

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Manipulating CRISPR-Cas Systems: New Strategies for Diagnostics and Delivery

Permalink

<https://escholarship.org/uc/item/4dj0f6tq>

Author

Meriwether, Amanda Isabel

Publication Date

2024

Peer reviewed|Thesis/dissertation

Manipulating CRISPR-Cas Systems: New Strategies for Diagnostics and Delivery

By

Amanda Isabel Meriwether

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Daniel A. Fletcher, Chair
Professor Sarah A. Stanley
Professor Seth Shipman

Summer 2024

Manipulating CRISPR-Cas Systems: New Strategies for Diagnostics and Delivery

Copyright © 2024

by Amanda Isabel Meriwether

Abstract

Manipulating CRISPR-Cas Systems: New Strategies for Diagnostics and Delivery

By

Amanda Isabel Meriwether

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Daniel A. Fletcher, Chair

The natural world encompasses an awesome diversity of organisms, each which have evolved molecules and molecular assemblies enabling them to survive. From an engineering perspective, these molecular tools compose a huge reservoir of evolutionarily-honed building blocks that can be harnessed to create solutions to emerging problems. In this dissertation, I describe two projects that leverage modules derived from bacteria and viruses to address problems in human health.

One strategy that has quickly been adopted and manipulated by the scientific community is the CRISPR-Cas system. CRISPR-Cas systems evolved in prokaryotes as an adaptive immune response against ever-evolving viruses. These systems function by complexing effector protein(s) with RNA that targets the complex to invasive genetic material for degradation. The effector proteins from these systems, namely Cas9, 12, and 13 have been repurposed for many uses – from therapeutic gene editing to molecular diagnostics. Cas13a is an RNA-guided RNase that targets specific RNA sequences and, once activated by its target, non-specifically cleaves other RNA segments. CRISPR-Cas13a has been repurposed as a tool for molecular diagnostics of SARS-CoV-2 and other viral RNA in patient samples. We sought to extend this strategy to detect small non-coding RNAs which play a key role in regulating cells. Specifically, we wanted to detect miRNAs (~17-26nt long) and smaller RNAs, which have emerged as promising biomarkers for disease diagnosis. In this work, I demonstrate that Cas13a can detect RNA segments as small as eight nucleotides by splitting the guide RNA. Critically, we show that the Cas13a split guide system can detect miRNAs in cell lysate – without the need to purify or amplify the target. This project introduces a new Cas-based tool for small RNA diagnostics and biology studies.

Another reservoir of biological building blocks is the world of viruses. Viruses have evolved to be excellent at entering hosts, infiltrating their cells, and commandeering the cell to make more virions. Consider Influenza A virus (IAV): an endemic virus and that has caused at least six major pandemics. Here, I present a strategy to turn that pathogen into a partner, leveraging its unique genetic and biophysical properties to create a delivery vehicle to transport gene editing tools to hard-to-reach mucous-protected cells. I focus on delivering CRISPR-Cas9 as a ribonucleoprotein (RNP) instead of as a gene cassette, addressing issues of inefficient and off-target gene editing. To achieve this goal, I trick the native IAV genome packaging machinery to incorporate Cas9 guide RNA disguised as viral RNA into a non-replicating Influenza A virus-like particle. The Cas9 protein binds to the chimeric guide RNA, piggybacking into the budding particle to form a Cas9 Influenza-based Delivery Vehicle (CasIDV). We show preliminary evidence that CasIDVs deliver Cas9 RNPs to cultured cells, resulting

in gene editing events. With further optimization, the CasIDV system has the potential to become a promising strategy to deliver therapeutic cargo like CRISPR-Cas gene editing tools to currently inaccessible cells.

Dedicated to my family – biological and chosen

Table of Contents

List of Acronyms.....	iii
Acknowledgements	v
Chapter 1 – Introduction	1
My approach to science.....	1
Bacteria and viruses: what are they good for?	1
CRISPR-Cas systems	1
Chapter 2 – MicroRNA detection with CRISPR-Cas13a using a split guide RNA....	3
Abstract.....	3
Introduction	3
Results	5
Discussion	15
Methods and Materials	29
Chapter 3 – Designing a Cas13a split guide assay	32
Preliminary rules for designing anchor-capRNA-miRNA triplets	32
Anchor Region – Design Principles.....	32
Capture Region.....	36
Capture RNA.....	37
Design Principles Summary & Next Steps.....	43
Chapter 4 – Turning pathogens into partners: Converting Influenza A Virus into a delivery platform for CRISPR-Cas gene therapy	46
Abstract.....	46
Introduction	46
Results	48
Discussion	55
Methods and Materials	58
Chapter 5 – Future directions	61
Cas13a split guide system.....	61
CasIDVs	67
References	70

List of Acronyms

actRNA	activating RNA
capDNA	capture DNA
capRNA	capture RNA
Cas	CRISPR-Associated
Cas9-EDV	Cas9-packaging Enveloped Delivery Vehicle
CasIDV	Cas9 Influenza-based Delivery Vehicle
cDNA	complementary DNA
cityRNA	cleavage-inducing tinyRNA
CRISPR	clustered regularly interspaced short palindromic repeats
ddPCR	digital droplet polymerase chain reaction
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
dsRNA	double stranded RNA
GFP	green fluorescent protein
gRNA	guide RNA
HA	hemagglutinin
HEK	human embryonic kidney
IAV	Influenza A Virus
isomiRs	microRNA isoforms
LbuCas13a	<i>L. buccalis</i> Cas13a
LNA	locked nucleic acid
LOB	limit of blank
LOD	limit of detection
MDCK	Madin-Darby canine kidney
miRNA	microRNA
NA	neuraminidase
NCR	non-coding region
PCR	polymerase chain reaction
pgRNA	packagene guide RNA
PhD	Doctor of Philosophy
piRNA	PIWI-interacting RNA
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNP	ribonuclear protein
rRNA	ribosomal RNA
RT-PCR	reverse transcription polymerase chain reaction
SARS-CoV-2	severe-accute-respiratory-syndrome coronavirus 2
siRNA	small interfering RNA

snoRNA	small nucleolar RNA
SNP	single nucleotide polymorphism
snRNA	small nuclear RNA
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
T _m	melting temperature
tRNA	transfer RNA
tyRNA	tinyRNA
UTR	untranslated region
VLP	virus-like particle
vRNA	viral RNA
vRNP	viral ribonucleoprotein
vtRNA	vault RNA
yRNA	Y RNA

Acknowledgements

Earning a PhD in Bioengineering is by far the hardest thing I have yet done in my life, and I certainly couldn't have achieved that feat without the support of my community. I am ever grateful to the web of people that I call my community: my friends, family, teachers, advisors, mentors, peers, and co-workers in Austin, Berkeley, and beyond. Thank you all.

Thank you to my advisor, Dan Fletcher, for letting me into his lab and giving me the intellectual freedom and support to pursue crazy ideas. Thank you to Tiama Hamkin-Indik for being such an amazing role model, collaborator, and friend. To Ari Joffe and Emily Suter, thank you both for supporting me in my early days as a PhD student and caring about me and my work. To the Languishing/Livid/Lugubrious Ladies of Lipid Lane – Caitlin Cornell, Liya Oster, and Lena Blackmon – thank you for commiserating over failed experiments and celebrating over small successes. To Joana Cabrera, Amy Lyden, and the rest of the molecular diagnostics team: Carlos Ng Pitti, Andres Dextre, Abdul Bhuiya. Thank you all for welcoming me onto the diagnostics team and letting me shower you with my crazy ideas. Thank you to Roni Weissman and Abinaya Anbuchelvan for sharing your time to help push my projects forward. Thank you to the rest of the Fletcher Lab, past and present, for making this space so special.

Finally, thank you to María Díaz de León Derby. Words cannot describe how grateful I am that you messaged me the summer before starting graduate school, asking to be my roommate. I have been so lucky to go to work for the past six years with my best friend. I couldn't imagine journeying through the PhD without you by my side.

To my friends and roommates Lauren O'Neal, Federico Mora, and María Díaz de León Derby thank you for putting up with me, letting me decorate the apartment and paint whatever walls I want. Thank you for feeding me, crying with me, and celebrating with me. Ya'll are my family.

To the Kuntabunta, who made the pandemic bearable, thank you: Andoni Mourdoukoutas, Hunter Johnson, Jordan Baker, Nina Yanagisawa, and Kwasi Amofa. To the friends who keep me grounded outside of lab and make sure I take time to live my life, thank you.

To my community in Austin, thank you all for being the foundation that launched me into PhD. Thank you to The University of Texas at Austin Biomedical Engineering Department for setting the example of what a stellar engineering program feels like. Thank you especially to my mentors: Dr. Jeanne Stachowiak, Dr. Tom Yankeelov, Dr. Kenneth Diller, and Dr. Brandi DeMont. Thank you to my teachers at L.C. Anderson High School for seeing and supporting my intellectual curiosity in the early days: Cathryne McNamara, Susan Devaney, Jeanne Westmoreland, and Dr. William Wofford. Thank you to my friends that made me smile then and now: Mackenzie, Jenny, Tessa, and Sachit.

Finally, thank you to my family. To my parents Suzy and Michael Meriwether. To my brother, Nick, sister-in-law Chrissy, and goofy niece Isabel. To my grandparents, my aunts, and uncles, and cousins. I am so grateful to have grown up with such a wonderful family that has believed in me and supported me since day one. You can all stop asking me if I'm still taking classes.

Chapter 1 – Introduction

With this introduction, I hope to motivate why and how I approach scientific research and how that has led to my two main research projects.

