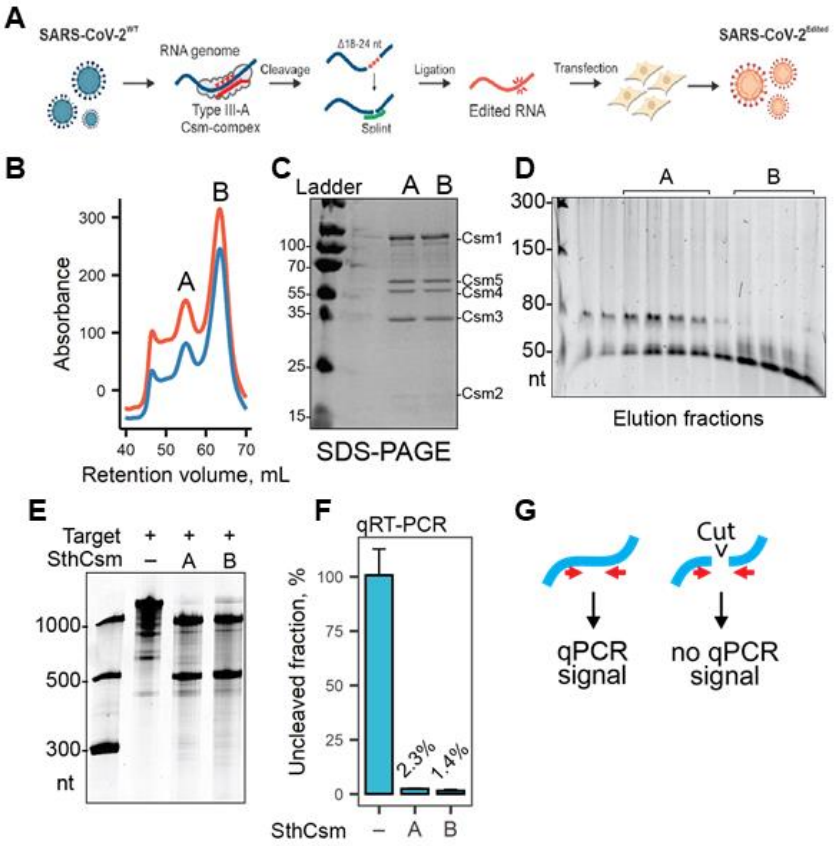


Figure 3.1 Purified SthCsm efficiently cleaves target RNA. (A) Outline of RNA editing. RNA genome editing is performed by cleaving target RNA by Csm, which excises 18-24 nucleotides. The resulting genome fragments are then ligated together via splint ligation, resulting in an edited genome with a target deletion. To then yield edited virions, this RNA genome must be transfected into cells and edited virions plaque purified. (B) Gel filtration profile of purified SthCsm complex. SthCsm complex purifies in two distinct sizes, labelled A and B. (C) Proteins of purified SthCsm complex were visualized by SDS-PAGE. (D) Urea-PAGE was used to visualize crRNA associated with fractions of each SthCsm complex. (E) Urea-PAGE of uncleaved target RNA and target RNA cleaved by SthCsm complexes. (F) RT-qPCR of uncleaved target RNA and target RNA cleaved by SthCsm complexes. Signal was normalized to uncleaved control by the ΔC_t method. (G) Design of RT-qPCR assay with target RNA shown in blue and qPCR primers in red.

Csm target RNA cleavage followed by splint ligation results in edited RNA in IVT RNA

After validating cleavage by Csm, we then set out to develop a protocol for ligating the resulting RNA fragments together. We planned to utilize commercially available RNA ligase enzymes to perform this ligation, most of which require a free 3' hydroxyl and 5' phosphate at the site of ligation. SthCsm cleaves RNA in a mechanism that is typical of a metal independent nuclease, that results in a 2'-3' cyclic phosphate group and a 5' hydroxyl group at the site of cleavage (**Figure 3.2A)**[24–27]. To remove this 2'-3' cyclic phosphate, we used T4 PNK, which replaces it with free 2' and 3' hydroxyl groups. This incubation with T4 PNK also adds a phosphate to the cleaved 5' terminus **(Figure 3.2A)**. These termini are compatible with T4 RNA ligase enzymes[28,29]. We also planned to use the RNA ligase enzyme, Rtcb, from the *E. coli* bacterium. This enzyme can ligate RNA molecules that have either a 3' hydroxyl terminus or a 2'-3' cyclic phosphate terminus to an RNA molecule with a 5' hydroxyl terminus[30].

RNA ligase I and RNA ligase II enzymes from the T4 bacteriophage have previously been applied for ligating together synthetic RNA molecules to yield high molecular weight or internally radiolabeled RNA molecules using splint ligation[28,29,31]. These previously described methods utilize synthetic single-stranded DNA oligonucleotide splints which are complementary to the two termini of RNA to be ligated together. To apply splint ligation to the RNA fragments cleaved by SthCsm, splints were designed that are complementary to the SthCsm cleavage products so that the ends of RNA to be ligated together are sequestered into close proximity via complementary base pairing to the DNA splint. Two different splint designs were utilized (**Figure 3.2B, Supplementary table 3.2)** which anneal to the RNA genome fragments and configure them into a bulge structure that imitates a cut tRNA anticodon loop (Splint A) or a

double-stranded structure that imitates nicked dsRNA (Splint B). These configurations imitate the natural substrates for T4 RNA ligase enzymes[28,29,32].

After annealing the RNA fragments to these DNA splints, hybrid structures were treated with RNA ligase enzymes T4 RNA ligase I, T4 RNA ligase II, or *E. coli* Rtcb RNA ligase. Ligation efficiency was assayed using the same RT-qPCR assay used previously (**Figure 3.1G**), with primers that amplify across the site of cleavage and ligation (i.e., RNA cleavage eliminates qPCR signal and ligation restores qPCR signal) **(Figure 3.2C)**. This assay demonstrated that PNK is essential for RNA ligase activity of T4 RNA ligase I and II, but not Rtcb. This assay further suggested that the most efficient splint and ligase combination was the double-stranded splint design (splint B) with T4 RNA ligase I.

To understand the sequence outcome of this cleavage and ligation protocol, long read Oxford Nanopore sequencing was used **(Figure 3.2D, E).** This analysis demonstrated that cleavage and ligation results in edited RNA with a deletion at the site of Csm cleavage, evidenced by a drop in sequencing depth at the Csm target site. Further, these sequencing results suggested that different splint designs could be used to generate different desired editing products. Splint design A facilitates the formation of either the 18 or 24 nucleotide deletion RNA molecule because either the Δ18 or the Δ24 cleavage products can anneal to this splint and be recognized by T4 RNA ligase I. In contrast, splint design B selects for either the Δ18 or the Δ24 nucleotide deletion RNA molecule, because only the specified cleavage product will anneal into the nicked, double-stranded RNA structure. Therefore, splint ligation using splint design A results in a mix of the 18 and 24 nucleotide deletion RNA, and splint ligation using splint design B results in either the 18 or 24 nucleotide deletion RNA, depending on how the splint is designed **(Figure 3.2E)**.

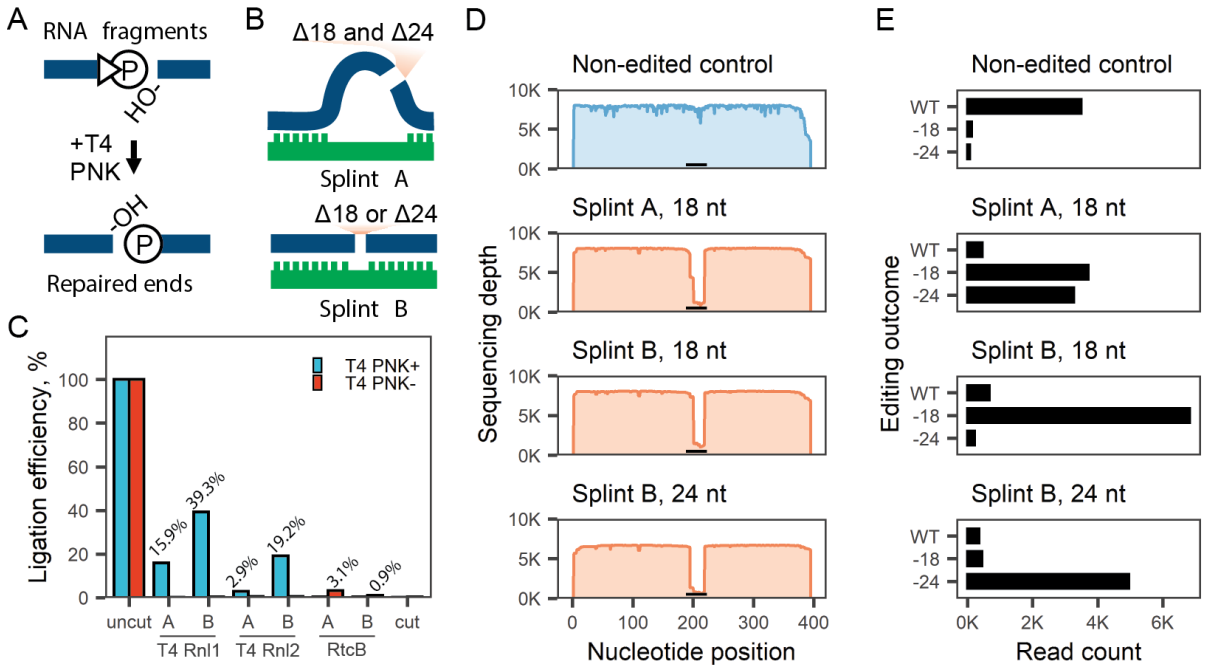


Figure 3.2 Ligation of Csm-cleaved RNA results in an edited product. (A) Csm cleavage of RNA results in a 2'-3' cyclic phosphate terminus and a 5' hydroxyl terminus at the site of cleavage. Treatment by T4 PNK repairs these termini to have a 3' hydroxyl group and a 5' phosphate group. (B) Single stranded DNA splints were designed to have 15-20 nucleotides of complementarity to the RNA termini to be ligated together. Hybridizing RNA to splint A results in a structure that imitates a cut tRNA anticodon loop. Hybridizing RNA to splint B results in a structure that imitates a nicked dsRNA structure. (C) RT-qPCR assay of the ligation products. Primers were used that flank the cleavage/ligation site so that cleavage eliminates qPCR signal and ligation restores qPCR signal. qPCR signal was normalized to the uncut control by the ΔC_t method. (D) Depth plots from deep sequencing the RNA cleavage/ligation products. Black bar indicates Csm target region of IVT RNA. (E) Quantification of editing outcomes shown in (D).

Csm cleavage and splint ligation of SARS-CoV-2 genome results in edited genome

After validating this approach for editing synthetic, IVT RNA, we then moved to editing the full-length RNA genome of SARS-CoV-2. This is a positive-sense, ~30-kilobase RNA genome. We targeted editing at the accessory gene, ORF7a, which has been shown to be a nonessential gene[5,33,34]. We chose ORF7a as the primary target in this study for two reasons. First, it had been previously demonstrated that ORF7a is not an essential gene, so an edited virus with no ORF7a would still be able to replicate. Second, it has also been shown that loss-of-function mutations in the ORF7a gene results in a fitness defect[35,36].

To pivot this editing platform validated on IVT N gene RNA to editing the ORF7a gene in the SARS-CoV-2 genome, we purified a new SthCsm complex with a guide RNA that is complementary to the ORF7a gene. We designed the gRNA of this complex so that the excision of 18 nucleotides by SthCsm cleavage followed by splint ligation would result in a novel stop codon after amino acid position 22 (**Figure 3.3A)**. We then used this SthCsm complex to perform the editing on full-length SARS-CoV-2 RNA purified from infected cell culture. Cleavage by SthCsm was >97% efficient, and splint ligation, using a double-stranded splint, was ~12% efficient (**Figure 3.3B**).

To preliminarily confirm editing, the qPCR product from **Figure 3.3B** was visualized via PAGE (**Figure 3.3C**). This visualization demonstrated the formation of a smaller qPCR product, indicative of the expected target 18 nucleotide deletion. To further confirm editing, we used deep sequencing. Deep sequencing confirmed that this CRISPR-based editing approach resulted in the expected 18 nucleotide deletion, resulting in a novel stop codon after amino acid position 22 (**Figure 3.3D, E**).

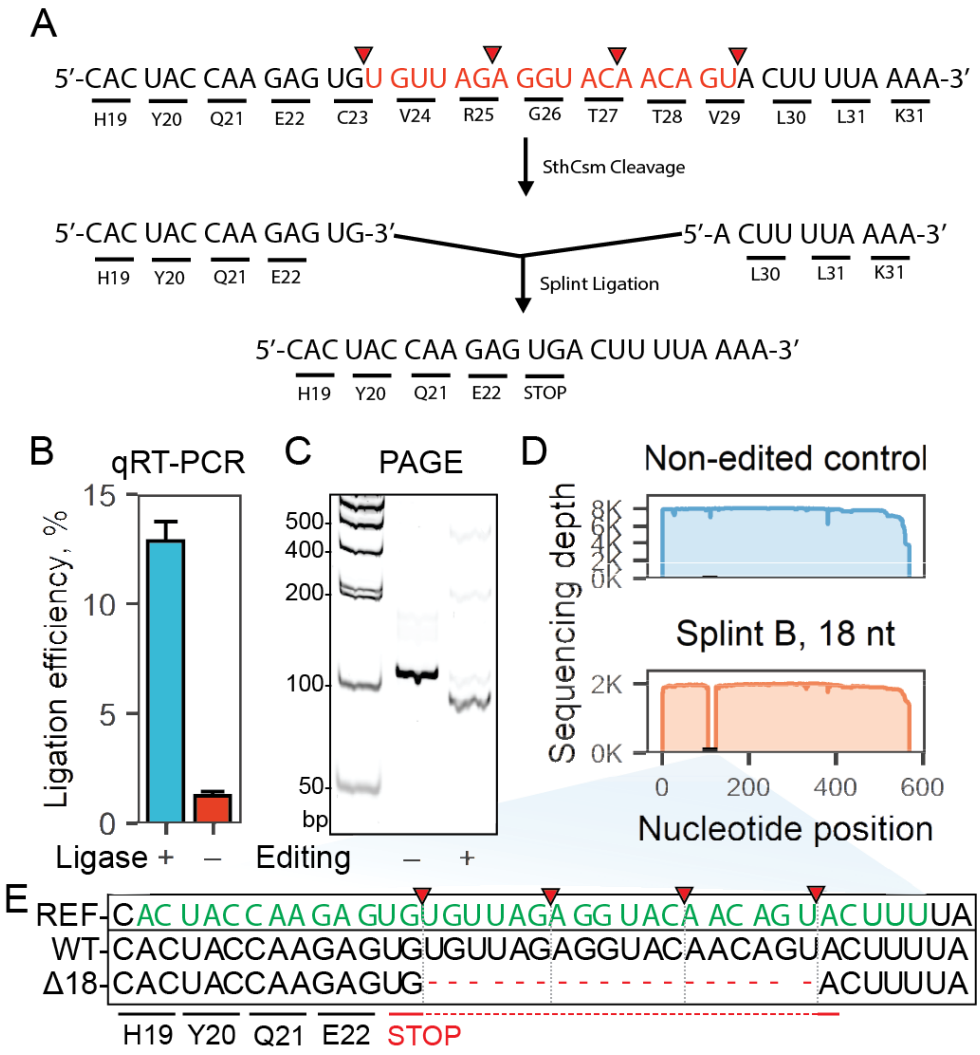


Figure 3.3 Editing of the SARS-CoV-2 genome. (A) Schematic of editing strategy. ORF7a gene sequence shown at editing target site. Csm cleavage sites shown as red triangles, nucleotides excised shown in red. (B) RT-qPCR assay of ligation reaction using primers that flank the cleavage and ligation sites. qPCR signals normalized to uncleaved control. (C) qPCR products were visualized on PAGE, which indicated the formation of the edited product. The expected qPCR product of the WT control is 116bp, and the expected qPCR product of the edited genome is 98bp. (D) Deep sequencing of the WT and edited genomes. Black bar indicates target site of SthCsm complex. (E) Sequence outcome of editing. Red triangles indicate cut sites of SthCsm complex. Green font in reference indicates RNA complementary to SthCsm crRNA.