My approach to science

I am simply amazed by nature. I am awed by the diversity of organisms that have evolved to thrive across so many distinct and complex environments. Consider the starfish, a vibrant creature that dwells in the ever-changing environment of the intertidal zone. Not only are they able to survive wet and dry conditions as the tide comes in and goes out, but also they are able to regenerate their own limbs and neurons¹! As an engineer, I see these starfish and other organisms broken down into their constitutive parts – functional modules that bestow specific skills and abilities upon the organism. If you take away the tube feet on a starfish, it can no longer adhere to rocky surfaces or grab prey to eat². But what if the starfish tube feet module was developed into a solution to reversibly adhere scaffolds to porous, wet surfaces like bones to facilitate bone repair? As a biomedical engineer, I like to think about how we can engineer biology to help solve problems in human health.

Building with biology, however, is tricky. The basic principles of biology are not as solid as those in physics. We do our best to engineer biology to answer questions and solve problems, but we are at the mercy of the biological systems we manipulate. The nature of this problem-solving approach lends itself to complex, risky projects, which is not an ideal situation for a PhD student. I believe, however, that academia is *the* place for creatively exploring new approaches to addressing big problems. I am grateful to my advisor, Dan Fletcher, for supporting me in pursuit of these biologically-driven engineering projects.

Bacteria and viruses: what are they good for?

One reservoir of biological building blocks that I draw from in my research is derived from bacteria and viruses. From a human perspective, bacteria and viruses are usually thought of as pathogens – agents that cause disease. From the perspective of the bacteria or virus, they are just trying to survive. On the path to survival, they have evolved characteristics that are marvelous to them, but deadly to humans.

Bacteria have given people invaluable tools – DNA polymerases that withstand high temperatures³, allowing scientists to easily synthesize DNA, and CRISPR-Cas enzymes⁴ that enable scientists to edit genes, to name a few. Viruses have also contributed to the scientific tool kit, from HIV-derived tools to express exogenous proteins in cells⁵ to adeno-associated viral vectors to deliver gene editing tools to cells⁶. In my thesis, I leverage molecular tools from both bacteria and viruses to address problems in molecular diagnostics and therapeutic delivery.

CRISPR-Cas systems

From bacteria, I repurpose CRISPR-Cas systems, which are elements of their adaptive immunity⁷. These systems are designed to detect and destroy foreign genetic material, such as that from invading viruses. There are many types of CRISPR-Cas systems that have evolved to protect bacteria and other prokaryotes, but they all are composed of two main components: nucleic acids that help identify foreign genetic material and effector proteins that cleave the identified genetic targets⁷. CRISPR-Cas systems are divided into two main groups: Class I, which involves a complex of multiple effector

proteins, and Class II, which uses one effector protein containing multiple functional domains⁷. In my work, I focus on two types of Class II CRISPR system, Cas9 and Cas13a.

CRISPR-Cas9 is composed of the Cas9 effector protein which complexes with two single-stranded RNA (ssRNA) segments, though scientists have engineered a single ssRNA segment that serves the same function⁴. This single piece of RNA is called the single guide RNA (sgRNA) or simply the guide RNA (gRNA). The gRNA contains two main functional motifs, one that facilitates binding to the Cas9 protein, and one, called the seed region, that determines what DNA sequence will activate the nuclease activity of the Cas9 protein. When the Cas9 and gRNA complex together, they form a Cas9 ribonucleoprotein (Cas9 RNP). This complex is then activated by a double-stranded DNA (dsDNA) segment matching the seed region of the gRNA. Once activated, the Cas9 RNP cleaves the target dsDNA. This RNA-guided DNA nuclease has been repurposed for applications in gene editing.

The other CRISPR-Cas system I work with is CRISPR-Cas13a. Like CRISPR-Cas9, CRISPR-Cas13a involves a single effector protein (Cas13a) that complexes with a gRNA⁸. Unlike a Cas9 RNP, a Cas13a RNP is activated by ssRNA. Once activated, the RNP cleaves both the target ssRNA and then starts non-specifically cleaving other ssRNA⁸. This RNA-guided RNA nuclease has been repurposed for applications in molecular diagnostics.

In my thesis, I present two projects that draw upon molecular tools from bacteria and viruses. I take CRISPR-Cas systems from bacteria and modify them to 1) detect small RNAs for disease diagnostics and 2) package inside delivery vehicles for transportation to and genetic editing of cells. To develop an effective delivery vehicle, I enlist the help of Influenza A Virus to create a delivery vehicle capable of accessing hard-to-reach cells and depositing therapeutic cargo inside the cell.

Chapter 2 – MicroRNA detection with CRISPR-Cas13a using a split guide RNA

Amanda Isabel Meriwether, Joana Cabrera, Amy Lyden, Melanie Ott, and Daniel A. Fletcher

Manuscript in submission.

Abstract

Small RNAs like microRNAs (miRNAs) play vital regulatory roles in cells and are promising biomarkers of disease. While RNA is typically detected using RT-PCR or sequencing methods, CRISPR-Cas13a has emerged as an alternative tool that does not require RNA purification, reverse transcription, or amplification steps. However, small RNA detection with Cas13a is challenging due to the short segment lengths (miRNAs are ~17-26 nucleotides long), which are comparable to the enzyme's seed region. Here we demonstrate that splitting the guide RNA (gRNA) enables Cas13a to directly detect single-stranded RNA segments ranging from 8 to 24 nucleotides. The Cas13a split guide system is sensitive to clinically-relevant miRNA concentrations (femtomolar) and rejects synthetic RNAs with sequence mismatches and misalignments. Finally, we show that Cas13a with a split guide RNA can directly detect miRNAs in unpurified cell lysate as well as in purified cellular RNA. This work extends the use of Cas13a to measurement of small RNAs, enabling new Cas-based applications in small RNA biology and diagnostics.

Introduction

Small RNAs are non-coding RNAs that are usually <200nt and play a regulatory role in cells⁹. Canonical small RNAs that mediate RNA silencing include small interfering RNA (siRNA, ~20-25nt), PIWI-interacting RNAs (piRNAs, ~18-30nt), miRNAs (~17-26nt), and a subset of tiny RNAs (tyRNAs, <18nt) called cleavage-inducing tyRNAs (cityRNAs, 14-15nt)⁹⁻¹³. Other smaller species of RNA (as small as 15nt) are enzymatically derived from small structural RNAs including transfer RNA (tRNA), Y RNA (yRNA), vault RNA (vtRNA), 5S ribosomal RNA (5S rRNA), small nuclear RNAs (snRNAs), and small nucleolar RNA (snoRNA)^{9,10}. The function of these small RNA derivatives is not well characterized, though many seem to resemble miRNAs⁹. Together, small RNAs form a complex cellular regulatory network that is an active area of study. Accurately and quantitatively detecting these small RNAs in different cellular contexts is necessary for elucidating their function and informing small RNA therapeutics and diagnostics. Current detection strategies, however, are challenged by the small size of the RNA species and their low abundance in samples.

Detection of circulating miRNAs is of special interest, as they are important molecular diagnostic targets that can serve as biomarkers for disease diagnosis and progression¹⁴. MicroRNAs are short, single-stranded RNA segments (~17-26nt) that bind to cellular Argonaute proteins to form RNA-induced silencing complexes (RISCs) that target and suppress cellular messenger RNA (mRNA)¹⁵. This miRNA-mediated regulatory network is altered in disease^{14,15}, providing a snapshot of dysregulated cells and tissues. Promising research over the last decade has identified unique panels of miRNAs correlated with early stages of disease^{16,17}, disease progression^{14,18}, and response to treatment¹⁴. Critically, these miRNA biomarkers can be collected from body fluids like blood and saliva without the need for invasive biopsies¹⁹. Though miRNA panels are a promising strategy for non-invasive disease diagnosis and monitoring, the complexity of current miRNA detection methods has limited their clinical use.

Existing strategies for small RNA detection, including miRNAs, typically rely on polymerase chain reaction (PCR)-based methods, including reverse transcription PCR (RT-PCR) and digital droplet PCR (ddPCR). These methods first require purification of RNA from a sample and then reverse transcription into cDNA before amplification and detection. The modifications involved in reverse transcription of miRNA to cDNA, such as the use of stem-loop primers and poly-A tailing to increase the length of the product, often introduce error and reduce the reliability and repeatability of the assay²⁰. Next, amplification of miRNA cDNA involves careful design of primers and probes that accommodate both the small size of miRNA (~17-26nt) and the sequence similarity within miRNA families. While RT-PCR and ddPCR methods have reported sensitivities ranging from 0.25 to 8 cDNA copies/uL for RT-PCR and 1.25 to 2 cDNA copies/uL for ddPCR²¹, the true limit of detection of miRNA copies depends on the efficiency of the reverse transcription step as well as the specificity of the RNA purification. Next-generation RNA sequencing (RNA-seq) is an alternate strategy for identifying and quantifying small RNAs, with the added potential for small RNA discovery. However, it is similarly flummoxed by the need to extract, purify, and reverse transcribe miRNAs²². Substantial error and bias can be introduced during library preparation when RNA is reverse transcribed to cDNA and amplified, as with PCR. Furthermore, there are multiple approaches to small RNA sample preparation, making it difficult to compare relative quantification across different sequencing methods. Techniques based on nanopore sequencing have been adapted to directly sequence miRNAs and are an exciting direction, but they are limited by a 15nt read-length and the base-calling error rate of nanopore technology²³.