Transfection of SARS-CoV-2 genome has not resulted viral replication

After editing the RNA genome of SARS-CoV-2, we then planned to transfect this genome into mammalian cells via electroporation and generate edited SARS-CoV-2 virions. To accomplish this, we followed recently published methods of transfecting SARS-CoV-2 genomic RNA into mammalian cells established for a reverse genetics system of SARS-CoV-2[4]. This method for transfection was developed using synthetic, *in vitro* transcribed RNA genomes of SARS-CoV-2, not RNA purified from infected cells. As such, these authors were able to use much higher levels of genomic RNA than we were able to purify from infected cell culture and edit. Nevertheless, we attempted transfection of the genomic RNA into mammalian cells, following the previously established protocol, and assaying for viral replication by RT-qPCR (**Figure 3.4B**). We transfected 1 pmol WT, uncleaved genomic RNA (sample 1), 1 pmol ORF7a edited genomic RNA (sample 2), and 3 pmol ORF7a edited genomic RNA (sample 3).

Further, we were curious if transfection of cleaved RNA was sufficient to produce edited virions with deletions at the cleavage site, as a similar phenomenon of viral RNA recombination and viral rescue has previously been reported in picornaviruses and flaviviruses (i.e., transfection of 5' and 3' genome segments result in recombined, full length RNA genome)[37,38]. To test this, we transfected 1 pmol cleaved genomic RNA directly after cleavage (sample 4), and after treatment with T4 PNK, annealed to a double-stranded splint (sample 5), or annealed to a single-stranded splint (sample 6).

We found that this transfection failed to result in viral replication in any of these samples, evidenced by a decrease of viral RNA in the supernatant of transfected cells over time. We expect that this may be due to two reasons. First, electroporation of such a large RNA molecule is likely very inefficient, and only a small fraction of the total RNA genomes is expected to be

successfully transfected into cells. Second, the large RNA genome of SARS-CoV-2 is likely susceptible to degradation through the process of viral RNA purification from cells, and the process of editing. As such, it is likely that much of the RNA used for electroporation was not full-length genome and could not produce infectious virions. This failure to produce any virions is evidenced by qPCR that demonstrates decreasing levels of viral RNA in the supernatant of electroporated cells (**Figure 3.4B**).

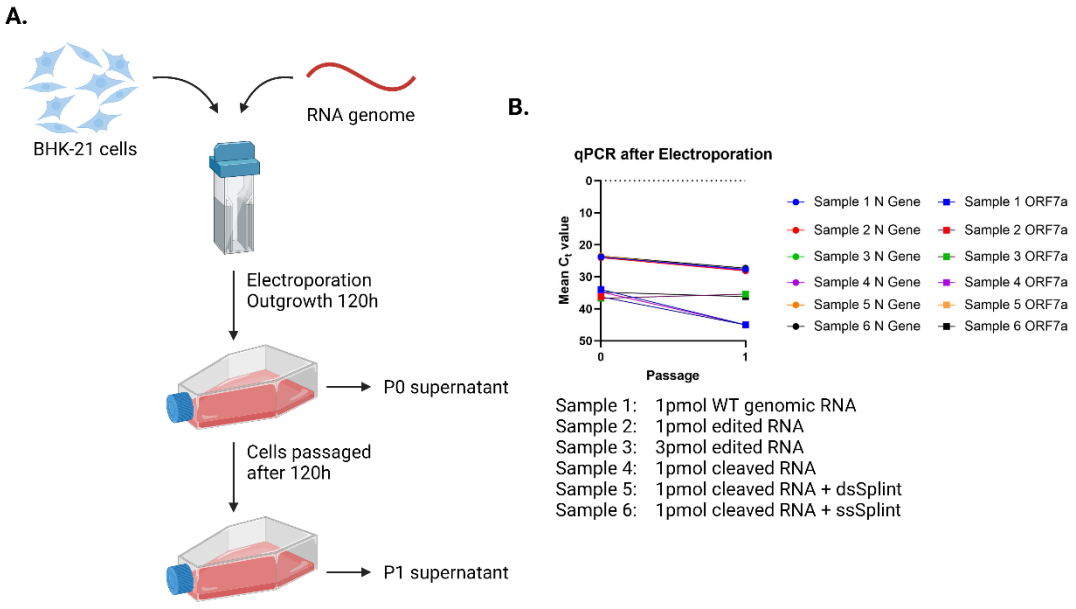


Figure 3.4 Transfection of edited genome has not resulted in viral replication. (A) Outline of transfection procedure. BHK-21 cells were transfected with either 1pmol or 5pmol of the RNA genome using the Mirus transfection system. Following transfection, BHK-21 cells were cocultured with VeroE6 cells which are permissive to SARS-CoV-2 infection and replication. After 120 hours of growth, supernatant was assayed for viral RNA and cells were split and passaged. Supernatant from P1 was assayed 120 hours after this passage. (B) RT-qPCR of supernatant from P0 and P1, circles are C_t values from N gene target, squares are C_t values from ORF7a target.

Materials and Methods

Plasmids for SthCsm complex

Synthetic gene sequences (gBlocks) coding for SthCsm complex were ordered and synthesized by ThermoFisher GeneArt gene synthesis. *E. coli* codon optimized genes for SthCsm3, SthCsm4, and SthCsm5 were coded for on one plasmid, SthCsm1 and SthCsm2 on a second plasmid, and the SthCRISPR array with SthCas6 on a third. The SthCsm3 protein includes a fused TwinStrep tag on its C-terminus. These proteins are under the control of a T7 promoter.

Purification of SthCsm complex

SthCsm was reconstituted using BL21(DE3) arabinose-inducible strain of *E. coli* as previously described[12,18,21] and purified by size exclusion chromatography. Briefly, three expression vectors coding for the gRNA, Cas6, Cas10, and Csm 2-5 were co-transformed into arabinose inducible BL21(DE3) *E. coli* chemically competent cells. These cells were grown to an OD_{600} of 0.5 and protein expression was induced with 250mM isopropylthio-β-galactoside (IPTG) and 0.1% L-arabinose then grown overnight at 16°C. Cells were lysed in buffer containing 20 mM Tris-HCl (pH 8.5), 500 mM potassium glutamate,1 mM DTT, 1 mM EDTA, and 1X Halt Protease Inhibitor (Thermo Scientific product number 1861279). SthCsm complex was first purified by binding to StrepTrap HP resin (Cytiva). Resin was washed with 30 column volumes of lysis buffer. Proteins were eluted in lysis buffer supplemented with 2.5 mM desthiobiotin and concentrated with centrifugation in 100K MWCO Corning Spin-X concentrators at 4°C. SthCsm complexes were further purified by size exclusion chromatography using a HiLoad Superdex 200 26/600 size-exclusion column (Cytiva) in 20 mM Tris-HCl pH

7.5, 1 mM DTT, 400 mM potassium glutamate, 5 % glycerol. Fractions containing the SthCsm complex were analyzed by SDS-PAGE and urea-PAGE to visualize protein components and crRNA sizes, respectively. The two species of SthCsm complex, A and B, were pooled separately, concentrated by centrifugation in 100K MWCO Corning Spin-X concentrators at 4°C, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

Cloning of N gene for *in vitro* transcription

SARS-CoV-2 RNA genome was reverse transcribed by SuperScript IV reverse transcriptase (ThermoFisher Scientific catalogue number: 1890010) following supplier's protocol. The SARS-CoV-2 N gene was cloned by PCR amplifying cDNA of the viral RNA genome by Q5 polymerase PCR (NEB catalogue number M0491) using primers JN_N1.T7.F and nCoV-96_R (**Supplementary table 3.1**), which introduces a T7 promoter site at the 5' end. *In vitro* transcription was performed using the MEGAscript T7 transcription kit (Invitrogen catalogue number AM133) following manufacturer's protocol. RNA was purified using the 500 ug Monarch RNA cleanup kit following manufacturer's protocol (NEB catalogue number T2050).

Cleavage of *in vitro transcribed* target RNA

RNA of the SARS-CoV-2 N gene was produced by *in vitro* transcription (IVT) and used as a target for the Csm complex (guide: AUUGCGACACGCUGAAGCGCUGGGGG CAAAUUGUGCAAUU). Cleavage was done at 37ºC for 30 minutes in buffer composed of 20 mM HEPES (pH 7.5), 50 mM KCl, 0.1 mg/ml BSA, 10 mM Mg-acetate, and 1 U/uL RNAse inhibitor. Following cleavage, RNA was purified using 10µg Monarch RNA cleanup kit (NEB

catalogue number: T2030). Cleavage products were resolved with denaturing 8M urea-PAGE (10%), and cleavage efficiency was quantified by RT-qPCR.

RT-qPCR assay for cleavage and ligation

After purification, RNA was reverse transcribed by SuperScript IV reverse transcriptase (ThermoFisher Scientific catalogue number: 1890010) following supplier's protocol. Before reverse transcription, RNA from all groups were normalized to 10ng/µL. cDNA was then diluted to 1:100 in nuclease free water before qPCR. For qPCR, PowerUp SYBR Green Master Mix (ThermoFIsher Scientific catalogue number: A25741) was used following manufacturer's protocol, along with custom qPCR primers, listed in **Supplementary table 3.1**. For N gene IVT RNA, N2.1_F and N2.1_R were used. For ORF7a editing in SARS-CoV-2 RNA, JN_ORF7a.F2 and JN_ORF7a.R1 were used.

qPCR reactions were pipetted in technical triplicates into MicroAmp Optical 96-Well Reaction Plate (Applied Biosystems reference number: 4306737) and the reactions run using the QuantStudio 3 Real Time PCR system (ThermoFisher Scientific catalogue number: A28567) with the following cycle settings: ramped up from room temperature at 1.6°C/second to 50°C, held for 2 minutes, ramped up to 95°C at 1.6°C/second and held for 10 minutes. Reactions were then cycled for 40 cycles by denaturing at 95°C, ramping down to 60°C at 1.6°C/second, then holding for 1 minute while collecting fluorescence data, then returning to denaturation step. Nuclease free water was used as a non-target control.

C_t values were averaged across technical replicates. ΔC_t method was used to quantify the relative amount of target RNA, normalized against an uncut control. Biologic replicates (n=2) were averaged, and data was visualized using GraphPad Prism.

Dephosphorylation of 2’-3’ cyclic phosphate

The cleavage of RNA by SthCsm results in a 5’ cleavage product which has 2’-3’ terminal cyclic phosphate. This phosphate group blocks the terminal nucleotide from participating in many biochemical reactions, including ligation with T4 RNA ligase. To remove the phosphate, the RNA was treated with T4 polynucleotide kinase (NEB catalogue number: M0201) in the manufacturer supplied T4 PNK buffer and incubated for 40 minutes at 37^{o}C in the absence of ATP.

Phosphorylation of 5’ end of 3’ cleavage product

The cleave of RNA by SthCsm results in a 3’ cleavage product which has a 5’ hydroxyl group. This hydroxyl group is incompatible with the ligation reaction catalyzed by T4 RNA ligase enzymes. To replace this hydroxyl group with a phosphate group, the dephosphorylation reaction was supplemented with 1mM ATP and incubated for 30 minutes at 37^{o}C.

Splint design and annealing

Cleaved RNA fragments were ligated through a directed method using splint ligation. DNA oligonucleotides splints were designed as previously described[28,29] to have 18-25 nucleotides of complementarity to the 5’ cleavage product and 18-25 nucleotides of complementarity to the 3’ cleavage product with compatible melting temperatures. Splint design A was designed such as to have 4-10 nucleotides of RNA overhang of the 5’ cleavage product and 2 nucleotides overhang of the 3’ cleavage product. The resulting complex resembles the natural tRNA substrate of T4 RNA ligase I[29]. Splint design B was designed to have no nucleotide

overhang and position the RNA to resemble the nicked dsRNA substrate for RNA ligase II (designs outlined in **Figure 3.1B**, splint sequences in **Supplementary table 3.2**).

Splints were mixed with the cleaved RNA fragments following the T4 PNK treatment in a 1:1 molar ratio of splint to RNA in 1X T4 RNA ligase buffer (NEB). The mixture was heated to 98^{o}C for 2 minutes then ramped down to 4^{o}C over 45 minutes.

Ligation of cleavage products

Following modification of 3' and 5' ends and annealing with either splint design A or B, the RNA cleavage products were ligated with T4 RNA ligase I (NEB catalogue number: M0204), T4 RNA ligase II (NEB catalogue number: M0239), or *E. coli* Rtcb RNA ligase (NEB catalogue number: M0458) following manufacturer protocol. Following ligation, RNA was purified by 10µg Monarch RNA cleanup kit (NEB catalogue number: T2030).