Cas13a offers an attractive alternative for direct detection of small single-stranded RNAs (ssRNAs). Cas13a is a class 2, type VI CRISPR-Cas system that complexes with a guide RNA (gRNA) to form a ribonucleoprotein (RNP) complex⁸. The gRNA contains a stem-loop region that mediates the interaction between the enzyme and gRNA and a programmable 20-28nt seed region. When a target RNA complements the seed region, the Cas13a RNP activates non-specific trans-cleavage of ssRNA⁸. To leverage the programmable Cas13a activation as a diagnostic tool, assays use a reporter molecule composed of a fluorophore linked to a quencher by a short ssRNA segment²⁴⁻²⁶. When the Cas13a RNP is activated by its target RNA, the RNP cleaves the ssRNAs of the reporters, unquenching the fluorophores and generating a signal that increases over time (Fig. 1a). The rate at which the signal increases is proportional to the number of active Cas13a RNPs and therefore enables quantification of the target RNA's abundance in a sample²⁶.

In addition to being quantitative, Cas13a-based assays have other characteristics that make them promising tools for RNA biology and diagnostic assays. They are rapid and sensitive, with run times typically under 30 minutes and limits of detection approaching that of PCR^{27,28}. Most important for detection of small RNAs is that Cas13a can be used without RNA purification, reverse transcription, and amplification steps that are required by PCR and sequencing-based methods^{24,28-31}. Recent studies have shown that Cas13a assays can detect 23nt miRNAs either directly³² or coupled to additional signal-amplification steps³³⁻³⁵. Chen et al developed a strategy to detect a 17nt miRNA using a guide RNA that folds back on itself³⁶. With this approach, they claim a limit of detection (LOD) of 15fM (9e3 copies/uL). Cas12a, which normally binds single-stranded (ss) and double-stranded (ds) DNA, has also been used to detect short RNA segments³⁷. Rananaware et al were able to detect 10-12nt sequences by splitting the activating nucleic acid into two parts, with one part ss- or dsDNA and the other part ssRNA. Using this approach, they detected a 23nt miRNA with an LOD of 767pM (4.6e8 copies/uL). However, clinically-relevant small RNAs can be present at lower concentrations (1e4 to 1e5 copies/uL for low-abundance miRNAs to 1e7 copies/uL for higher abundance miRNAs^{19,20}) and

smaller sizes (14-15nt cityRNAs^{12,13}). Thus, additional methods for detecting short and low-abundance small RNAs are needed.

Here we present an alternate strategy to detect small RNAs (8-24nt) that flips the traditional paradigm of Cas13a-based RNA detection by splitting the gRNA. This strategy uses a Cas13a enzyme, split gRNA, and sequence-specific target RNA to detect ssRNA with lengths as small as 8nt. We split the gRNA within the seed region, replacing the 3' end with a short RNA of interest, which we call the activating RNA (actRNA). The split gRNA and actRNA are fully complemented by the added target RNA, which we call the capture RNA (capRNA) and include as a reagent in the system. To target different actRNAs, we change the sequence of the 5' end of the capRNA; the split gRNA remains constant. To demonstrate the utility and flexibility of the Cas13a split guide system, we show that the system can directly detect multiple miRNA species in cell lysate.

Results

Redesigning the Cas13a RNP to detect small RNAs

We first used a conventional Cas13a RNP (Cas enzyme and gRNA) to characterize the lower limit of target RNA length that Cas13a can robustly detect. We used *L. buccalis* Cas13a (LbuCas13a) with a gRNA containing a 20nt seed region and tested three different target RNA lengths: (i) a 36nt target RNA containing a 20nt sequence complementary to the seed region, (ii) a 20nt target RNA complementary to the seed region, and (iii) a 10nt target RNA complementary to the first 10nt of the gRNA seed (Fig. 1b, top; see Supplementary Information Table 1 for all sequences). We quantified Cas13a activation for each target RNA at a range of concentrations by extracting the slope of the linear region of the reaction curve (fluorescence vs time) and plotting the slope (reaction rate) vs target concentration (Fig. 1b, bottom; see Extended Data Fig. 1a and Methods). We saw significantly reduced activation between the 36nt and 20nt targets at all concentrations tested and no activation from the 10nt target ($p < 0.001$ ANOVA of slope between all slopes).

For our bulk assay conditions, we estimated a limit of detection (LOD) of 2.7×10^4 copies/uL for the 36nt target RNA and an LOD of 1.6×10^5 copies/uL for the 20nt target RNA, roughly 6x higher. Critically, the 10nt target was not detected over the background signal at any of the measured concentrations. We tested a different 10nt target complementary to the second half of the gRNA seed region (10nt target B, position 11-20) and found that it was also undetectable above background, even at a concentration of 1×10^8 copies/uL (Extended Data Fig. 1b). This reduced sensitivity for 20nt ssRNA targets and inability to detect 10nt ssRNA targets limits the utility of conventional Cas13a RNPs in applications where short ssRNAs are of interest, such as detecting miRNAs and cityRNAs.

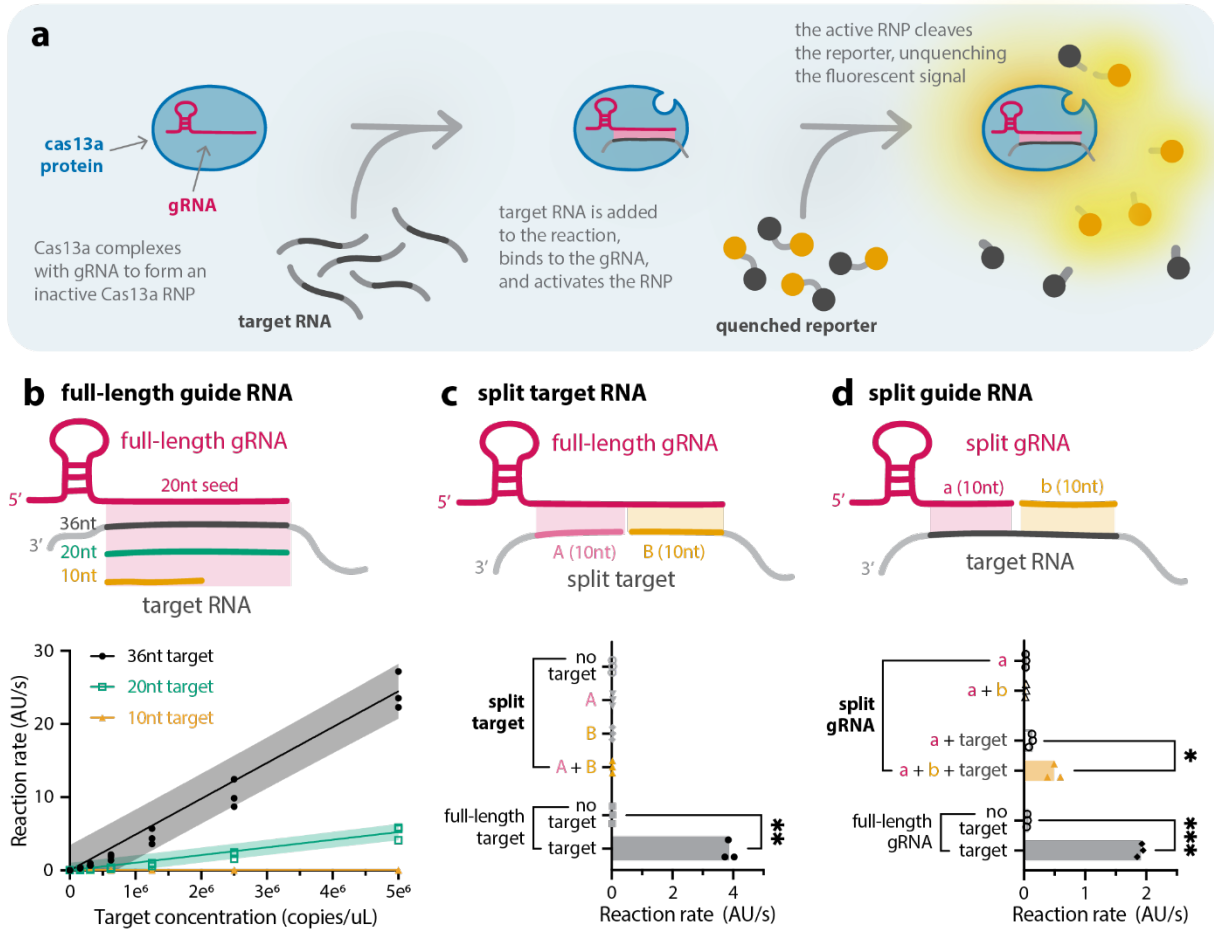


Figure 1. Cas13a is activated by ssRNA <20nt with a split gRNA.

A: Schematic illustrating ssRNA detection by Cas13a. Cas13a (light blue) complexes with a gRNA (hot pink) to form a Cas13a ribonucleoprotein complex (RNP). The RNP binds to target RNA (grey), activates, and cleaves quenched fluorophore reporters. B, Top: Schematic illustrating a Cas13a full-length gRNA with a 20nt seed region (hot pink) complemented with a 36nt (grey), 20nt (green), and 10nt (yellow) target RNA. Bottom: Scatter plot shows the reaction rate (AU/s) for Cas13a RNPs activated by 36nt, 20nt, and 10nt targets across a range of target concentrations ($n = 3$ technical replicates per target and concentration). The reaction rate is calculated as the slope of the linear region of the reaction progress curve (AU vs. time). The reaction rates for each target length are fit to a linear curve to compare limit of detection across different target lengths. The linear regression is plotted with 95% prediction bands. Cas13a and gRNA are equimolar at 25nM. C, Top: Schematic illustrating split target approach, where gRNA (hot pink) complexes to two pieces of target RNA (A, in light pink, and B, in yellow), that together complement the full 20nt seed region. The two pieces of target RNA are the limiting reagents of the system. Bottom: The reaction rates (AU/s) of different combinations of the split target system are compared to the full-length gRNA and 36nt target RNA. Split target parts A and B are both at $1e6$ copies/uL in the final reaction, as is the full-length target. Cas13a and gRNA are equimolar at 25nM. The mean reaction rate (AU/s) with three technical replicates is plotted. D, Top: Schematic illustrating 'split guide' approach, where gRNA is split in two parts in the seed region. The first part, a (hot pink), contains the 30nt stem-loop region and 10nt of the seed region. The second part, b (yellow), contains the remaining 10nt of the seed region and is the limiting reagent of the system. Together split gRNA parts a and b complement to the full-length target RNA (36nt, with 20nt sequence complementary to the seed region, shown in grey). Bottom: The reaction rates (AU/s) are plotted for the split guide system tested with all combinations of RNA components. Cas13a, split gRNA a, and target RNA (36nt) are all equimolar at 25nM. The 10nt split gRNA b is at $1e6$ copies/uL. For the full-length gRNA system, gRNA is equimolar to Cas13a at 25nM and the 36nt target is at $1e6$ copies/uL. The mean reaction rate (AU/s) with three technical replicates is plotted. Stats: For C and D, we used pairwise analysis of the reaction rates between experimental conditions and negative controls using a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for multiple comparisons. We report multiplicity-adjusted p-values. Non-significant interactions ($p > 0.05$) are not plotted. * is $p \leq 0.05$, ** is $p \leq 0.01$. See Extended Data Fig. 1 for more details.

To overcome these limitations, we tested two different approaches to detect shorter ssRNA targets: (i) splitting the target RNA into two parts, where either portion could act as the short ssRNA target

(Fig. 1c, top) and (ii) splitting the gRNA within the seed region, where the segment without the stem-loop acts as the short ssRNA target (Fig. 1d, top). To test the ‘split target’ approach, we partitioned the 36nt target at the midpoint of the 20nt seed complement region. Split target part A (light pink) is 15nt long (5nt 3’ flanking region and 10nt seed complement). Split target part B (yellow) is 21nt long (10nt seed complement and 11nt 5’ tail). We then added split target parts A, B, or both at a concentration of $1e7$ copies/uL to the Cas13a RNP mixture. None of the conditions generated a reaction rate above the no target control (Fig. 1c, bottom).

We next tested the ‘split guide’ approach (Fig. 1d) by similarly splitting the gRNA into two parts in the seed region. Split gRNA part a (hot pink) contains the 30nt stem-loop region and the first 10nt of the seed region. Split gRNA part b (yellow) contains the remaining 10nt of the seed region. For the ‘split guide’ approach to be a useful strategy to detect short ssRNA, it must work when split gRNA part b is the limiting reagent, as would be the case when detecting a low-abundance miRNA. To test the ‘split guide’ strategy, we formed the RNPs by incubating Cas13a with an equimolar amount of split gRNA part a and then added an equimolar amount of 36nt target RNA and a limiting amount of split gRNA part b. Only when all three RNA segments were present (split gRNAs parts a and b, and target RNA) was a reaction rate generated that was significantly above the no target controls (Fig. 1d, bottom). The background-subtracted reaction rate for the ‘split guide’ approach with $1e6$ copies/uL of 10nt split gRNA part b was roughly 20% that of the full-length guide system with a 36nt target RNA ($0.37 \pm \text{SEM } 0.06$ AU/s and $1.9 \pm \text{SEM } 0.03$ AU/s, respectively; Extended Data Fig. 1d). We also tested the ‘split guide’ approach where the 36nt target RNA was the limiting reagent and the two parts of the split gRNA were equimolar and found that that ratio of reagents also worked (Extended Data Fig. 1c).

Together, these experiments suggest that splitting the gRNA is a viable strategy to detect short ssRNA, where the split gRNA part b is the target.

Split guide system enables detection of 10nt RNA targets that are reaction limiting

To clarify the new roles of the RNA segments in the Cas13a split guide system (Fig. 2a), we call the short ssRNA of interest the activating RNA, or ‘actRNA’ (yellow), and the remaining portion of the standard gRNA the ‘split gRNA’ (hot pink). The split gRNA contains the 30nt stem-loop region that mediates the interaction with the Cas13a enzyme and a portion of the seed region that we term the ‘anchor region’ (light pink). The target RNA from the conventional guide system is now called the capture RNA, or ‘capRNA’ (grey). It fully complements the anchor region of the split gRNA to become part of the RNP. The capRNA also contains a sequence complementary to the actRNA (short ssRNA or miRNA of interest), which we call the ‘capture region’ (light yellow). The capRNA thus holds the split gRNA and actRNA together, completing the double stranded RNA seed region. With the Cas13a split guide system, we change only the capture region sequence to detect different actRNAs.

While detecting a 10nt ssRNA with the split guide system is an improvement over the conventional Cas13a approach, the sensitivity of the split guide system is limited in part by the background activity of the RNP and capRNA in the absence of actRNA, which was significantly higher than that of the full-length guide system ($0.12 \pm \text{SEM } 0.02$ AU/s and $0.051 \pm \text{SEM } 0.003$ AU/s, respectively, from Fig. 1d). Since the capRNA could be activating the Cas13a RNP in the absence actRNA, we tested

whether different lengths and sequences of the capRNA (Fig. 2b) and concentrations (Fig. 2c) could reduce the background signal.

We tested four different capRNAs with varied lengths and sequences but constant anchor and capture regions (Fig. 2b). We compared our original 36nt capRNA (capRNA 1) with a new 36nt capRNA (capRNA 2). Both had the same 10nt anchor and 10nt capture regions, but different sequences for the 3' flanking region (5nt) and the 5' tail (11nt). CapRNAs 1 and 2 had drastically different levels of background activity ($1.2 \pm \text{SEM } 0.1 \text{ AU/s}$ and $0.08 \pm \text{SEM } 0.01 \text{ AU/s}$, respectively) but similar levels of activity in the presence of the 10nt actRNA ($1.5 \pm \text{SEM } 0.04 \text{ AU/s}$ and $1.6 \pm \text{SEM } 0.1 \text{ AU/s}$, respectively), such that their background-subtracted reaction rates were significantly different ($p = 0.0050$; Extended Data Fig. 2a). Of note, capRNA 2 had such a high background that the positive reaction rate was not significantly different from the background ($p = 0.0503$). Similarly, when the 5' tail was increased to 18nt (capRNA 3), the 10nt actRNA reaction rate was not significantly different from the background ($p = 0.3969$). However, when the 5' tail was reduced to 7nt (capRNA 4), the background reaction rate decreased such that the 10nt actRNA reaction rate was significantly higher ($p = 0.0003$). When considering the difference in the positive signal over the background, capRNA 4 (7nt 5' tail) performed significantly better than sequence-matched capRNA 2 (11nt 5' tail) and capRNA 3 (18nt 5' tail) but performed similarly to capRNA 1 (original 11nt 5' tail) (Extended Data Fig. 2a). We further tested capRNA 1 to evaluate the importance of the 3' flanking sequence and the 5' tail and found that deleting either 3' flank or 5' tail reduced the 10nt actRNA reaction rate (Extended Data Fig. 2c).