RT-PCR and RNA Sequencing

For sequencing edited IVT RNA or SARS-CoV-2 genomic RNA, sample was reverse transcribed using SuperScript IV (Thermo Fisher Scientific) according to the supplier's protocol. The protocol developed by ARTIC Network, with some modifications, was used to generate and sequence amplicon library that covers the edited region of the IVT RNA or the SARS-CoV-2 genome on Oxford Nanopore using ligation sequencing kit (SQK-LSK109) (https://artic.network/ncov-2019). Briefly, PCR reactions were performed with custom primers (**Supplementary Table 2.3**) flanking the site of editing (nCOV-96_L, nCOV-96_R for N gene IVT, JN_ORF7a_F1 and JN_ORF7a_R1 for editing ORF7a in SARS-CoV-2 genome) using Q5 High-Fidelity DNA Polymerase (New England Biolabs). Reactions were performed with the following thermocycling conditions: 98°C for 2min, 30 cycles of 98°C for 15 s and 65^{o}C for 5

min. Resulting amplicons were purified with DNA Clean & Concentrator kit (Zymo Research) and then used for library preparation. Next, 50 ng of each amplicon was used to prepare sequencing libraries with NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs catalogue number: E7645S). DNA was end-repaired with NEBNext Ultra II End Prep Enzyme Mix (New England Biolabs catalogue number: E7546S) and barcoded with Native Barcoding Expansion kit (ONT, EXP-NBD104) using Ultra II Ligation Module (New England Biolabs catalogue number: E7595S). Barcoded DNA fragments were pooled together, purified with magnetic beads (Omega Bio-tek, M1378-01), and Nanopore sequencing adaptors were ligated. After clean-up, 20 ng of multiplexed library DNA was loaded onto the MinION flowcell (FLO-MIN 106) for sequencing. A total of 10-20 K reads were collected per sample. Nuclease-free water was used as a negative control for library preparation and sequencing. Non-specific amplification in negative control was additionally checked using an agarose gel. N gene IVT RNA or ORF7a from SARS-CoV-2 Isolate USA-WA1/2020 (BEI, NR-52281) was used as a positive control and reference for genome sequencing.

RNA sequence Assembly

Nanopore reads were basecalled in high-accuracy mode and demultiplexed with MinKNOW software (Oxford Nanopore Technologies). Demultiplexed reads were aligned with minimap2 v2.17-r954-dirty (ax map-ont mode) to the reference sequence. Resulting alignments (BAM files) were sorted and indexed using samtools v1.13. Sequencing depth was calculated with samtools *depth* function and plotted with ggplot2 package in RStudio. To analyze RNA editing efficiency, sequencing errors were corrected in Nanopore reads using isONcorrect computational method[39]. Corrected reads were aligned to the reference with minimap2 and

analyzed with CrispRVariants package in RStudio. Frequencies of RNA editing outcomes were quantified and plotted with ggplot2 package in RStudio.

Cell culture

Vero E6 cells (ATCC® cat. #CRL-1586) were maintained at 37°C and 5% CO2 in Eagle's Modified Eagle Medium (EMEM, ATCC®, cat. #30-2003) supplemented with 10% fetal bovine serum (FBS, ATLAS Biologicals, Lot. #F31E18D1) and 50 I.U./mL penicillin and 50 µg/mL streptomycin.

BHK-21 cells (ATCC® cat. #CCL-10) were maintained at 37°C and 5% CO2 in Eagle's Modified Eagle Medium (EMEM, ATCC®, cat. #30-2003) supplemented with 5% fetal bovine serum (FBS, ATLAS Biologicals, Lot. #F31E18D1) and 50 I.U./mL penicillin and 50 µg/mL streptomycin.

SARS-CoV-2 infected cell culture and RNA genome purification

Vero E6 cells were inoculated in SARS-CoV-2 in a bio-safety level-3 laboratory at an MOI of 0.1 following standard protocols[40,41]. Vero E6 cells were grown in T75 flasks, as described above in DMEM media supplemented with 10% FBS and 1% penicillin/streptomycin. At 80% confluence, cells were infected with SARS-CoV-2 isolate WA-1 at an MOI of 0.1. Virus was propagated for 5 days, which were then harvested by aspirating supernatant from infected cell culture media. Supernatants were centrifuged at 1,000 RCF for 5 minutes to remove cellular debris. Virus in supernatant were inactivated by heating culture media to 60°C for 30 minutes before removing supernatant from BSL3. Virions were pelleted by ultracentrifugation at 100,000xg, after which the pellet was resuspended in 1X tris-EDTA buffer. RNA was extracted

from viral lysate by TRIzol RNA extraction (Invitrogen catalogue number: 15596026) following supplier protocol.

Transfection of RNA genome into BHK-21 cells

Transfection of purified SARS-CoV-2 RNA was done as previously described[4], with slight modifications of number of cells in suspension and cuvette size, as to follow manufacturer's protocol of the Mirus EZPorator Electroporation System (MirusBio catalogue number: MIR 51000). Briefly, BHK-21 cells were grown to 80-90% confluence in a T75 flask as described above. Cells were harvested and pelleted by centrifugation at 420xg for 5 minutes at 4°C. The cell pellet was then resuspended to 10 x 10^6 cells / mL in cold PBS and kept on ice. In a BSL-3 lab, 2µg N gene mRNA and either 1pmol or 5pmol SARS-CoV-2 genomic RNA was mixed with 100µL of the BHK-21 cell suspension (10 x 10^5 cells) and added to a cold 0.2cm Mirus electroporation cuvette. Electroporator was then pulsed at 150 V. Cells were recovered by flushing cuvette with Vero E6 cell culture DMEM media, which was then added to flasks with Vero E6 cells. After cells seeded to plate, culture supernatant was aliquoted and assayed for SARS-CoV-2 RNA by qPCR. Cells were checked for cytopathic effect (CPE) every day, and supernatant was assayed by qPCR 5 days post transfection, when cells were passaged to a new cell culture flask.

Discussion

Here, we show the development of SthCsm RNA editing, and apply it to perform a targeted edit in the RNA genome of SARS-CoV-2. This CRISPR-based approach demonstrates a leap in RNA editing technology and will facilitate future genetic studies of viral genetics and virology in general. In the context of the SARS-CoV-2 pandemic, and rapidly emerging novel genetic variants, traditional approaches to viral genetic studies are time consuming and not well suited to generating a rapid understanding of virus biology[4–6]. Currently, to study a novel viral variant, genetic studies require the generation of cDNA from the viral RNA genome, performing site directed genetic manipulations of the DNA, then *in vitro* transcribing the resultant synthetic DNA into an RNA genome, and finally transfecting the IVT RNA into mammalian cells[4–6]. The CRISPR-Csm approach outlined here is more streamlined, and only requires that new gRNA and DNA splints are designed for novel viral variants. RNA editing using SthCsm cleavage followed by splint RNA ligation is a simplified, robust approach for generating desired edits in an RNA genome. More efficient transfection protocols will enable facile manipulation of the SARS-CoV-2 genome and other RNA viruses.

This RNA editing pipeline may also enable researchers to integrate transgenes into viral genomes. This may be done by adding synthetic RNA coding for such a transgene to the cleaved RNA genome fragments and direct the ligation of the synthetic RNA into the viral genome by splint ligation. Future studies will focus on developing this gene integration strategy further and applying it to RNA viruses and RNA bacteriophage.

This RNA editing approach may even evolve to a point in which it can enable new strategies for RNA virotherapy, as it may be used to generate viruses that express therapeutic transgenes or oncolytic immunostimulatory genes in tumor cells. Currently, most viruses

approved or in trials for gene therapy and oncolytic virotherapy are DNA viruses, as they are easier to genetically engineer. The majority of gene therapy trials utilize viruses from the DNA virus family, adeno-associated viruses[42]. The only FDA approved oncolytic virotherapy is Talimogene Laherparepvec, a genetically engineered oncolytic virus based on the DNA virus, herpes simplex virus-1[43,44]. The advancement of novel CRISPR-based technologies for engineering RNA viruses could make it possible to rapidly genetically engineer new therapeutic viruses based on Coronaviruses, paramyxoviruses, or any other single-stranded RNA virus. This is especially promising for gene therapy applications in which transient expression is optimal (i.e., therapeutic CRISPR gene editing).

References Cited

1. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, (2014).
2. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
3. Wilkinson, R. A., Martin, C., Nemudryi, A. A. & Wiedenheft, B. CRISPR RNA-guided autonomous delivery of Cas9. *Nat Struct Mol Biol* **26**, 14–24 (2019).
4. Xie, X. *et al.* Engineering SARS-CoV-2 using a reverse genetic system. *Nature Protocols* **16**, 1761–1784 (2021).
5. Thi Nhu Thao, T. *et al.* Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).
6. Ye, C. *et al.* Rescue of SARS-CoV-2 from a Single Bacterial Artificial Chromosome. *mBio* **11**, (2020).
7. Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).
8. Hodcroft, E. B. *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **595**, 707–712 (2021).
9. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
10. Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J. & Staals, R. H. J. Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Biosci Rep* **38**, BSR20170788 (2018).
11. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67–78 (2017).

12. Mogila, I. *et al.* Genetic Dissection of the Type III-A CRISPR-Cas System Csm Complex Reveals Roles of Individual Subunits. *Cell Rep* **26**, 2753-2765.e4 (2019).
13. Jia, N. *et al.* Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Mol Cell* **73**, 264-277.e5 (2019).
14. 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).
15. Benda, C. *et al.* Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. *Molecular Cell* **56**, 43–54 (2014).
16. Spilman, M. *et al.* Structure of an RNA Silencing Complex of the CRISPR-Cas immune system. *Mol Cell* **52**, 10.1016/j.molcel.2013.09.008 (2013).
17. Lin, J., Shen, Y., Ni, J. & She, Q. A type III-A CRISPR–Cas system mediates co-transcriptional DNA cleavage at the transcriptional bubbles in close proximity to active effectors. *Nucleic Acids Research* **49**, 7628–7643 (2021).
18. Santiago-Frangos, Andrew. Intrinsic Signal Amplification by Type-III CRISPR-Cas Systems Provides a Sequence-Specific Viral Diagnostic | medRxiv. https://www.medrxiv.org/content/10.1101/2020.10.14.20212670v1.full.
19. 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).
20. 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).
21. Fricke, T. *et al.* Targeted RNA Knockdown by a Type III CRISPR-Cas Complex in Zebrafish. *CRISPR J* **3**, 299–313 (2020).

22. Zebec, Z., Manica, A., Zhang, J., White, M. F. & Schleper, C. CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. *Nucleic Acids Res* **42**, 5280–5288 (2014).

23. Tamulaitis, G. *et al.* Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell* **56**, 506–517 (2014).

24. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529–536 (2011).

25. Hochstrasser, M. L. & Doudna, J. A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences* **40**, 58–66 (2015).

26. Nam, K. H. *et al.* Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure* **20**, 1574–1584 (2012).

27. Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).

28. Stark, M. R. & Rader, S. D. Efficient Splinted Ligation of Synthetic RNA Using RNA Ligase. in *Spliceosomal Pre-mRNA Splicing: Methods and Protocols* (ed. Hertel, K. J.) 137–149 (Humana Press, 2014). doi:10.1007/978-1-62703-980-2_10.

29. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12**, 2014–2019 (2006).

30. Tanaka, N. & Shuman, S. RtcB Is the RNA Ligase Component of an Escherichia coli RNA Repair Operon *. *Journal of Biological Chemistry* **286**, 7727–7731 (2011).

31. Bain, J. D. & Switzer, C. Regioselective ligation of oligoribonucleotides using DNA splints. *Nucleic Acids Res* **20**, 4372 (1992).
32. Amitsur, M., Levitz, R. & Kaufmann, G. Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. *EMBO J* **6**, 2499–2503 (1987).
33. Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **182**, 429-446.e14 (2020).
34. Xie, X. *et al.* An Infectious cDNA Clone of SARS-CoV-2. *Cell Host & Microbe* **27**, 841-848.e3 (2020).
35. Addetia, A. *et al.* Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol* **129**, 104523 (2020).
36. Nemudryi, A. *et al.* SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Reports* **35**, (2021).
37. Bentley, K. & Evans, D. J. Y. 2018. Mechanisms and consequences of positive-strand RNA virus recombination. *Journal of General Virology* **99**, 1345–1356.
38. Agol, V. I. & Gmyl, A. P. Emergency Services of Viral RNAs: Repair and Remodeling. *Microbiol Mol Biol Rev* **82**, (2018).
39. Sahlin, K. & Medvedev, P. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nat Commun* **12**, 2 (2021).
40. Harcourt, J. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg Infect Dis* **26**, 1266–1273 (2020).
41. Loveday, E. K. *et al.* Effect of Inactivation Methods on SARS-CoV-2 Virion Protein and Structure. *Viruses* **13**, 562 (2021).

42. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **31**, 317–334 (2017).

43. Conry, R. M., Westbrook, B., McKee, S. & Norwood, T. G. Talimogene laherparepvec: First in class oncolytic virotherapy. *Human Vaccines & Immunotherapeutics* **14**, 839–846 (2018).

44. Goradel, N. H. *et al.* Oncolytic virotherapy: Challenges and solutions. *Current Problems in Cancer* **45**, 100639 (2021).

Supplementary Figures

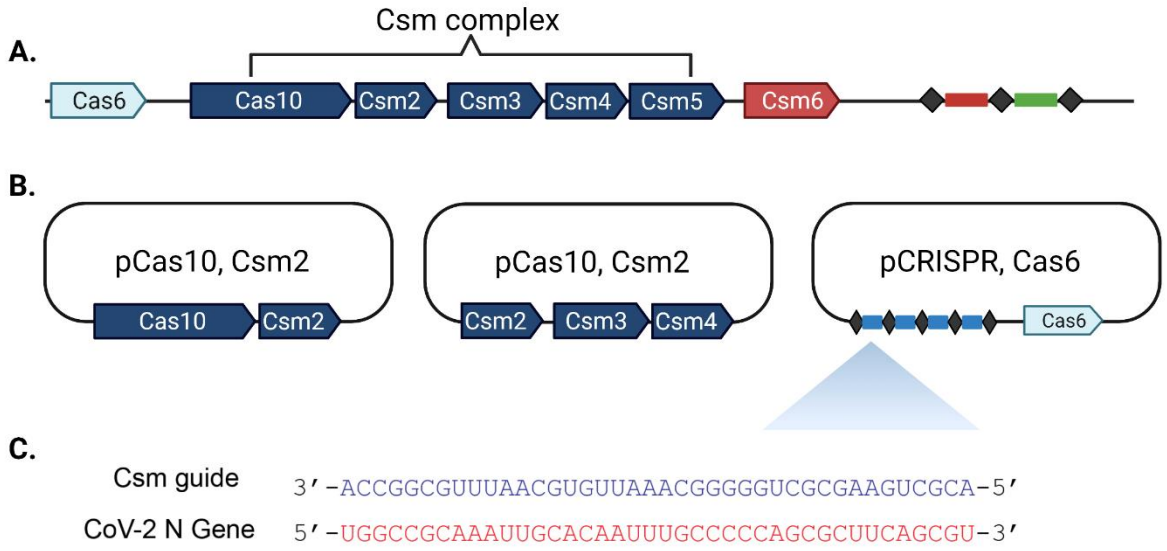


Figure S3.1 Purification strategy for SthCsm. (A) Genetic organization of the Type III-A CRISPR system of *Streptococcus thermophilus*. (B) Expression vectors were produced that code for each component of the SthCsm complex, along with an expression vector with a synthetic CRISPR that includes four identical spacers that are complementary to the SARS-CoV-2 N gene. (C) Four identical CRISPR spacers were designed to have 39 nucleotides complementary to the target RNA.