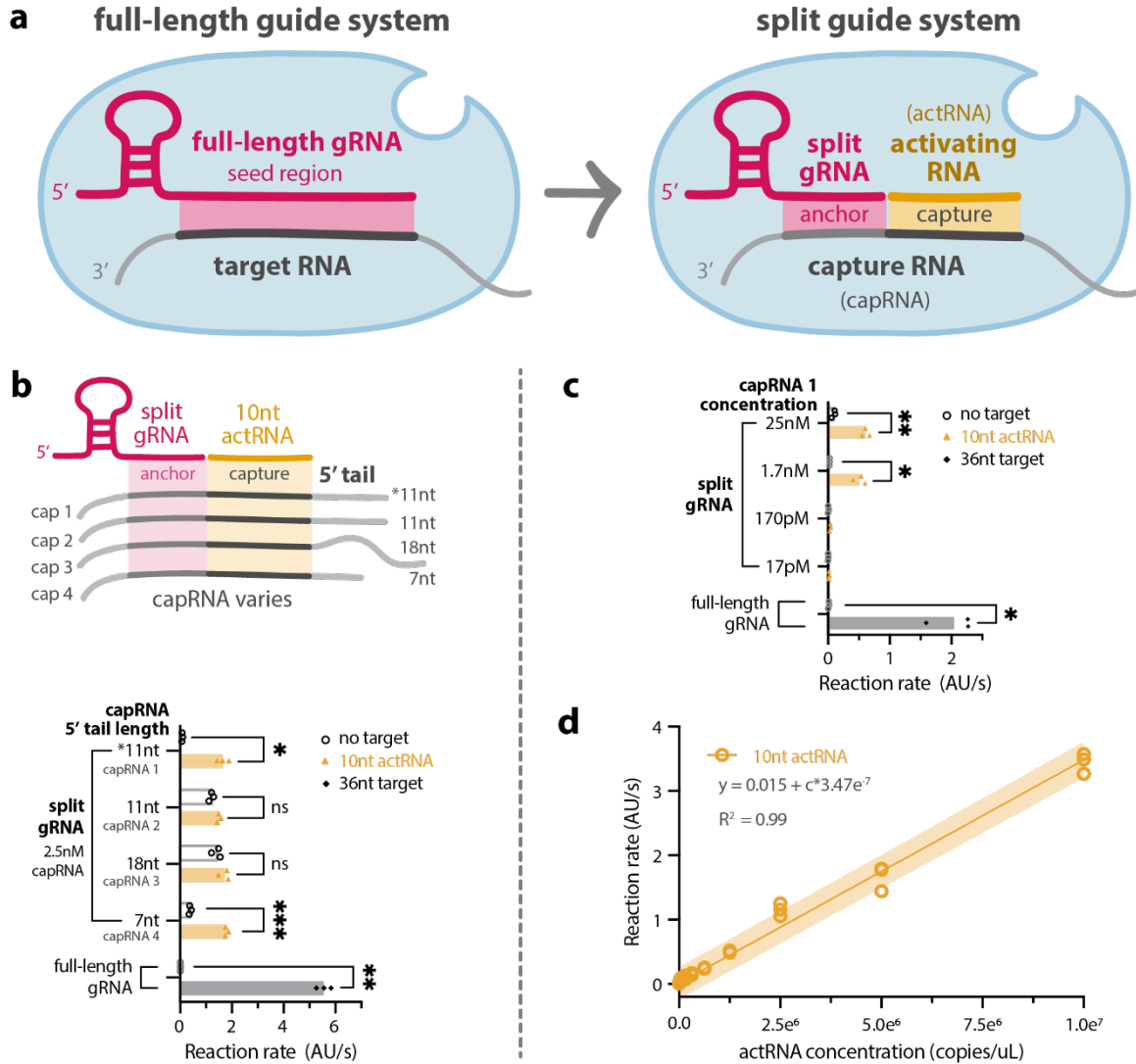


Figure 2. Cas13a split guide system detects 10nt ssRNA at femtomolar levels.

A, Schematic illustrating the standard full-length guide system (left) and introducing the split guide system and nomenclature (right). The full-length gRNA is split in two in the seed region. The first part of the gRNA is referred to as the split gRNA (hot pink). The second part of the gRNA is now called the activating RNA, or actRNA (yellow); it is the limiting reagent of the system. In the split guide system, the target RNA becomes the capture RNA, or capRNA. The capRNA complements to the split gRNA via the anchor region (light pink) and the actRNA via the capture region (pale yellow). The capRNA is provided as a reagent in the system. B, Top: Schematic illustrating the split guide system with split gRNA (10nt anchor) and 10nt actRNA bound to different capRNAs. The sequence of the anchor and capture regions are held constant, but the length and sequence of the 5' tails varies. CapRNA 1 has a 5nt 3' flanking sequence and an 11nt 5' tail. CapRNAs 2, 3, and 4 share the same sequence for the 3' flank and 5' tail, but they are different sequences than capRNA 1. Bottom: For each capRNA species, the reaction was tested with 10nt actRNA (yellow bars, triangles) or without actRNA (grey bars, empty circles). The Cas13a and split gRNA are present at 25nM, and the capRNA is at 2.5nM. The reaction rate (AU/s) is plotted with and without 10nt actRNA at 1e6 copies/uL. For the full-length guide system, the gRNA is equimolar to Cas13a at 25nM, and the 36nt target is at 1e6 copies/uL. The mean reaction rate (AU/s) with three technical replicates is plotted. C, Test of the split guide system with varied capRNA concentrations to determine the effects on actRNA detection over background signal. For the split guide system, capRNA 1 concentration was tested ranging from 25nM to 17pM. At each capRNA concentration, the reaction was tested with actRNA (yellow bars, triangles) or without actRNA (grey bars, empty circles). Cas13a and split gRNA are at 25nM. The 10nt actRNA is at 1e6 copies/uL in the final reaction. For the full-length guide system, gRNA is equimolar to Cas13a at 25nM and the 36nt target is at 1e6 copies/uL. The mean reaction rate (AU/s) with three technical replicates is plotted. D, Limit of detection of the 10nt actRNA determined with optimized reaction conditions. The LOD of the 10nt actRNA was tested with Cas13a equimolar to the split gRNA (10nt anchor) at 25nM and capRNA 1 at 2.5nM. The concentration of 10nt actRNA was tested at 2-fold

dilutions from 1e7 copies/uL to 4e4 copies/uL. The reaction rate (AU/s) is plotted for each of three technical replicates across the range of actRNA concentrations. The linear regression fit to the reaction rate data is plotted with 95% prediction bands. Stats: For B and C, we used pairwise analysis of the reaction rates for each experimental condition with and without actRNA using a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for multiple comparisons. We report multiplicity-adjusted p-values. Non-significant interactions ($p > 0.05$) are not plotted. * is $p \leq 0.05$, ** is $p \leq 0.01$, *** is $p \leq 0.001$. See Extended Data Figs. 2 and 3 for more details.

To further optimize the split guide system, we tested capRNA 1 at range of concentrations in the reaction mixture. We found that decreasing the capRNA concentration from 25nM to 1.7nM (~15x decrease) reduced the activity in the presence and absence of the 10nt actRNA (Fig. 2c), such that when we subtracted the background activity, the reaction rates were not significantly different between the 25nM and 1.7nM capRNA conditions ($0.52 \pm \text{SEM } 0.04 \text{ AU/s}$ and $0.49 \pm \text{SEM } 0.06 \text{ AU/s}$, respectively; Extended Data Fig. 2b). Further decreasing the capRNA concentration to 170pM and 17pM decreased the background to practically zero, but that also decreased the positive signal such that it was not significantly higher than the background. In subsequent experiments, we used 25nM or 2.5nM capRNA, depending on whether we wanted to maximize the positive signal (25nM) or minimize the background signal (2.5nM).

With capRNA 1 at 2.5nM, our estimated LOD for a 10nt actRNA was 1.2e5 copies/uL (Fig. 2d), significantly better than our inability to detect a 10nt target RNA with a full-length gRNA and comparable to the LOD of a 20nt target RNA in the full-length guide system (Fig. 1b). Unless otherwise noted, we use capRNA 1 for subsequent experiments.

Split guide system detects 8 to 24nt-long actRNA

We next investigated the limits of short ssRNA detection with the split guide system and its robustness to variations in the anchor and capture regions. The initial configuration for testing the split guide system used a split gRNA with a 10nt anchor region and a 10nt actRNA, which together replaced the 20nt seed region of a full-length gRNA. We tested whether the split gRNA system could detect actRNAs with different lengths by varying the ratio of anchor to actRNA length while maintaining a constant 20nt combined length and sequence (Fig. 3a, top). We varied the length of the anchor region from 5nt to 12nt, complemented with an actRNA ranging from 15nt to 8nt. When adding actRNA at 1e7 copies/uL, we saw significant activation over background from an anchor region as small as 9nt and as long as 12nt (Fig. 3a, bottom, and Extended Data Fig. 4a). The greatest activation over background was from the 10nt anchor plus 10nt actRNA and the 12nt anchor plus 8nt actRNA ($2.9 \pm \text{SEM } 0.2 \text{ AU/s}$, and $3.1 \pm \text{SEM } 0.4 \text{ AU/s}$, respectively). The 12nt anchor, however, had the highest background activation, so we selected the 10nt anchor for further testing (Extended Data Fig. 4a).

With the 10nt anchor split gRNA, we next tested detection of actRNA of increasing length (Fig. 3b, top). The Cas13a split guide system was able to detect actRNA ranging from 10 to 24nt long when paired with an appropriate capRNA (Fig. 3b, bottom). The background-subtracted signal varied for each actRNA depending on which capRNA it was paired with (Extended Data Fig. 4c). Of note, the 20nt actRNA resulted in a background-subtracted signal much greater than a 20nt target RNA detected with the conventional full-length guide system (Fig. 3b, bottom, and Fig. 1b), indicating that the split guide system can outperform the full-length guide system. Together, these results demonstrate the key role the capRNA plays in the effectiveness of the split guide system. With judicious selection of the split gRNA (anchor length and sequence) and capRNA (length and sequence), the split guide system can detect actRNA ranging from 8nt to at least 24nt.

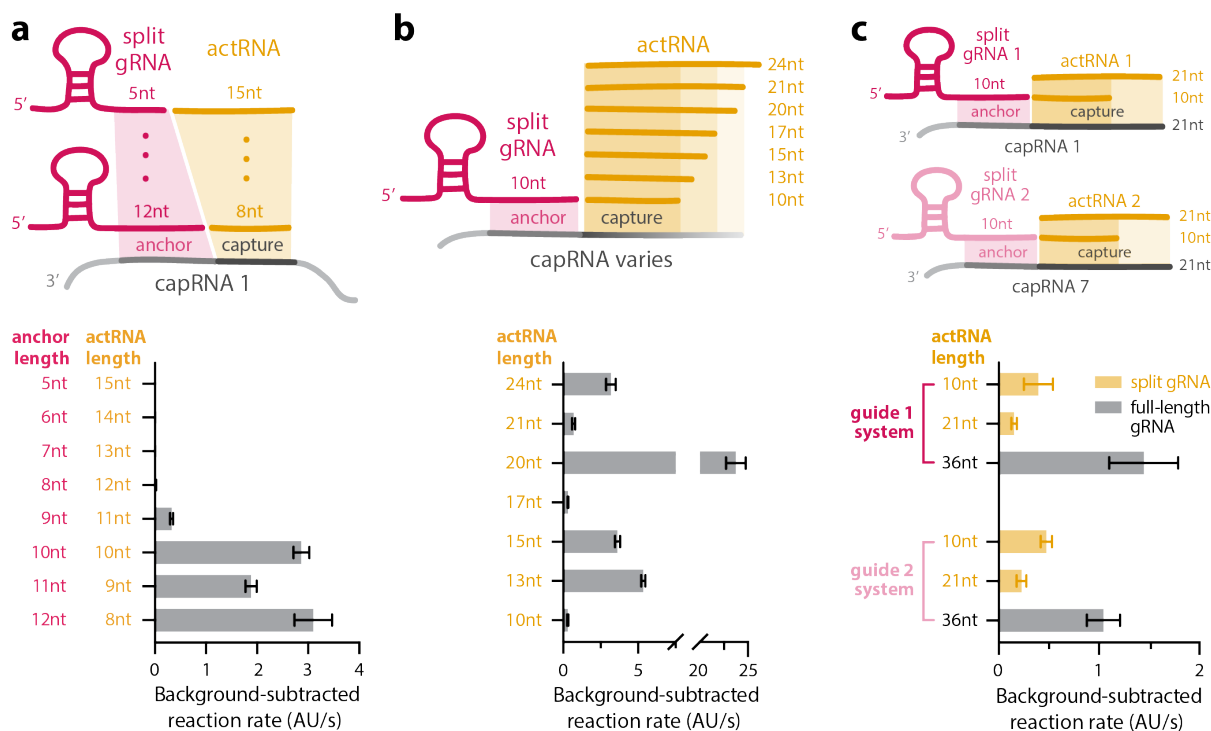


Figure 3. The split guide system works across a range of RNA lengths and sequences.