Supplementary Tables

Primer Name	**Primer sequence (5'-3')**
JN_N1.T7.F	GCTCGATAATACGACTCACTATAGGGCTGCTGAGGCTTCTAAGA AGCCTC
nCoV-96_R	TAGGCTCTGTTGGTGGGAATGT
N2.1F	GGGACCAGGAACTAATCAGAC
N2.1R	TGTGACTTCCATGCCAATG
JN_ORF7a.F2	AGATGAAGAGCAACCAATGGAGA
JN_ORF7a.R1	CCGTCAGGACAAGCAAAAGC

Table S3.1. Primers used for cloning of N gene RNA (JN_N1.T7.F + nCoV-96_R) and qPCR assays (N2.1F and N2.1R for N gene; JN_ORF7a.F2 and JN_ORF7a.R1 for ORF7a).

Primer Name	Splint sequence (5'-3')
N gene dsSplint 18	GAAGAACGCTGAAGCGATTTGCGGCCAATGTTT
N gene dsSplint 24	GAAGAACGCTGAAGCGGGCCAATGTTTGTAATCAGT
N gene ssSplint	TCCGAAGAACGCTGAAGTTTGTAATCAGTTCCTTGTCTG
ORF7a dsSplint 18	AGAGCAAGGTTCTTTTAAAAGTCACTCTTGGTAGTGATAAAGC

Table S3.2. DNA oligonucleotide splint designs. Red portion of the DNA sequence corresponds to the region complementary to the 5' terminus of the 3' RNA cleavage product. Blue sequence is complementary to the 3' terminus of the 5' RNA cleavage product.

Primer Name	**Primer sequence (5’-3’)**
nCoV-96_L	GCCAACAACAACAAGGCCAAAC
nCoV-96_R	TAGGCTCTGTTGGTGGGAATGT
JN_ORF7a_F1	AGATGAAGAGCAACCAATGGAGA
JN_ORF67ab_R	GTGATGATTCCTAAGAAAACAAGAAATTTCATGTTCGT

Supplementary table 3.3 primers used for Nanopore amplicons. nCOV-96 primers were used for N gene Nanopore amplicons, JN_ORF7a_F1 and JN_ORF67ab_R were used for ORF7a Nanopore amplicons.

CHAPTER FOUR

SUBSTRATE PREFERENCES OF TYPE III-A ANCILLARY CAS NUCLEASES AND THEIR IMPACT UPON BACTERIAL GROWTH KINETICS

Abstract

Type III CRISPR-Cas systems are the only group of CRISPR systems that utilizes a signaling pathway to activate ancillary effectors. In addition to the CRISPR RNA (crRNA)-guided, sequence-specific binding and cleavage of target RNAs, such as phage transcripts, the type III-A effector, Csm, initiates a signaling pathway that activates ancillary nucleases, triggering nonspecific RNA and DNA cleavage. This nonspecific nuclease activity likely induces bacterial cell dormancy, thus protecting the bacterial population from uncontrolled phage replication and spread. We show here that the Type III-A CRISPR system of *Thermus thermophilus* encodes for three ancillary nucleases, each of which are activated by the cyclic oligoadenylate (cOA) molecules produced by Csm to cleave nucleic acids. The activation patterns of these ancillary effectors are nonredundant, and likely signify independent pathways for inducing growth arrest to block phage replication. We attempted to demonstrate the effect of activation of these ancillary CRISPR effectors upon growth kinetics of a heterologous expression system and planned to subsequently identify which RNA or DNA species are preferentially cleaved by these nucleases. We found only a subtle change in growth kinetics between active and inactive immune systems in this expression system, suggesting that this heterologous expression system is not a faithful model for studying the Type III-A CRISPR system of *T. thermophilus*.

Introduction

CRISPR systems are adaptive immune defenses present in about half of all prokaryotic genomes that have been sequenced to date[1]. These defenses evolved to block infection by mobile genetic elements such as bacteriophage and parasitic plasmids. Because these systems evolved to counter such diverse and rapidly evolving agents, CRISPR systems exhibit a massive amount of diversity[2]. This diversity is borne out in the two classes (Class I and Class II) and the six types (I-VI) of CRISPR systems, distributed between the two classes. Some of the most intricate CRISPR systems are those of the Type III group[3,4].

Like all other CRISPR systems, Type III systems utilize an RNA-guided interference enzyme complex to surveil the cell, searching for nucleic acid complementary to its guiding CRISPR RNA (crRNA)[4]. Type III systems stand out among CRISPR immune systems though because instead of targeting DNA, like most CRISPR systems, the Type III interference complexes (Csm for Type III-A systems, Cmr for Type III-B systems) target RNA[5]. The only other RNA-targeting CRISPR systems are from the Type VI group[6]. After Csm identifies target RNA within the cell, it cleaves the RNA at the site complementary to the crRNA[4–8]. This target recognition allosterically activates two domains within the Cas10 subunit of the Csm complex, the HD nuclease and Palm polymerase domains[4–8]. The HD domain, when activated, is a nonspecific single-stranded DNase which is proposed to cleave DNA that is being actively transcribed[4]. The Palm domain of Cas10 is a polymerase-cyclase which, when activated, synthesizes cyclic oligoadenylate (cOA) molecules from ATP[3,5,9]. These cyclic messenger molecules bind to ancillary nucleases within the cell, activating them. These nucleases are predicted to be indiscriminate RNases or DNases[3,9–11].

The ancillary nucleases of Type III systems are predicted to be cOA activated nucleases because they contain a CRISPR-Associated Rossmann Fold (CARF) domain[10]. CARF domains bind to cOA produced by the Cas10 subunit of the Type III effector complex[10]. Such binding allosterically activates effector domains of the enzymes[10,12]. In the Type III-A system we focus on here, from the *Thermus thermophilus* bacterium, CARF domains are fused to nucleases[13,14]. Interestingly, this CRISPR system has three CARF-nucleases, each of which can target ssRNA, and one that can also cleave dsDNA[3,13–15]. This CRISPR locus is illustrated in Figure 4.1A. The aim of this project is to determine the substrate preferences of these ancillary nucleases when they are activated by cOA *in vivo*, and to identify the effect of such nuclease activation on bacterial growth kinetics. We hypothesize that these nucleases hold nonredundant substrate preferences, that they signify independent strategies of inducing growth arrest upon recognition of bacteriophage transcripts.

We hypothesize that activation of these ancillary nucleases will result in growth arrest of the bacterial host, as activated RNAses will destroy essential RNA transcripts, tRNAs, or rRNAs and activated DNases within the cell will result in sheering of the bacterial genome. To test this hypothesis, we developed heterologous expression systems, in which *E. coli* expresses the Csm complex, its activating target (or non-target) RNA, and each of the ancillary CARF nucleases from the *T. thermophilus* Type III-A CRISPR system. We then induced expression of these immune components and compare the growth kinetics between groups expressing target or non-target RNA.

Results

Activation of TtCsm *in vitro* produces a spectrum of cOA species

The Csm complex of the bacterium *Streptococcus thermophilus* produces a range of differently sized cyclic adenylate molecules (cAx; x = 2-6)[16]. We therefore hypothesized that the Csm complex of *Thermus thermophilus* also produced a spectrum of cyclic oligoadenylates. To test this, we purified the TtCsm complex and activated it *in vitro* with target RNA. Resulting cAx molecules were purified by phenol-chloroform extraction and quantified by mass spectroscopy. This analysis demonstrated that the majority of cOA produced by TtCsm is cA_4, followed by cA_3. The other species, cA_2, cA_5, and cA_6 are produced, but in lower quantities (Figure 4.1C).

The bacterium *T. thermophilus* is a thermophile that lives at 50-70°C, but the host expression system used here grows at 37°C. To ensure that these immune components remain active at 37°C, we incubated the purified TtCsm complex with target RNA at 37°C with ^{32}P radiolabeled ATP and assayed for formation of cyclic nucleotides with thin-layer chromatography, as previously described[16] (Supplemental Figure 4.1). This analysis demonstrated that the TtCsm complex is still active at this temperature, but at a lower level than at 65°C.

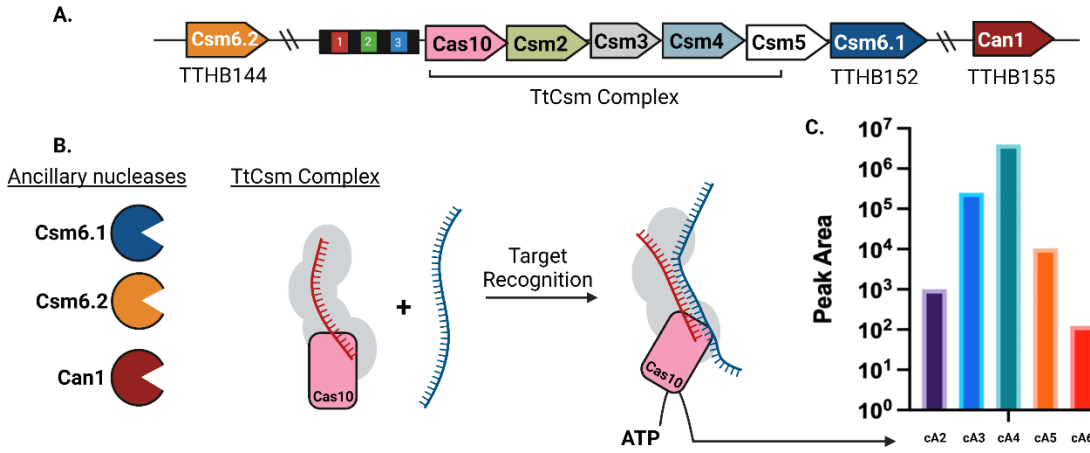


Figure 4.1 Outline of *Thermus thermophilus* Type III-A CRISPR locus and production of cOA molecules by TtCsm. (A) The Type III-A CRISPR locus of *Thermus thermophilus* includes a CRISPR locus, genes coding for the TtCsm complex, and three CARF ancillary nucleases. (B) Effector proteins of the *Thermus thermophilus* Type III-A CRISPR system. TtCsm complex, guided by a crRNA, identifies target RNA, and produces cOA molecules which then activate the ancillary Cas nucleases. (C) Mass spectroscopy data identifies major products produced by the TtCsm complex upon target recognition. Samples were prepared for mass spectroscopy by Laina Hall, who also prepared Figure 3.1C. Mass spectroscopy was performed by Joshua Jones, a graduate student in the lab of Kristin Koutmou at the University of Michigan.

CARF-containing nucleases are cOA activated nucleases *in vitro*

Given that TtCsm produces various species of cyclic oligo adenylate, and that this CRISPR system has various cOA-activated nucleases, we wondered which species activated the RNAse activity of each of the CARF nucleases present in the *Thermus thermophilus* Type III-A CRISPR locus. To test this, we purified the three CARF nucleases (TtCsm6.1, TtCsm6.2, TtCan1) and assayed their RNAse activity in the presence of the different cOA effectors. Gel filtration profiles of these proteins are shown in Figure 4.2 A-C. We determined that TtCsm6.1 RNase activity is modestly activated by cA_4, TtCsm6.2 RNase activity is modestly activated by cA_4 and strongly activated by cA_6, and TtCan1 RNase activity is strongly activated by cA_4. Additionally, TtCan1 has double-stranded DNase activity that is activated by cA_3. Again, because this heterologous expression system grows at 37°C, we also tested this cOA activated nuclease activity at 37°C (Supplemental Figure 4.2). This analysis determined that these nucleases were still active at 37°C, but at a lower level than at 65°C.

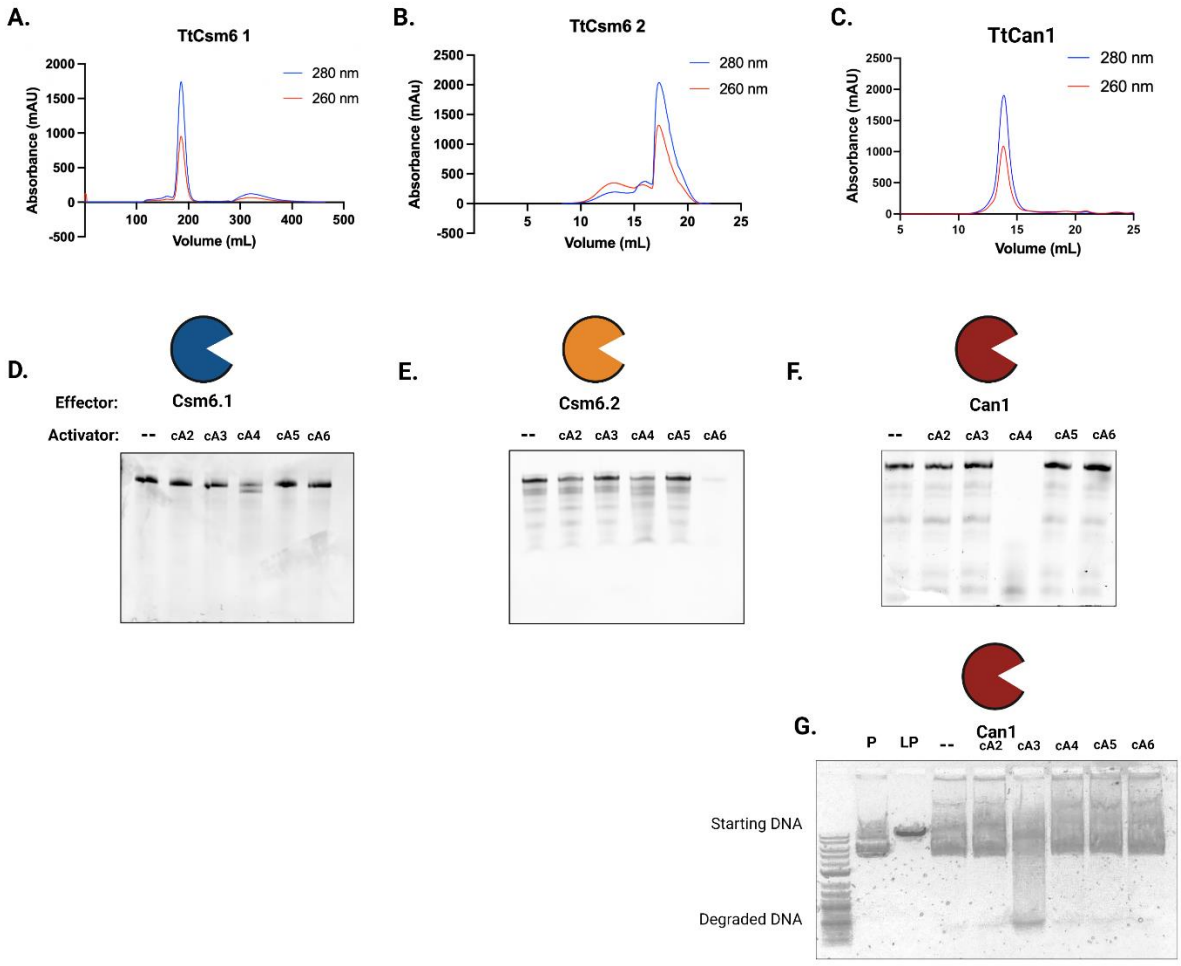


Figure 4.2 CARF nucleases are activated by cOA molecules *in vitro*. (A) Size-exclusion chromatogram for purification of TtCsm6.1. (B) Size-exclusion chromatogram for purification of TtCsm6.2. (C) Size-exclusion chromatogram for purification of TtCan1. (D) RNase assay of Csm6.1 visualized by urea-PAGE. (E) RNase assay of Csm6.2 visualized by urea-PAGE. (F) RNase assay of Can visualized by urea-PAGE. (G) dsDNase assay of Can1 visualized by agarose gel electrophoresis. P = plasmid, LP = linear plasmid.