A, Top: Schematic illustrating the split gRNA (hot pink) with varied anchor region length (ranging from 5 to 12nt) and the corresponding actRNA (yellow, ranging from 15 to 8nt) such that the anchor plus capture regions remains constant at 20nt. Bottom: The reaction rates (AU/s) of different split gRNA and actRNA pairs are plotted with the mean background activity of the Cas13a RNP and capRNA alone subtracted. The Cas13a, split gRNA, and capRNA 1 are equimolar at 25nM. All actRNA is at 1e7 copies/uL in the final reaction. The mean reaction rate (AU/s) with mean background activity subtracted is plotted with SEM (n=3). B, Top: Schematic illustrating the split guide system with actRNA (yellow) of increasing length, ranging from 10nt to 24nt. Each actRNA is paired with a capRNA for optimal signal. Bottom: The reaction rates (AU/s) of different actRNA-capRNA pairs are plotted with the mean background activity of the Cas13a RNP and capRNA alone subtracted. The 10, 13, 15, and 17nt actRNAs are paired with capRNA 4; the 20, 21, and 24nt actRNAs are paired with capRNA 2. The Cas13a is equimolar to the split gRNA (10nt anchor) at 25nM. All capRNA is at 2.5nM, and all actRNA is at 1e6 copies/uL in the final reaction. The mean reaction rate (AU/s) with mean background activity subtracted is plotted with SEM (n=3). C, Top: Schematic illustrating two different split guide systems. Split gRNA 1 (hot pink) with actRNA 1 (yellow) and capRNA 1 (grey) are the same sequence as previously tested in this paper. Split gRNA 2 (light pink) with actRNA 2 (yellow) and capRNA 7 (grey) are distinct RNA sequences from split guide system 1. Both systems are tested with a 10nt and 21nt actRNA and compared to their sequence-matched full-length gRNA variants. Bottom: The split guide system works with different RNA sequences. Split gRNA 1 is tested with a 10nt and 21nt actRNA and compared to full-length gRNA 1 with a 36nt target. The same is shown for split gRNA 2 and full-length gRNA 2. All reaction rates are reported with mean background activity subtracted. For the split guide systems, Cas13a is equimolar to split gRNA at 25nM, capRNA is at 2.5nM, and actRNA is at 1e6 copies/uL. For full-length guide systems, Cas13a is equimolar to gRNA at 25nM, and target RNA is at 1e6 copies/uL. The mean reaction rate (AU/s) with mean background activity subtracted is plotted with SEM (n=3). Stats: In B, for each actRNA, the background-subtracted reaction rate is compared between the two capRNA conditions using a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for multiple comparisons. We report multiplicity-adjusted p-values. Non-significant interactions (p>0.05) are not plotted. * is p≤0.05, ** is p≤0.01. See Extended Data Figs. 4 and 5 for more details.

Finally, we wanted to confirm that the split gRNA performance was not sequence dependent and would work with different capture and anchor region sequences. We built a second split guide system with a capRNA and split gRNA based on a previously validated gRNA (full-length gRNA 2, 20nt seed; see Supplementary Information Table 1 for sequence information). Split gRNA 2 had the same stem-loop sequence as split gRNA 1, but a different 10nt anchor region (Fig. 3c, top, and Extended Data Fig. 5a and b). The capture regions for the capRNAs in both systems were the same length but different sequences. For both split gRNAs, we tested a 10nt and 21nt actRNA at 1e6 copies/uL. Split

guide system 2 detected both actRNAs above background and showed the same trend as the split guide system 1 where the 10nt actRNA resulted in higher activity than the 21nt actRNA (Fig. 3c, bottom, and Extended Data Fig. 5c). However, when comparing the activity of the 10nt and 21nt actRNAs relative to the full-length control, the split guide system 2 performed proportionally better than the split guide system 1 (Extended Data Fig. 5d). These experiments demonstrate that the split guide system can work with alternative capture and anchor region sequences, and that screening for high-activity sequences could improve detection of miRNAs.

Split guide system is specific against RNA mismatches and misalignments

For the Cas13a split guide system to be a useful strategy for molecular diagnostics and basic RNA biology, it must be specific to the sequence of interest. Specificity is particularly important for small RNA detection, since many small RNAs exist in modified form. MicroRNAs, for example, are present in many isoforms (isomiRs) that are functionally different³⁸. These isoforms may include mutations within the canonical sequence, like A-to-I or C-to-U RNA edits, as well as nucleotide additions and truncations on both the 5' and 3' ends³⁸. Furthermore, miRNAs and other small RNAs may contain RNA modifications which also alter their function⁹. We sought to characterize the specificity of the split guide system to on-target vs off-target actRNA, looking at mismatches between the actRNA and capRNA, as well as extensions and truncations of the actRNA.

We first tested our system for specificity in the presence of 2nt consecutive mismatches. We tiled 2nt mismatches across the first 10nt of a 17nt actRNA and paired them with capRNA 5 (10nt capture region) (Fig. 4a, top). When we added actRNA at 1e7 copies/uL, which produces a strong signal in the no mismatch condition, we saw a total reduction in signal for all the mismatch conditions (Fig. 4a, bottom). We also tested specificity for mismatches with capRNA 4 (17nt capture region) that binds the entire 17nt actRNA and saw a partial reduction in activity for all mismatch conditions except when the mismatch was at positions 3-4 (Extended Data Fig. 6a, right).

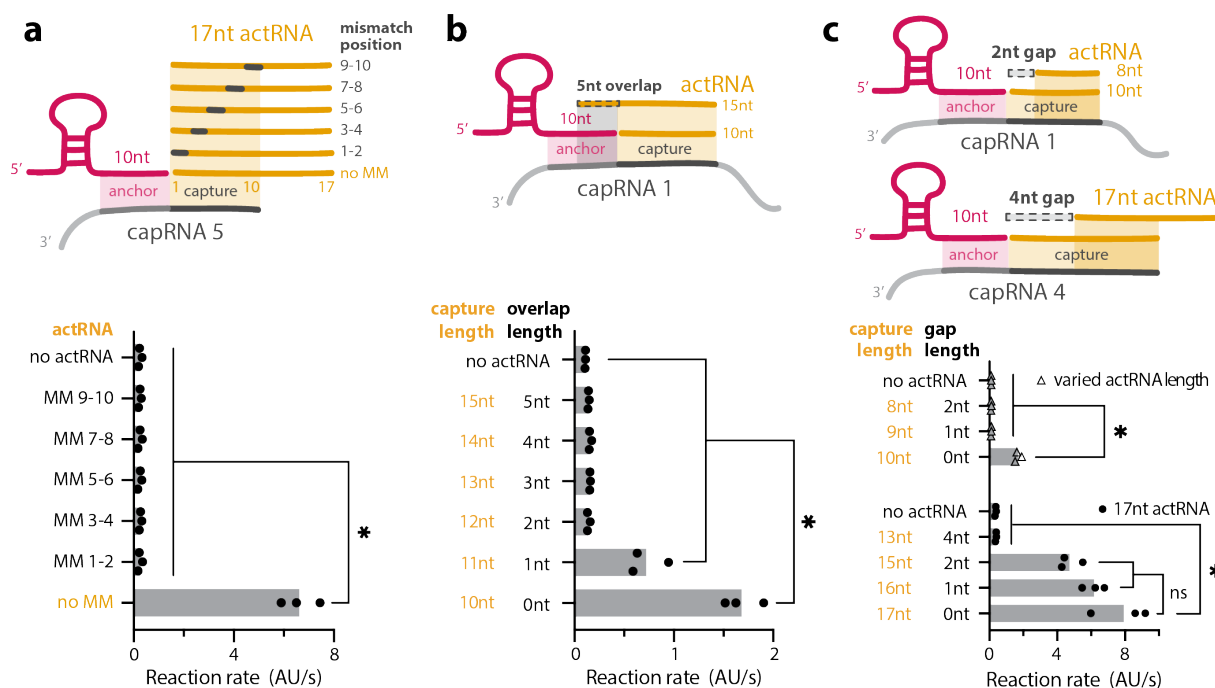


Figure 4. The split guide system is specific against RNA mismatches and misalignments.

A, Top: Schematic illustrating test of mismatch sensitivity. The split gRNA (hot pink, 10nt anchor) and 17nt actRNA (yellow) containing a 2nt mismatch tiled across the first 10nt of the actRNA are bound to capRNA 5. The capRNA contains a 10nt anchor region and a 10nt capture region, such that only the first 10nt of the actRNA bind to the capRNA, leaving 7nt unbound on the 3' end of the actRNA. Bottom: The split guide system is tested with 17nt actRNA containing 2nt mismatches. Cas13a is equimolar to the split gRNA at 25nM, capRNA 5 is at 2.5nM, and all actRNA is at 1e7 copies/uL. The mean reaction rate (AU/s) is plotted with three technical replicates. All reactions are compared to the actRNA with no mismatches. B, Top: Schematic illustrating test of overlap sensitivity. Split gRNA (hot pink, 10nt anchor) is paired with actRNA (yellow) ranging from 10 to 15nt, with the actRNA overlapping the anchor region of the split gRNA up to 5nt. The region of the actRNA that overlaps the split gRNA has the same sequence as the split gRNA. The split gRNA and actRNA complement to capRNA 1. Bottom: The split guide system is tested with actRNA that increasingly overlaps the anchor region of the split gRNA. Cas13a is equimolar to split gRNA at 25nM, capRNA 1 is at 2.5nM, and all actRNA is at 1e7 copies/uL. The mean reaction rate (AU/s) is plotted with three technical replicates. All reactions are compared to the actRNA with no overlap (overlap length = 0). C, Top: Schematic illustrating test of gap sensitivity. In top schematic, split gRNA (hot pink, 10nt anchor) is shown with actRNA (yellow) of decreasing lengths: a 10nt actRNA with no gap and an 8nt actRNA with a 2nt gap between the 3' end of the split gRNA and the 5' end of the actRNA. All are in complex with capRNA 1. The bottom schematic shows the same split gRNA adjacent to a 17nt actRNA. Multiple 17nt actRNAs are tested with a gap between the 3' end of the split gRNA and the 5' end of the actRNA. The gap size ranges from 0nt to 4nt. The split gRNA and 17nt actRNAs are in complex with capRNA 4 containing a 17nt capture region. Thus, as the gap size increases, the portion of the actRNA that binds to the capRNA decreases (for a gap of 4nt, only the first 13nt of the actRNA bind). Bottom: The split guide system is tested for sensitivity to gaps between split gRNA and actRNA. The top four bars (triangles) show the split gRNA with actRNA of decreasing length (10nt to 8nt) bound to capRNA 1. The bottom five bars (circles) show the split gRNA with a 17nt actRNA bound to capRNA 4. Cas13a is equimolar to the split gRNA at 25nM, capRNA is at 2.5nM, and all actRNA is at 1e7 copies/uL. The mean reaction rate (AU/s) is plotted with three technical replicates. All reactions are compared to the actRNA with no gap (gap length = 0). Stats: In A, B, and C, the mean reaction rate of experimental conditions is compared to the appropriate control (for A, the no mismatch actRNA, for B, the no gap actRNA, and for C, the no overlap actRNA) using a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for multiple comparisons. We report multiplicity-adjusted p-values. Non-significant interactions are $p > 0.05$, * is $p \leq 0.05$. See Extended Data Figs. 6 for more details.

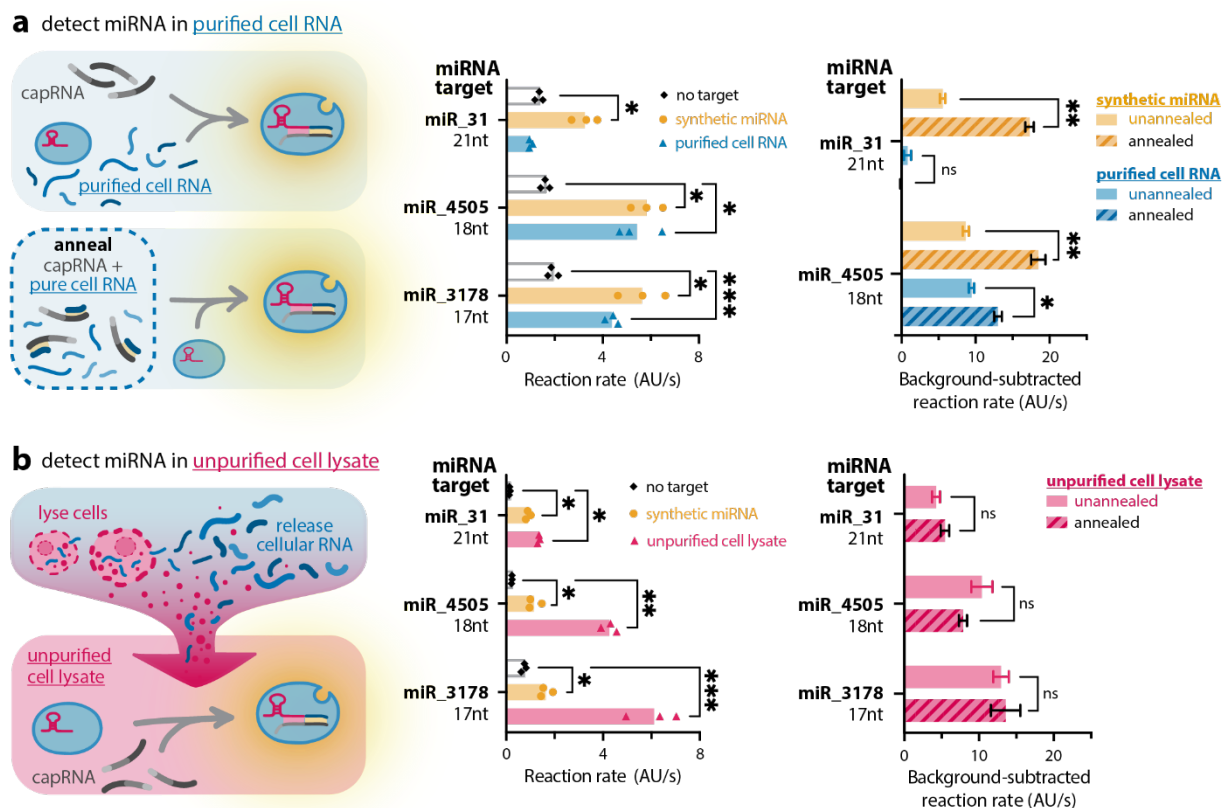
We next investigated how the split guide system performed when probed with ssRNA segments that were misaligned with the capture region. We first tested actRNA that extended into the split gRNA, resulting in a 1-5nt overlap (Fig. 4b, top). We hypothesized that this overlap would sterically hinder the actRNA from nestling in the groove of the Cas13a RNP's nuclease lobe, since the spacing between the split gRNA and actRNA is controlled by the sequence of the capRNA. At 1e7 copies/uL of actRNA, we saw significant reduction in signal with a 1nt overlap and complete reduction in signal for 2-5nt overlaps (Fig. 4b, bottom). We then tested the effect of truncated and misaligned actRNA that introduced gaps between the end of the split gRNA and the start of the actRNA. We either truncated the 5' end of a 10nt actRNA to create 1-2nt gaps or misaligned a 17nt actRNA to introduce 1-4nt gaps (Fig. 4c, top). The truncated actRNAs with a 1 or 2nt gap both showed a complete reduction in activity relative to the full-length actRNA with no gap (Fig. 4c, bottom). Importantly, since both the 8 and 9nt actRNAs are detectable with our system (Fig. 3a), the presence of a gap between the actRNA and split gRNA is likely responsible for the loss of activity. With the misaligned 17nt actRNA, we saw a partial reduction in activity for a 1 and 2nt gap, and a complete reduction to background levels for a 4nt gap (Fig. 4c, bottom). This sensitivity to gaps is consistent with other studies of Cas13a mismatch sensitivity³⁹. These results show that the split guide system has high specificity against actRNA containing mismatches and modest specificity against actRNA that are misaligned with the capture region.

Split guide system detects synthetic and endogenous cellular miRNA

To demonstrate the split guide system's ability to work in a more complex matrix, we first tested synthetic RNA detection in a cellular RNA background, which introduced many off-target RNA species. We extracted total cellular RNA from Lenti-X HEK 293T cells and added it to our split guide reaction at 1ng/uL (Extended Data Fig. 6b, right panel). To that matrix, we spiked in synthetic 20nt actRNA at a range of concentrations. The presence of the cellular RNA background resulted in a

slightly increased signal compared to the no cellular RNA control (linear regression fit to each condition have significantly different slopes, $p = 0.0013$) (Extended Data Fig. 6b, right panel). Spiking in higher concentrations of cellular RNA resulted in a reduced signal, which is expected, because the Cas13a RNP cleaves both the reporter molecule and cellular RNA, diluting the effective concentration of the reporter molecule (Extended Data Fig. 6b, middle panel).

Next, we tested whether the split guide system could detect known cellular miRNAs. We selected six miRNA targets ranging in length from 17 to 23nt that are found in HEK 293 cells⁴⁰. We designed a capRNA for each miRNA containing a 5nt 3' flank, 10nt anchor region, and a capture region that fully complements the miRNA with no 5' tail. We then screened the capRNA to identify those that had low activity in the absence of the cellular miRNA (Extended Data Fig. 6c). We selected the three capRNAs with the lowest background activity for further testing. For each capRNA, we added either synthetic miRNA at $1e7$ copies/uL or purified cellular RNA from Lenti-X HEK 293T cells at $1ng/uL$ (Fig. 5a, left and middle panels). For all three miRNA targets, the synthetic miRNAs were detected above background (yellow bars). MiR_3178 (17nt) and miR_4505 (18nt) were also detected above background in the purified cellular RNA sample (blue bars). Pre-annealing the capRNA with either the synthetic miRNA or purified cellular RNA further boosted the signal for all conditions tested, except miR_31 in purified cellular RNA (Fig. 5a, right panel, striped vs solid bars).