Activation of immune components in cells results in a subtle change in growth kinetics

To test how activating these immune components affected the growth kinetics of bacteria, we developed a heterologous expression system of *E. coli* in which we could induce expression of TtCsm, corresponding target (or non-target) RNA, and each of the CARF effectors. We first set out to use the WT TtCsm complex in this expression system. This Csm complex maintains its RNA-guided RNase activity, HD nonspecific ssDNase activity, and its Palm cyclase activity. To test the effect of each of the CARF nucleases on growth kinetics, we compared cells that were expressing either target or non-target RNA (which is the same plasmid, only with the protospacer removed), along with the TtCsm complex and a CARF nuclease. This comparison showed very subtle differences between the target and non-target groups (Supplemental Figure 4.3C-E). We expected that this may be because the Csm complex was highly efficient at depleting the target RNA. Other groups had previously reported that Csm complexes halt production of cOA soon after the release of cleaved target RNA[9].

To get around this problem, we then switched to using the backbone mutant of TtCsm ($TtCsm^{Csm3\ D34A}$) in our growth curve analyses, which lacks crRNA-guided RNA cleavage activity. This complex has been previously demonstrated to bind to target RNA, produce cOA and not cleave target RNA[17]. Further, we thought that we may be able to observe a more substantial change in growth kinetics by inducing protein expression in the early log phase, instead of in the lag phase, as done in earlier experiments. We again compared groups expressing target RNA with a control group expressing non-target RNA and found a more substantial change in growth kinetics (**Figure 3.3C-E),** although the difference was still not as dramatic as previous groups had demonstrated[18,19]. We thought it likely that activation of such nonspecific

CARF-nucleases would result in shredding of essential RNA transcripts and structures, like rRNA or tRNA. As such, we expected this activation would result in dramatic changes to the growth kinetics between the groups expressing target and non-target RNA. We saw, though, that activating the production of cOA with target RNA did not dramatically affect the growth kinetics in this system. This may be because the TtCsm complex or the CARF nucleases are not as active in this expression system as they are in their native host.

Alternatively, this experimental design may be selecting for mutant bacteria that have evolved methods for avoiding activation of these Type III-A immune components. We thought that the most likely strategy for avoiding activation of the Csm complex was for bacteria to mutate the protospacer region of the pTarget plasmid. To test this, we grew the *E. coli* expressing the Type III-A components, activated expression of the components, then subsequently purified plasmids via miniprep, PCR amplified the protospacer region of pTarget, and sequenced the resulting PCR product by Sanger sequencing. This sequencing demonstrated that there were no dramatic mutations in the protospacer. Resulting PCR products and sequencing chromatograms are shown in Supplemental Figure 4.4.

We further wondered if the TtCsm complex was not being expressed or assembled in this system. To test this, we performed a thermophilic protein preparation, in which non-thermophilic proteins were crashed out of solution by heat denaturation, then the TtCsm complex was pulled out of solution with magnetic nickel-NTA beads that bind to the fused 6XHis tag. We visualized this protein purification using SDS-PAGE (Supplemental Figure 4.5**)**, which demonstrated that the expected Csm proteins were copurified together is the appropriate stoichiometry. This suggests that the TtCsm complex is being expressed and assembled.

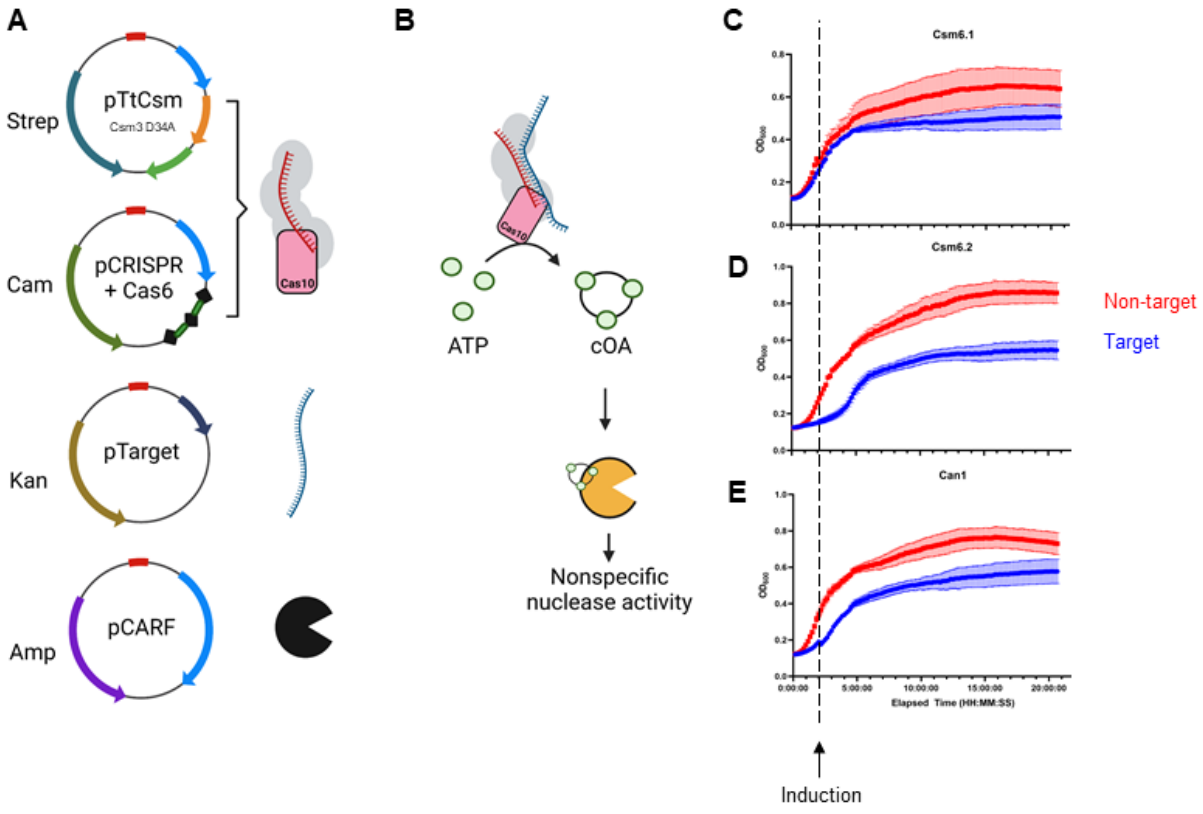


Figure 4.3 Activation of Type III-A immune components in *E. coli* result in subtle changes of growth kinetics. (A) Schematic outline of heterologous expression system. BL-21 (DE3) arabinose-inducible *E. coli* were transformed with four expression vectors. These vectors coded for the TtCsm$^{Csm3\ D34A}$ complex, crRNA, Cas6, which processes nascent CRISPR transcripts into mature crRNA, target RNA that is complementary to the crRNA, and CARF-containing nucleases from the *T. thermophilus* Type III-A locus. (B) Schematic of proposed outcome of this heterologous expression system. (C-E) Growth curves of heterologous expression system with TtCsm6.1 (C), TtCsm6.2 (D), or TtCan1(E). Non-target control coded for the same RNA as pTarget, but with the protospacer removed. Expression of Type III-A immune components and target RNA was induced when OD_{600} = 0.2-0.3 by addition of IPTG and L-arabinose. Error bars are ± SDM of three biologic replicates.

Discussion

Prokaryotes have evolved various routes to protect themselves from selfish genetic elements. Here, we explored one such system, the Type III-A CRISPR system from *Thermus thermophilus*. We demonstrate that the CARF nucleases of this CRISPR system hold non-redundant activation patterns *in vitro* (Figure 4.2A-C) and that TtCan1 can target different nucleic acids depending on the cyclic nucleotide activator present (Figure 4.2C). It is still unclear what the evolutionary strategy of these differential nuclease activities are. It is possible that early in the infection, TtCsm produces cA_4, which activates RNase activities of CARF nucleases, and later in the infection, the complex preferentially produces cA_3, to activate DNase activity, sheering the host genome, and killing the cell.

Interestingly, we did not observe the expected growth phenotype when these CRISPR systems were expressed and activated in live cells. We expected to see a dramatic crash in cell viability due to the activity of nonspecific RNase and DNase activity of these CARF-nucleases, as other groups have reported previously[18,19]. What we observed, though, was a subtle difference in growth kinetics, and a difference in final optical density. It is still unclear why this immune activation did not lead to a more dramatic difference, and why the major phenotypic difference between the target and non-target groups is the final OD_{600} reading at the stationary phase.

Had we seen a more robust growth defect in the cells with activated CRISPR systems, we would have moved ahead with determining which RNA species are preferentially targeted by these CARF nucleases using total RNA sequencing. Given the subtle difference between cells expressing active and non-active CRISPR systems, we thought it unlikely that we would be able to identify significant differences in RNA levels between the two groups.

Materials and Methods

Purification of TtCsm complex and CARF nucleases

These proteins were purified as previously described for the Csm complex[20] and for the CARF nucleases[13–15]. Briefly, expression vectors for these proteins were transformed into arabinose-inducible BL-21 (DE3) *E. coli* and grown at 37°C in LB broth (Lennox) (Thermo Fisher Scientific) to an OD_{600} of around 0.5 at which point they were induced with 1mM IPTG and 0.1% arabinose. After induction, they were transferred to an incubator at 25°C and grown overnight. Cells were harvested by centrifugation and resuspended in lysis buffer consisting of 25mM HEPES, 150 mM KCl, 10 mM imidazole, 1 mM tris(2-carboxyethyl) phosphine (TCEP), 0.01% Triton, 1 mM PMSF, 5% glycerol (v/v) for the TtCsm complex or 20mM Tris HCl (pH = 8.0), 500 mM NaCl, 1 mM DTT for the CARF nucleases. Both lysis buffers were supplemented to 1X with Halt Protease Inhibitor (Thermo Scientific product number 1861279). Lysis was then completed by sonication at 30% amplitude, 6 minutes total, 1 second on, 3 seconds off. Lysate was then clarified by centrifugation at 4°C for 25 minutes at 10,000 x g.

His-tagged TtCsm complex was then bound to HisTrap HP resin (Cytiva). Resin was washed with 30 column volumes of wash buffer containing 50mM HEPES (pH = 7.5), 150 mM KCl, 1 mM TCEP, 5% glycerol, and 20 mM imidazole. Protein was then eluted with lysis buffer supplemented with 300 mM imidazole. This eluate was concentrated by centrifugation in 100k MWCO Corning Spin-X concentrators following manufacturer's protocol. Protein was then further purified by size exclusion chromatography over HiLoad Superdex 200 26/600 GL size-exclusion columns (Cytiva) in a buffer consisting of 25 mM HEPES (pH = 7.5), 150 mM NaCl, 5% glycerol, 1 mM TCEP. Fractions containing the TtCsm complex were pooled, concentrated, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

After lysis and lysate clarification, TwinStrep-tagged CARF nucleases were purified first by binding to StrepTrap HP resin (Cytiva). Resin was washed with 30 column volumes of lysis buffer. Proteins were eluted in lysis buffer supplemented with 2.5 mM desthiobiotin and concentrated with centrifugation in 10K MWCO Corning Spin-X concentrators at 4°C. TwinStrep tags were cleaved from the proteins using small ubiquitin-related modifier (SUMO) protease (Invitrogen catalogue number: 12588018; 100 μL of 2.5 mg/ml protease per 20 mg of CARF nuclease substrate) during dialysis against SUMO digest buffer (30 mM Tris-HCl pH 8, 500 mM NaCl 1 mM DTT, 0.15% Igepal) at 4°C overnight. Cleaved His6-TwinStrep tag and uncleaved His6-TwinStrep-TtCsm6 were removed by binding to HisTrap HP resin (Cytiva), and the flow-through was concentrated using Corning Spin-X concentrators at 4°C. The CARF nucleases were finally purified using a HiLoad Superdex 200 26/600 size-exclusion column (Cytiva) in 20 mM Tris-HCl pH 7.5, 1 mM DTT, 400 mM potassium glutamate, 5 % glycerol. Fractions containing the CARF nucleases were pooled, concentrated, aliquoted, flash frozen in liquid nitrogen, and stored at −80°C.

In vitro production and purification of cOA

Cyclic oligoadenylate formation assays were prepared in a reaction buffer consisting of 20 mM Tris HCl (pH = 7.9), 200 mM potassium glutamate, 10 mM ammonium sulphate, 8 mM magnesium sulphate and 1mM TCEP supplemented with 500 uM ATP. Reaction was run with1 x 10^{10} copies of target RNA /uL, and 500 nM TtCsm complex. Reaction was incubated for 60 minutes at 60°C. Cyclic oligoadenylates were extracted using a phenol chloroform extraction and sent to collaborator Joshua Jones at the university of Michigan (Koutmou lab) for liquid chromatography-mass spectroscopy.

Thin-layer chromatography of cOA

Cyclic oligoadenylate assays were prepared as described above, but with α32P-ATP. Again, nucleotides were purified by phenol-chloroform extraction. Thin-layer chromatography was completed as previously described[16].

Cleavage assays of CARF nucleases

Purified CARF nucleases were used in cleavage assays as previously described with few modifications[13–15]. Briefly, the reaction was performed in a buffer containing 20 mM Tris-HCl (pH = 7.8), 200 mM potassium glutamate, 10 mM ammonium sulfate, 5 mM magnesium sulfate (for TtCsm6.1 and TtCsm6.2 assays) or 5 mM manganese chloride (for TtCan1 assays), and 2 mM TCEP. CARF nuclease was added to the reaction to 5 nM. Substrate RNA was added to 1.7 µM. Substrate plasmid DNA for TtCan1 cleavage was added to 80ng/µL. Cyclic oligoadenylates were added to each reaction to 400 nM. Reaction was incubated for 1 hour at the specified temperature. RNA was visualized by 8M urea-PAGE stained with SYBR Gold nucleic acid gel stain (ThermoFisher Invitrogen catalogue number S11494). Gels were imaged on an Amersham Typhoon imager (GE) with excitation at 488 nm and detection through a Cy2 filter.

Growth curves of heterologous expression system

Heterologous expression system was initiated by transforming four plasmid constructs that coded for the TtCsm complex, crRNA, TtCas6, target (or non-target) RNA, and each of the effectors into BL21(DE3) arabinose inducible *Escherichia coli* cells. First, the plasmid coding for the TtCsm complex and the plasmid coding for the crRNA and TtCas6 were transformed into chemically competent cells. Positive transformants were selected for on agarose plates containing chloramphenicol at 30 µg/mL and streptomycin at 30 µg/mL. Colonies from selection

plates were then picked, grown in selective LB broth (Lenox) (Thermo Fisher Scientific) (supplemented with chloramphenicol and streptomycin at 30 µg/mL each), then made electrocompetent using established protocols[21]. Two plasmids, one coding for one of the CARF nucleases, and one for the target or non-target RNA, were then electroporated into these cell lines, making quadruple transformants. Quadruple transformant colonies were selected for on quadruple antibiotic LB-agar plates (ampicillin at 100 µg/mL, chloramphenicol at 30 µg/mL, kanamycin at 30 µg/mL, streptomycin at 30 µg/mL). Positive transformants from each group were then grown overnight in 5 mL LB broth with all four antibiotics at the same concentrations that were in the agar plates.

After overnight growth, liquid cultures were then diluted in quadruple antibiotic LB broth to OD_{600} = 0.1. For experiments with induction from the start of the growth curve (**Supplemental figure 3.3**), LB media was supplemented to 1 mM IPTG and 0.2% arabinose at this dilution step. 200 µL of diluted culture was then aliquoted in triplicate in a 96 well plate. This plate was placed in a Cytation 5 cell imaging plate reader (Agilent) to generate a growth curve. Cells were grown at 37°C with constant orbital shaking at 250 rpm with OD_{600} measurements taken every 10 minutes for 20 hours. For growth curve experiments that cells were induced at early log phase (**Figure 3.3**), IPTG and arabinose were supplemented to the media of all groups when the first non-target control group reached an OD_{600} of 0.3.

Crude thermophilic protein purification

Thermophilic proteins were purified from the heterologous expression system to assay for TtCsm complex expression and assembly. This was done by growing the quadruple transformants in 1 mL selective LB media (streptomycin, kanamycin, ampicillin, and

chloramphenicol at levels reported above) overnight. The culture was then diluted to 25 mL selective media and grown at 37°C. When OD_{600} reached 0.5, protein expression was induced by the addition of IPTG to 1 mM and 0.2% arabinose. The culture was then returned to the incubator for overnight growth at 37°C.

Cells were harvested by centrifugation (3,000xg for 25 minutes at 4°C) then resuspended in 500 µL lysis buffer consisting of 25 mM HEPES (pH = 7.5), 200 mM potassium chloride, 1 mM DTT, and 5% glycerol. Cells were lysed by sonication at 30% output power for 20 seconds, lysate was allowed to settle, then was sonicated again at 30% output power for 20 seconds. Lysate was then clarified by centrifugation at 16,000xg for 20 minutes at 4°C. Supernatant was then transferred to a fresh tube and the pellet was discarded. Non-thermophilic proteins were heat denatured by heating lysate to 55°C for 45 minutes. Lysate was then cooled on ice, then clarified by centrifugation at 16,000xg for 20 minutes at 4°C. Supernatant was then transferred to a fresh supernatant tube and assayed by SDS-PAGE.

Magnetic Ni-NTA bead pulldown of TtCsm complex

The TtCsm complex was purified from the crude thermophilic protein purification using HisPur Ni-NTA magnetic beads (ThermoFisher Scientific catalogue number 88832) following manufacturer procedure. Briefly, magnetic beads were prepared by aliquoting 40 µL of homogenized Ni-NTA to a microcentrifuge tube. 160 µL of Equilibration buffer (25 mM HEPES at pH = 7.5, 1 mM TCEP, 150 mM NaCl, .05% Tween-20) was added to beads. Beads were collected to the side of the tube by magnet, and supernatant was discarded. 400 µL of Equilibration buffer was then added to beads, mixed by vortex, then the supernatant was discarded. Crude thermophilic protein purification was then added to beads and incubated for 30 minutes, vortexing every 10 minutes for 10 seconds. Beads were collected with magnet, and the

supernatant was removed and saved for later SDS-PAGE analysis. 400 µL Wash buffer (25 mM HEPES, pH = 7.5, 1 mM TCEP, 150 mM NaCl) was then added to beads and mixed by vortex. Beads were collected and supernatant was discarded. This was step was repeated. TtCsm complex was then eluted from the beads by adding 25 µL of Elution buffer (25 mM HEPES, pH = 7.5, 1 mM TCEP, 150 mM NaCl, 300 mM imidazole) and incubating 15 minutes, vortexing every 5 minutes for 15 seconds. Beads were collected by magnet and supernatant was removed and analyzed by SDS-PAGE.

References Cited

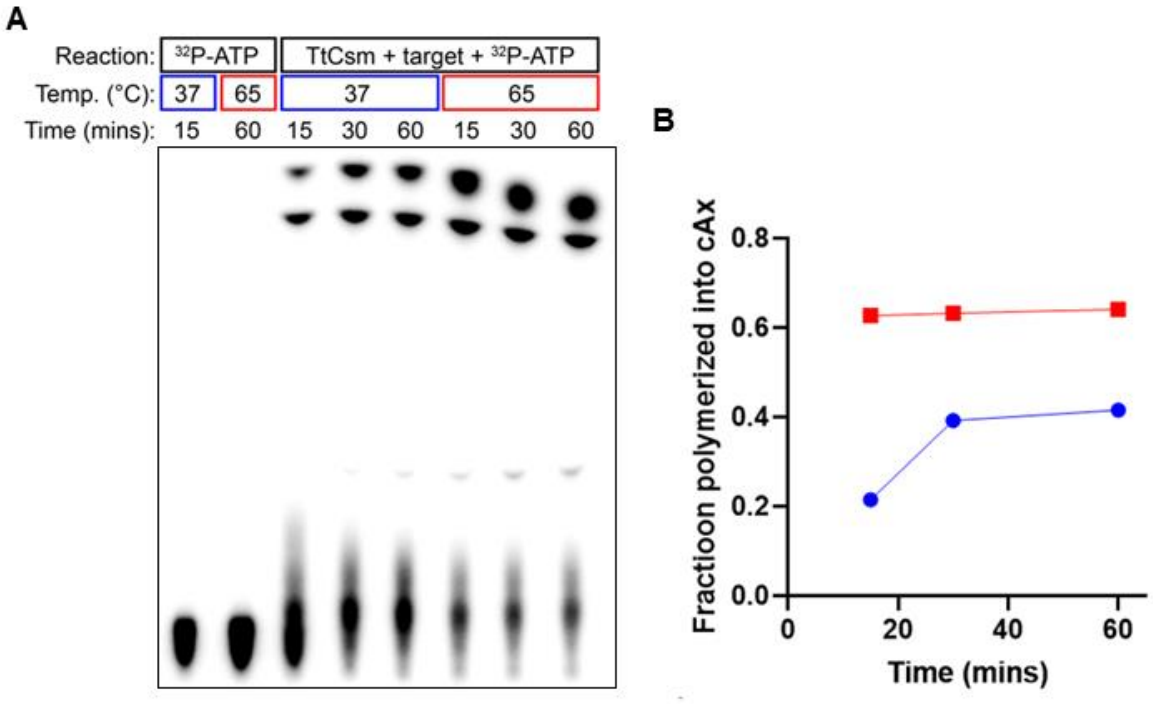
1. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).
2. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
3. Athukoralage, J. S. & White, M. F. Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. *RNA* **27**, 855–867 (2021).
4. Molina, R., Sofos, N. & Montoya, G. Structural basis of CRISPR-Cas Type III prokaryotic defence systems. *Curr. Opin. Struct. Biol.* **65**, 119–129 (2020).
5. Jia, N. *et al.* Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Mol. Cell* **73**, 264-277.e5 (2019).
6. Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J. & Staals, R. H. J. Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Biosci. Rep.* **38**, BSR20170788 (2018).
7. Tamulaitis, G., Venclovas, Č. & Siksnys, V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends Microbiol.* **25**, 49–61 (2017).
8. 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).
9. 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).
10. 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).

11. 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).
12. Koonin, E. V. & Makarova, K. S. Discovery of Oligonucleotide Signaling Mediated by CRISPR-Associated Polymerases Solves Two Puzzles but Leaves an Enigma. *ACS Chem. Biol.* **13**, 309–312 (2018).
13. McMahon, S. A. *et al.* Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat. Commun.* **11**, 500 (2020).
14. 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).
15. 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).
16. 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).
17. 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).
18. Ichikawa, H. T. *et al.* Programmable type III-A CRISPR-Cas DNA targeting modules. *PLOS ONE* **12**, e0176221 (2017).
19. Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. *Nat. Microbiol.* **4**, 656–662 (2019).

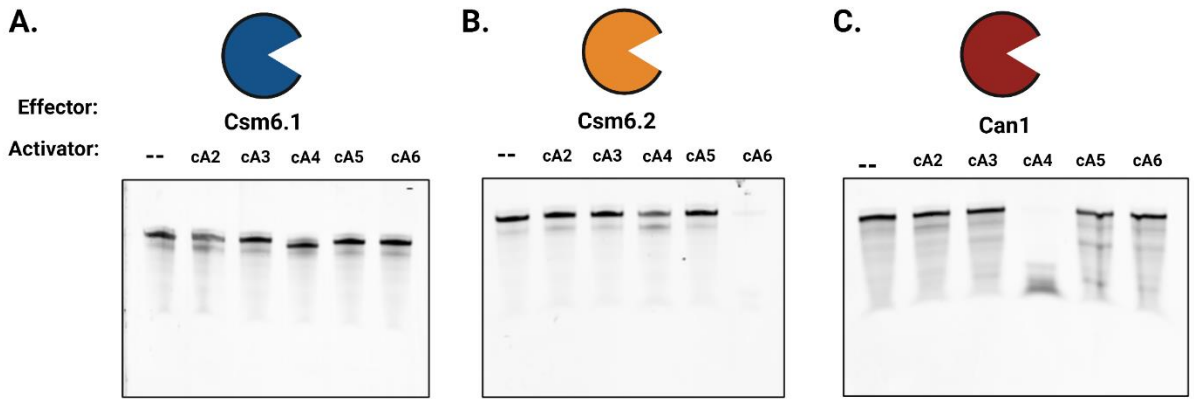
20. Liu, T. Y., Liu, J.-J., Aditham, A. J., Nogales, E. & Doudna, J. A. Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. *Nat. Commun.* **10**, 3001 (2019).

21. Miller, E. M. & Nickoloff, J. A. Escherichia coli Electrotransformation. in *Electroporation Protocols for Microorganisms* (ed. Nickoloff, J. A.) 105–113 (Humana Press, 1995). doi:10.1385/0-89603-310-4:105.

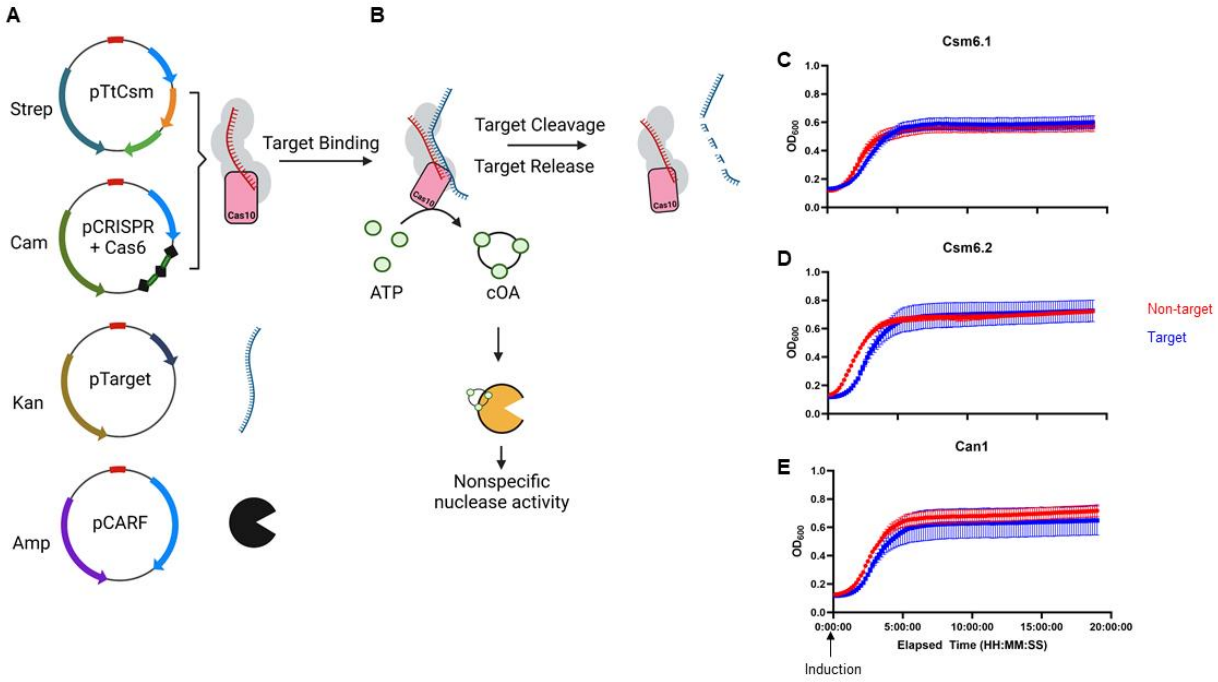
Supplemental Figures



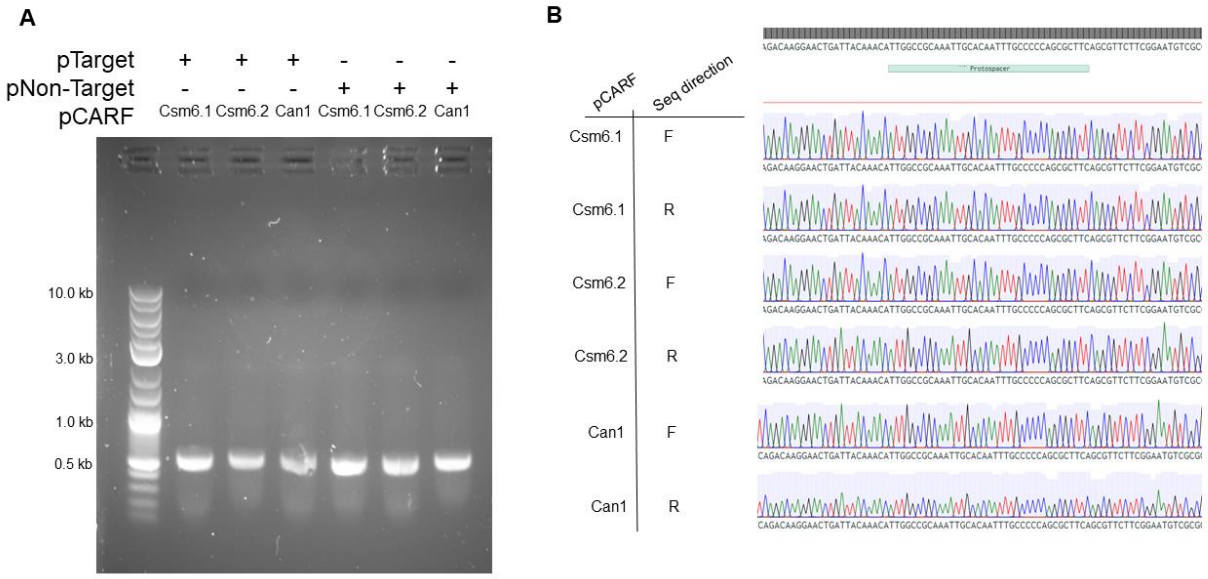
Supplemental Figure 4.1 TtCsm complex produces cAx at 37°C. (A) TLC plate visualization of cAx formation by TLC at both 37°C and 65°C. (B) Quantification of product formation was completed using Image J. Figure courtesy of Andrew Santiago-Frangos.



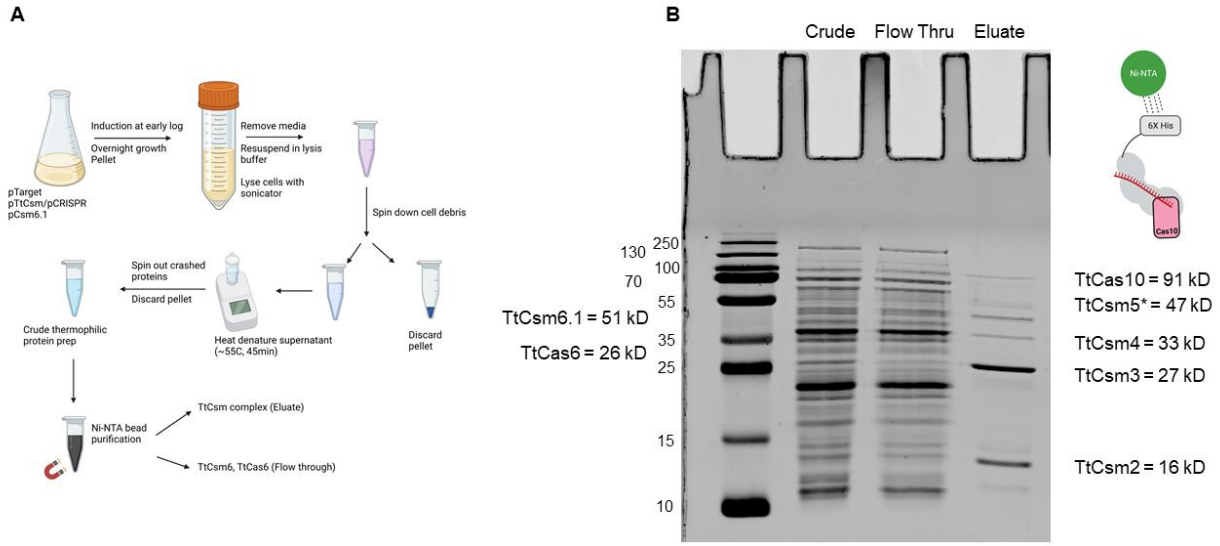
Supplemental Figure 4.2 CARF nucleases retain cOA-activated nuclease activity at 37°C. The nuclease activity of Csm6.1 (A), Csm6.2 (B), and Can1(C) were each assayed at 37°C. RNA was visualized by urea-PAGE.



Supplemental Figure 4.3 Activation of Type III-A immune components in *E. coli* result in subtle changes of growth kinetics. (A) Schematic outline of heterologous expression system. BL-21 (DE3) arabinose-inducible *E. coli* were transformed with four expression vectors. These vectors coded for the WT TtCsm complex, crRNA, Cas6, target RNA, and CARF-containing nucleases from the *T. thermophilus* Type III-A locus. (B) Schematic of proposed outcome of this heterologous expression system. Following target cleavage and target release, TtCsm stops producing cOA. (C-E) Growth curves of heterologous expression system with TtCsm6.1 (C), TtCsm6.2 (D), or TtCan1(E). Non-target control coded for the same RNA as pTarget, but with the protospacer removed. Expression of Type III-A immune components and target RNA was induced at the start of this growth curve with IPTG and arabinose.



Supplemental Figure 4.4 Protospacer in pTarget is not being mutated in this heterologous expression system. (A) PCR amplification of the protospacer region of pTarget or pNon-Target after overnight growth of heterologous expression system. Expected PCR product size was 590bp and 560bp for pTarget and pNon-Target, respectively. (B) Sanger sequencing of PCR products from (A). Aligning Sanger reads to reference shows no signs of mutations in the protospacer region of pTarget.



Supplemental Figure 4.5 TtCsm is being expressed and assembled in our heterologous expression system. (A) Schematic outline for thermophilic protein purification. (B) Proteins from heterologous expression system were visualized by SDS-PAGE. Lane 1 shows the crude protein mixture following the heat denaturation step. Lane 2 shows the flow through, proteins that were not bound to the Ni-NTA beads. Lane 3 shows the proteins bound to and eluted off the Ni-NTA beads. Schematic to right illustrates how the 6XHis tag that is fused to Csm5 binds Ni-NTA beads.

CONCLUSION

The constant bombardment of viruses that bacteria face has resulted in a dynamic interplay between viruses and the host defenses involved in battling these genetic parasites. One such immune defense is CRISPR, which utilizes crRNAs to identify complementary sequences within genomes or transcripts of bacteriophage[1,2]. For millennia, CRISPR immune systems have been useful tools for bacteria to protect themselves from viral predation, but in the last decade, humans have repurposed the RNA-guided nuclease activity of CRISPR-Cas systems for research, clinical, and biotechnological use[2–4]. One such application has been the interrogation of gene function by editing DNA genomes. This has provided insight into the genetic basis of molecular, cellular, and organismal biology of a wide array of organisms from across all domains of life. In contrast, there have been few tools developed to interrogate the role of specific RNA molecules in cellular functions. In this thesis, I have outlined work done during my master's training toward investigating RNA viruses (Chapter 2), RNA targeting CRISPR systems (Chapter 4), and applying those RNA targeting CRISPR tools to perform genome engineering in an RNA virus (Chapter 3).

Looking ahead, CRISPR technologies will likely continue to revolutionize biomedical science. The advancement of new RNA-targeting CRISPR tools will be central to gaining a better understanding of RNA biology and the mechanisms that drive RNA metabolism and repair, which may lead to new therapies for the current unmet needs of patients facing debilitating disease. The work outlined here is a starting point towards the goal of developing new tools for engineering RNA viruses and better understanding CRISPR systems that target RNA, which may be applied to further interrogating RNA processes in live cells.

Investigating SARS-CoV-2

Throughout the COVID-19 pandemic, genomic surveillance has been critical to our understanding of SARS-CoV-2 pathogenesis and evolution. Our lab performs this genomic surveillance for our county, giving public health officials insight into disease dynamics in our community. Through this surveillance, we identified a recurring truncation of an accessory protein, ORF7a. Using bioinformatics, we found that this truncation occurs in viruses isolated from around the globe and that these mutations do not persist over time. We hypothesized that this truncation alters protein structure and function because it initiates a stop codon that eliminates three of the protein's beta-sheets, the transmembrane domain, and the ER retrieval signal.

To determine the effect of this truncation on protein function, we developed mammalian expression vectors for the WT and mutant ORF7a genes with C-terminus peptide tags. Transfection of these expression vectors into mammalian cells, followed by immunocytochemistry and fluorescent confocal microscopy demonstrated that WT ORF7a has perinuclear localization, and the mutant has non-specific localization in the cytoplasm. This difference in cellular localization supported our hypothesis that the truncation altered ORF7a's function.

To further test our hypothesis, we focused on ORF7a's role in inhibiting Type I IFN signaling[5]. To determine if the mutant maintains this function, we generated HEK293T cells with an integrated luciferase reporter controlled by the IFN-Stimulated-Response-Element. In this system, IFN-response is correlated with luciferase activity. We transfected these cells with the ORF7a vectors, induced IFN-response with IFNα2b, and assayed for luciferase activity. We

determined that the WT ORF7a significantly reduced IFN-response over control cells, and the mutant did not.

ORF7a is not essential to viral replication, but we showed that the virus with truncated ORF7a has a fitness defect. We hypothesized that this lower fitness was due to a higher IFN-response in the cells infected by the mutant virus. To test this hypothesis, we quantified the expression of a panel of IFN-stimulated-genes (ISGs) in infected cell lysates using RT-qPCR. With this analysis, we demonstrated that cells infected by the mutant virus had significantly higher ISG expression than cells infected by the WT virus.

The importance of studying RNA viruses has never been more obvious, as the current COVID-19 pandemic continues to interrupt daily life around the globe. Our research group has contributed to the scientific fight against SARS-CoV-2 by developing methods of viral RNA detection in wastewater[6], developing novel CRISPR-based diagnostic technologies[7], by studying mutated variants of the virus circulating in our community[8], and by working toward the development of novel RNA genome engineering technologies.

Investigating RNA-Targeting CRISPR Systems

While most CRISPR systems exclusively target DNA, two groups, Type III and Type VI, target RNA. Type III systems are further distinguished, because they produce signaling molecules, cyclic oligoadenylates (cOA), upon recognition of target RNA. This signaling molecule then activates ancillary Cas effectors within the cell which transition the cell into an antiviral state[9,10]. In a model Type III-A CRISPR system, from the *Thermus thermophilus* bacterium, these ancillary Cas effectors are nucleases. We show here that the three cOA-activated nucleases have nonredundant activating molecules. We hypothesize that these different

nucleases represent nonredundant pathways for the cell to enter an antiviral growth arrest upon recognition of a phage transcript. To test this hypothesis, we developed a heterologous expression system, in which *Escherichia coli* expresses Type III-A immune components from *T thermophilus*. We induced expression of these immune components within these cells, along with target RNA to activate immune signaling, and measured their activity upon bacterial growth kinetics.

We have demonstrated that the surveillance complex, TtCsm, is being expressed and assembled in this heterologous host, but have yet to demonstrate a reliable phenotype tied to immune activation of this CRISPR system. It is possible that these immune complexes and their signaling pathway are not as active or efficient in this system as they are in their natural setting, and this heterologous system is not a faithful representation of Type III immunity.

Developing Tools for RNA Engineering

We applied the RNA targeting Csm complex from a Type III-A to develop a completely novel approach for RNA genome engineering. We show that the RNA targeting CRISPR enzyme complex is programable and that it is efficient in cleaving target RNA. Further, we repurpose splint ligation methods to repair the cleaved RNA *in vitro*, resulting in an edited RNA molecule. We applied this CRISPR-based technology to perform a targeted edit in the RNA genome of SARS-CoV-2. We show that this technology is capable of knocking out a gene in an RNA genome and expect that this approach to editing an RNA genome may also be applied to integrate transgenes, fluorescent peptide tags, or to delete large portions of an RNA genome. While we have not yet been successful in recovering edited viruses after transfection of edited genomes, we were successful in producing edited RNA without the use of any DNA

intermediate. Future work will be focused on applying this RNA editing technology to other RNA viruses for which transfection protocols are more well established, such as the *E. coli* RNA bacteriophage, MS2.

Outlook

In the relatively short period of time from the discovery of CRISPR as a prokaryotic immune system to now, CRISPR technologies have spearheaded a monumental revolution in nearly every field of biology. Perhaps most promising is the prospect CRISPR holds for curing genetic disease. In the last decade, CRISPR has redefined gene therapy. While conventional (pre-CRISPR) approaches to gene therapy depended upon delivering transgenes to cells and either maintaining those transgenes episomally, or integrating them into a cell's genome, CRISPR DNA editing approaches target an exact DNA locus in a genome and perform a programmable edit. This is important because expression of episomal DNA can be transient, and random integration of DNA can be oncogenic. CRISPR enables gene therapy, as it can directly knock-out a gene, direct homology-directed repair (HDR) to replace a disease allele with a healthy one, or direct base-editing to correct a point mutation[11,12].

Here we have focused on RNA-targeting CRISPR systems from the Type-III family. These CRISPR systems have yet to be applied therapeutically, but they hold great promise to treat disease, because they are the only system that can specifically cleave RNA without triggering collateral nuclease activity[13]. We show that these systems are useful for developing new strategies for RNA genome engineering, as we were able to use them to edit an RNA genome without the use of a DNA intermediate. Further, these systems are interesting, as they are the only family of CRISPR defense systems to use a signaling molecule to activate ancillary

effectors[10], a mechanism that nearly parallels eukaryotic immune strategies and blurs the lines between eukaryotic and prokaryotic immunity against viruses.

Looking ahead, it's likely that RNA-targeting CRISPR systems will drive a revolution in our understanding of RNA processes, such as RNA metabolism and repair. Further, I expect Type III CRISPR systems to have a major impact on gene therapy, as many diseases are rooted in the dysregulation of RNA and could be ameliorated by RNA knockdown or RNA editing[14,15]. As RNA editing technology advances to the point at which it can enable new forms of gene therapy, an outstanding challenge will be to make those therapies widely available to patients regardless of income level or geographic location. Gene therapies are extremely promising for improving the lives patients with low access to medical care because they are a one-time treatment that can prevent or reverse what would otherwise be a chronic illness, but currently available gene therapies are astronomically expensive and only available at academic medical institutions. As such, for genetic therapies to fulfill their great promise for revolutionizing medicine, advances in this field must be focused upon democratizing access to gene therapy and making the technology scalable to treat patients in low-and-middle-income countries.

REFERENCES CITED

1. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67–78 (2017).
3. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nat Rev Mol Cell Biol* **20**, 490–507 (2019).
4. Wilkinson, R. A., Martin, C., Nemudryi, A. A. & Wiedenheft, B. CRISPR RNA-guided autonomous delivery of Cas9. *Nat Struct Mol Biol* **26**, 14–24 (2019).
5. Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports* **33**, 108234 (2020).
6. Nemudryi, A. *et al.* Temporal Detection and Phylogenetic Assessment of SARS-CoV-2 in Municipal Wastewater. *Cell Reports Medicine* **1**, 100098 (2020).
7. 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).
8. Nemudryi, A. *et al.* SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Rep* **35**, 109197 (2021).
9. Athukoralage, J. S. & White, M. F. Cyclic oligoadenylate signaling and regulation by ring nucleases during type III CRISPR defense. *RNA* **27**, 855–867 (2021).
10. 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).

11. Yang, B., Yang, L. & Chen, J. Development and Application of Base Editors. *The CRISPR Journal* **2**, 91–104 (2019).
12. Uddin, F., Rudin, C. M. & Sen, T. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Front Oncol* **10**, 1387 (2020).
13. Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J. & Staals, R. H. J. Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Biosci Rep* **38**, BSR20170788 (2018).
14. Pollex, R. & Hegele, R. Hutchinson–Gilford progeria syndrome. *Clinical Genetics* **66**, 375–381 (2004).
15. Liu, E. Y., Cali, C. P. & Lee, E. B. RNA metabolism in neurodegenerative disease. *Disease Models & Mechanisms* **10**, 509–518 (2017).

REFERENCES CITED

Addetia, A. *et al.* Identification of multiple large deletions in ORF7a resulting in in-frame gene fusions in clinical SARS-CoV-2 isolates. *J Clin Virol* **129**, 104523 (2020).

Almazán, F. *et al.* Coronavirus reverse genetic systems: Infectious clones and replicons. *Virus Res* **189**, 262–270 (2014).

Amarilla, A. A. *et al.* A versatile reverse genetics platform for SARS-CoV-2 and other positive-strand RNA viruses. *Nat Commun* **12**, 3431 (2021).

Amitsur, M., Levitz, R. & Kaufmann, G. Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. *EMBO J* **6**, 2499–2503 (1987).

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat Med* **26**, 450–452 (2020).

Andolfo, I. *et al.* Common variants at 21q22.3 locus influence MX1 and TMPRSS2 gene expression and susceptibility to severe COVID-19. iScience 24, 102322 (2021).

Arunachalam, P. S. *et al.* Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).

Athukoralage, J. S., Graham, S., Grüschow, S., Rouillon, C. & White, M. F. A Type III CRISPR Ancillary Ribonuclease Degrades Its Cyclic Oligoadenylate Activator. *Journal of Molecular Biology* **431**, 2894–2899 (2019).

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).

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

Aubry, F., Nougairède, A., Gould, E. A. & de Lamballerie, X. Flavivirus reverse genetic systems, construction techniques and applications: A historical perspective. *Antiviral Research* **114**, 67–85 (2015).

Avanzato, V. A. *et al.* Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901-1912.e9 (2020).

Bain, J. D. & Switzer, C. Regioselective ligation of oligoribonucleotides using DNA splints. *Nucleic Acids Res* **20**, 4372 (1992).

Banerjee, A. *et al.* Experimental and natural evidence of SARS-CoV-2-infection-induced activation of type I interferon responses. *iScience* **24**, 102477 (2021).

Bedford, T. *et al.* Cryptic transmission of SARS-CoV-2 in Washington state. *Science* **370**, 571–575 (2020).

Benda, C. *et al.* Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. *Molecular Cell* **56**, 43–54 (2014).

Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**, 1036-1045.e9 (2020).

Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).

Choi, B. *et al.* Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *New England Journal of Medicine* **383**, 2291–2293 (2020).

Das, U. & Shuman, S. Mechanism of RNA 2′,3′-cyclic phosphate end healing by T4 polynucleotide kinase–phosphatase. *Nucleic Acids Research* **41**, 355–365 (2013).

Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **372**, eabg3055 (2021).

Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, (2014).

Emeny, J. M. & Morgan, M. J. Y. 1979. Regulation of the Interferon System: Evidence that Vero Cells have a Genetic Defect in Interferon Production. *Journal of General Virology* **43**, 247–252.

Fricke, T. *et al.* Targeted RNA Knockdown by a Type III CRISPR-Cas Complex in Zebrafish. *CRISPR J* **3**, 299–313 (2020).

Gamage, A. M. *et al.* Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional profiles. *PLOS Pathogens* **16**, e1009130 (2020).

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

Grubaugh, N. D. *et al.* An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biology* **20**, 8 (2019).

Hachim, A. *et al.* ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. *Nat Immunol* **21**, 1293–1301 (2020).

Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).

Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).

Harcourt, J. *et al.* Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg Infect Dis* **26**, 1266–1273 (2020).

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. A. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* **329**, 1355–1358 (2010).

Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).

Hochstrasser, M. L. & Doudna, J. A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences* **40**, 58–66 (2015).

Hodcroft, E. B. *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* **595**, 707–712 (2021).

Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e8 (2020).

Holland, L. A. *et al.* An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *Journal of Virology* (2020) doi:10.1128/JVI.00711-20.

Holshue, M. L. *et al.* First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine* **382**, 929–936 (2020).

Honda, S., Morichika, K. & Kirino, Y. Selective amplification and sequencing of cyclic phosphate–containing RNAs by the cP-RNA-seq method. *Nat Protoc* **11**, 476–489 (2016).

Hou, Y. J. *et al.* SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **182**, 429-446.e14 (2020).

Ichikawa, H. T. *et al.* Programmable type III-A CRISPR-Cas DNA targeting modules. *PLOS ONE* **12**, e0176221 (2017).

Jia, N. *et al.* Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Mol Cell* **73**, 264-277.e5 (2019).

Jiang, H. *et al.* SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol* **17**, 998–1000 (2020).

Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529–536 (2011).

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).

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. *Molecular Cell* **62**, 295–306 (2016).

Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921.e10 (2020).
Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885 (2020).

Konno, Y. *et al.* SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Reports* **32**, 108185 (2020).

Koonin, E. V. & Makarova, K. S. Discovery of Oligonucleotide Signaling Mediated by CRISPR-Associated Polymerases Solves Two Puzzles but Leaves an Enigma. *ACS Chem. Biol.* **13**, 309–312 (2018).

Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67–78 (2017).

Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).

Kotenko, S. V. & Durbin, J. E. Contribution of type III interferons to antiviral immunity: location, location, location. *Journal of Biological Chemistry* **292**, 7295–7303 (2017).

Lei, X. *et al.* Activation and evasion of type I interferon responses by SARS-CoV-2. *Nat Commun* **11**, 3810 (2020).

Lemieux, J. E. *et al.* Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* **371**, eabe3261 (2021).

Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
Lin, J., Shen, Y., Ni, J. & She, Q. A type III-A CRISPR–Cas system mediates co-transcriptional DNA cleavage at the transcriptional bubbles in close proximity to active effectors. *Nucleic Acids Research* **49**, 7628–7643 (2021).

Lin, P. *et al.* Type III CRISPR-based RNA editing for programmable control of SARS-CoV-2 and human coronaviruses. *Nucleic Acids Research* gkac016 (2022) doi:10.1093/nar/gkac016.

Liu, T. Y., Liu, J.-J., Aditham, A. J., Nogales, E. & Doudna, J. A. Target preference of Type III-A CRISPR-Cas complexes at the transcription bubble. *Nat Commun* **10**, 3001 (2019).

Loveday, E. K. *et al.* Effect of Inactivation Methods on SARS-CoV-2 Virion Protein and Structure. *Viruses* **13**, 562 (2021).

Lucas, C. *et al.* Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).

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).

Mantlo, E., Bukreyeva, N., Maruyama, J., Paessler, S. & Huang, C. Antiviral activities of type I interferons to SARS-CoV-2 infection. *Antiviral Research* **179**, 104811 (2020).

Martin-Sancho, L. *et al.* Functional landscape of SARS-CoV-2 cellular restriction. *Molecular Cell* **81**, 2656-2668.e8 (2021).

McMahon, S. A. *et al.* Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nat Commun* **11**, 500 (2020).

McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat Rev Immunol* **15**, 87–103 (2015).

Miller, E. M. & Nickoloff, J. A. Escherichia coli Electrotransformation. in *Electroporation Protocols for Microorganisms* (ed. Nickoloff, J. A.) 105–113 (Humana Press, 1995). doi:10.1385/0-89603-310-4:105.

Miorin, L. *et al.* SARS-CoV-2 Orf6 hijacks Nup98 to block STAT nuclear import and antagonize interferon signaling. *PNAS* **117**, 28344–28354 (2020).

Mogila, I. *et al.* Genetic Dissection of the Type III-A CRISPR-Cas System Csm Complex Reveals Roles of Individual Subunits. *Cell Rep* **26**, 2753-2765.e4 (2019).

Mogila, I. *et al.* Genetic Dissection of the Type III-A CRISPR-Cas System Csm Complex Reveals Roles of Individual Subunits. *Cell Reports* **26**, 2753-2765.e4 (2019).

Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147 (2016).

Molina, R., Sofos, N. & Montoya, G. Structural basis of CRISPR-Cas Type III prokaryotic defence systems. *Current Opinion in Structural Biology* **65**, 119–129 (2020).

Muth, D. *et al.* Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep* **8**, 15177 (2018).

Nam, K. H. *et al.* Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure* **20**, 1574–1584 (2012).

Nelson, C. A., Pekosz, A., Lee, C. A., Diamond, M. S. & Fremont, D. H. Structure and Intracellular Targeting of the SARS-Coronavirus Orf7a Accessory Protein. *Structure* **13**, 75–85 (2005).

Nemudryi, A. *et al.* SARS-CoV-2 genomic surveillance identifies naturally occurring truncation of ORF7a that limits immune suppression. *Cell Rep* **35**, 109197 (2021).

Osada, N. *et al.* The Genome Landscape of the African Green Monkey Kidney-Derived Vero Cell Line. *DNA Research* **21**, 673–683 (2014).

Pati, A., Padhi, S., Suvankar, S. & Panda, A. K. Minor Allele of Interferon-Induced Transmembrane Protein 3 Polymorphism (rs12252) Is Covered Against Severe Acute Respiratory Syndrome Coronavirus 2 Infection and Mortality: A Worldwide Epidemiological Investigation. *The Journal of Infectious Diseases* **223**, 175–178 (2021).

Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261–1276 (2017).

Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **5**, 1403–1407 (2020).

Riva, L. *et al.* Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **586**, 113–119 (2020).

Rosenthal, S. H. *et al.* Identification of eight SARS-CoV-2 ORF7a deletion variants in 2,726 clinical specimens. 2020.12.10.418855 (2020) doi:10.1101/2020.12.10.418855.

Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. *Nat Microbiol* **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).
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).

Schoggins, J. W. Interferon-Stimulated Genes: What Do They All Do? *Annual Review of Virology* **6**, 567–584 (2019).

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

Sola, I., Almazán, F., Zúñiga, S. & Enjuanes, L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* **2**, 265–288 (2015).

Spilman, M. *et al.* Structure of an RNA Silencing Complex of the CRISPR-Cas immune system. *Mol Cell* **52**, 10.1016/j.molcel.2013.09.008 (2013).

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).

Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12**, 2014–2019 (2006).
Stark, M. R. & Rader, S. D. Efficient Splinted Ligation of Synthetic RNA Using RNA Ligase. in *Spliceosomal Pre-mRNA Splicing: Methods and Protocols* (ed. Hertel, K. J.) 137–149 (Humana Press, 2014). doi:10.1007/978-1-62703-980-2_10.

Stoltz, M. & Klingström, J. Alpha/Beta Interferon (IFN-α/β)-Independent Induction of IFN-λ1 (Interleukin-29) in Response to Hantaan Virus Infection. *Journal of Virology* **84**, 9140–9148 (2010).

Su, Y. C. F. *et al.* Discovery and Genomic Characterization of a 382-Nucleotide Deletion in ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *mBio* (2020) doi:10.1128/mBio.01610-20.

Tamulaitis, G. *et al.* Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell* **56**, 506–517 (2014).

Tamulaitis, G., Venclovas, Č. & Siksnys, V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends in Microbiology* **25**, 49–61 (2017).

Tan, Y., Schneider, T., Leong, M., Aravind, L. & Zhang, D. Novel Immunoglobulin Domain Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related Viruses. *mBio* (2020) doi:10.1128/mBio.00760-20.

Tanaka, N. & Shuman, S. RtcB Is the RNA Ligase Component of an Escherichia coli RNA Repair Operon *. *Journal of Biological Chemistry* **286**, 7727–7731 (2011).

Tay, M. Z., Poh, C. M., Rénia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* **20**, 363–374 (2020).

THE CHINESE SARS MOLECULAR EPIDEMIOLOGY CONSORTIUM. Molecular Evolution of the SARS Coronavirus During the Course of the SARS Epidemic in China. *Science* **303**, 1666–1669 (2004).

Thi Nhu Thao, T. *et al.* Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).

Thi Nhu Thao, T. *et al.* Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561–565 (2020).

Tyson, J. R. *et al.* Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv* 2020.09.04.283077 (2020) doi:10.1101/2020.09.04.283077.

van Dorp, L. *et al.* No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun* **11**, 5986 (2020).

Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. 2020.12.30.20249034 (2021) doi:10.1101/2020.12.30.20249034.

Wang, C. *et al.* The Coronavirus PEDV Evades Type III Interferon Response Through the miR-30c-5p/SOCS1 Axis. *Frontiers in Microbiology* **11**, (2020).

Wang, F. *et al.* Advances in CRISPR-Cas systems for RNA targeting, tracking and editing. *Biotechnology Advances* **37**, 708–729 (2019).

Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).

Wilkinson, R. A., Martin, C., Nemudryi, A. A. & Wiedenheft, B. CRISPR RNA-guided autonomous delivery of Cas9. *Nat Struct Mol Biol* **26**, 14–24 (2019).

Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and North America. *Science* **370**, 564–570 (2020).

Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

Xia, H. *et al.* Evasion of Type I Interferon by SARS-CoV-2. *Cell Reports* **33**, 108234 (2020).

Xie, X. *et al.* Engineering SARS-CoV-2 using a reverse genetic system. *Nature Protocols* **16**, 1761–1784 (2021).

Xie, X. *et al.* An Infectious cDNA Clone of SARS-CoV-2. *Cell Host & Microbe* **27**, 841-848.e3 (2020).

Ye, C. *et al.* Rescue of SARS-CoV-2 from a Single Bacterial Artificial Chromosome. *mBio* **11**, (2020).

Young, B. E. *et al.* Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *The Lancet* **396**, 603–611 (2020).

Zang, R. *et al.* Cholesterol 25-hydroxylase suppresses SARS-CoV-2 replication by blocking membrane fusion. *PNAS* **117**, 32105–32113 (2020).

Zebec, Z., Manica, A., Zhang, J., White, M. F. & Schleper, C. CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. *Nucleic Acids Res* **42**, 5280–5288 (2014).

Zhang, B. *et al.* Clinical characteristics of 82 cases of death from COVID-19. *PLOS ONE* **15**, e0235458 (2020).

Zhang, Y. *et al.* The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. *PNAS* **118**, (2021).

Zhang, Y. *et al.* Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C Associated With Disease Severity in Coronavirus Disease 2019. *The Journal of Infectious Diseases* **222**, 34–37 (2020).

Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* **395**, 1054–1062 (2020).

Zhou, Z. *et al.* Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience* **24**, 102187 (2021).

Zhu, Y., Klompe, S. E., Vlot, M., van der Oost, J. & Staals, R. H. J. Shooting the messenger: RNA-targetting CRISPR-Cas systems. *Biosci Rep* **38**, BSR20170788 (2018).

Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological* https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).

Opposing activities of IFITM proteins in SARS-CoV-2 infection. *The EMBO Journal* **40**, e106501 (2021).