A, Left: Schematic illustrating miRNA detection by the split guide system. CapRNA and purified cellular RNA were added to Cas13a RNPs containing a split gRNA (10nt anchor) (top schematic). Alternately, capRNA and purified cell RNA were annealed before adding to the Cas13a RNP and reaction mixture (bottom schematic). Middle: Three miRNAs were tested: miR_31 (21nt), miR_4505 (18nt), and miR_3178 (17nt). For each miRNA target, the mean reaction rate (AU/s) is plotted with three technical replicates for the condition with no activating RNA (Cas13a RNP and capRNA only; grey bars, rhombuses), synthetic miRNA (yellow bars, circles), and purified

cellular RNA (blue bars, triangles). For all conditions, Cas13a is equimolar to the split gRNA at 25nM, and the capRNA is at 2.5nM. Synthetic miRNA is added at a final concentration of 1e7 copies/uL; purified cellular RNA from Lenti-X HEK 293T cells is added at a final concentration of 1ng/uL. All reactions are compared to their specific no target control. Right: Detection of miR_31 and miR_4505 was tested with a pre-annealing step. The mean background-subtracted reaction rates (AU/s) of annealed and unannealed conditions are plotted with three technical replicates. The graph shows the background-subtracted reaction rate of synthetic miRNA (yellow bars), synthetic miRNA annealed to capRNA (striped yellow bars), purified cellular RNA (blue bars), and purified cellular RNA annealed to capRNA (striped blue bars). For all conditions, Cas13a is equimolar to the split gRNA at 25nM, and the capRNA is at 2.5nM. Synthetic miRNA is added at a final concentration of 1e7 copies/uL; purified cellular RNA from Lenti-X HEK 293T cells is added at a final concentration of 1ng/uL. Detection of miRNA is compared with and without annealing for each miRNA target and source. B, Left: Schematic illustrating addition of unpurified cell lysate, containing all cellular RNA to the Cas13a reaction mixture. Middle: Detection of miR_31, 4505, and 3178 in whole cell lysate. For each miRNA target, the mean reaction rate (AU/s) is plotted with three technical replicates for the condition with no target (Cas13a RNP and capRNA only; grey bars, rhombuses), synthetic miRNA (yellow bars, circles), and unpurified cell lysate (pink bars, triangles). For all conditions, Cas13a is equimolar to the split gRNA at 25nM, and the capRNA is at 2.5nM. Synthetic miRNA is added at a final concentration of 1e7 copies/uL; unpurified cell lysate from Lenti-X HEK 293T cells is added at a final concentration of 10 cells/uL. All reactions are compared to their specific no target control. Right: Detection of all three miRNAs was tested with pre-annealing of the unpurified cell lysate and capRNA. The mean background-subtracted reaction rate (AU/s) of annealed and unannealed cell lysate conditions is plotted with three technical replicates. The graph shows the background-subtracted reaction rate of unpurified cell lysate (pink bars) and unpurified cell lysate annealed to capRNA (striped pink bars). For all conditions, Cas13a is equimolar to the split gRNA at 25nM, and the capRNA is at 2.5nM. Synthetic miRNA is added at a final concentration of 1e7 copies/uL; unpurified cell lysate from Lenti-X HEK 293T cells is added at a final concentration of 10 cells/uL. Detection of miRNA is compared with and without annealing for each miRNA target in unpurified cell lysate. Stats: We used pairwise analysis of the reaction rates for each experimental conditions with and without actRNA (D, F middle) or with and without annealing (E, F right) using a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for multiple comparisons. For all data sets, we report multiplicity-adjusted p-values. Non-significant values are $p > 0.05$, * is $p \leq 0.05$, ** is $p \leq 0.01$, and *** is $p \leq 0.001$. See Extended Data Fig. 6 and 7 for more details.

Finally, we tested the detection of endogenous cellular miRNAs in unpurified cell lysate (Fig. 5b). We lysed Lenti-X HEK 293T cells and added unpurified cell lysate to the reaction mixture at a final concentration of 10 cells/uL (Fig. 5b, left panel). In the cell lysate, we detected all three miRNAs (miR_31, 4505, and 3178) above background (Fig. 5b, middle panel, pink bars), confirming the ability of the Cas13a split guide system to directly detect miRNAs without RNA purification. To confirm that the signal was not due off-target activity, we tested a capRNA that shouldn't bind any ssRNA in HEK cells. As expected, we did see a modest increase in background signal when adding cell lysate (Extended Data Fig. 6d), but even accounting for that increase in background activity, we still observe detection of the miRNA targets in cell lysate. We also tested whether pre-annealing the capRNA with the cell lysate would boost the signal, but we found no significant difference in the background-subtracted reaction rate between the unannealed and annealed conditions (Fig. 5b, right panel, striped vs solid bars).

Discussion

Accurately detecting the full range of small RNAs, down to the smallest tyRNA, is necessary to leverage the potential of small RNAs for disease diagnosis and to better study their biology, function, and regulation. Cas13a has been used to detect ssRNA targets with high sensitivity, but detecting RNA targets shorter than 21nt has been a challenge. Critically, this lower-length limit of detection excludes many small RNAs; an estimated 15.5% (411 of 2656) of documented human miRNAs are fewer than 21nt (see Extended Data Fig. 8, data from miRbase release 22.1⁴¹). Furthermore, it is likely that there are other small RNAs, like tyRNAs that have yet to be studied for their diagnostic potential due to their small size.

Here we present a Cas13a split guide system that extends the lower-length limit of target RNA detectable by Cas13a, enabling the use of Cas13a for detection of known miRNAs and tyRNAs without the need for purification, reverse transcription, or amplification. The Cas13a split guide system uses a gRNA with a short seed region (split gRNA) paired with a capture RNA (capRNA) to

detect RNA targets ranging from 8 to 24nt. Only the capRNA needs to change to detect a new short RNA target; thus, a unique panel of small RNAs can be detected using the same Cas13a RNPs mixed with a customized set of capRNAs. Furthermore, the split guide system retains sufficient sensitivity for diagnostic use, even with small targets. We show that it can detect a 10nt target with an estimated LOD of 1.2e5 copies/uL (200fM), which is within the range of concentrations that clinically-relevant miRNAs are found in serum^{19,20}. We also show that the split guide system is specific for the desired actRNA, exhibiting no or limited activity for 2nt mismatches and actRNA misalignments. We further demonstrate applications of the split guide system to directly detect known human miRNAs within purified cellular RNA and unpurified cell lysate, proving that the split guide system can detect miRNAs without RNA purification. For the miRNAs detected in unpurified cell lysate, we estimate their concentration to be approximately 1-2e6 copies/cell. Together, the split guide system expands the repertoire of RNA detectable with Cas13a, raising the possibility of direct detection of miRNAs and other small RNA for diagnostic purposes and to support functional studies of small RNAs.

As with conventional uses of Cas13a, sensitive ssRNA detection by the split guide system depends on low activity of the Cas13a RNP in the absence of the activating RNA. We found that this background activity increased with the length of the anchor region when paired with the same capRNA (Extended Data Fig. 4a). This increased background may be due in part to the increased stability of the split gRNA bound with the capRNA, as suggested by an increase in melting temperature of the anchor region (Extended Data Fig. 4b). However, melting temperature does not fully explain the stability, which may also be due to the Cas13a enzyme itself. Tambe et al. found that the binding affinity of a target RNA containing consecutive mismatches was affected by the location of mismatches, with a maximum reduction of binding affinity occurring when the mismatch was in the central portion of the gRNA seed region (positions 9-12)³⁹. This is consistent with our findings that the split gRNA with a 12nt anchor region (no unbound nucleotides in the 9-12 position) had a background activity more than 10x that of the split gRNA with an 11nt anchor region (1 unbound nucleotide in position 12) (Extended Data Fig. 4a).

Background activity of the split guide RNP is also affected by details of the capRNA. We varied the length and sequence of the capRNA outside of the anchor and capture regions and found no consistent trends with either variable (Fig. 2b). Rather, the length and sequence contribute to capRNA secondary and tertiary structure, which could sterically affect complexing with the split gRNA and actRNA. For example, capRNA that forms stable same-species duplexes would be unlikely to bind either the split gRNA or actRNA, resulting in poor actRNA sensitivity. Alternately, capRNAs that form stem-loops in the capture region may be sufficient to activate the RNP when bound via the anchor region, such as with capRNAs 3 and 7, along with most capRNAs designed for the cellular miRNA targets, resulting in high background activity (Extended Data Figs. 3, 5b, and 7). Further capRNA design iterations to reduce the probability of a stem-loop structure forming may boost the signal-to-background ratio. However, secondary and tertiary structures, as predicted with NUPACK⁴², are not sufficient to explain all the variability we observe. For example, capRNA 1 and 2, which both have a 5nt 3' flank and 11nt 5' tail but with different sequences, have dramatically different background activities (Fig. 2b). The predicted structure of these capRNAs bound with the split gRNA does not include any stem-loop structures that might lead to unwanted activation (Extended Data Fig. 3). This suggests that the sequence of the 3' flank and 5' tail may also play a role in background activation.

We observed that different combinations of the complete split guide system – actRNA, capRNA, and split gRNA – resulted in varied reaction rates, consistent with the variation in activity observed in the conventional Cas13a system for different guide and target RNA combinations. This may be partially

explained by differences in the stability of binding in the capture region, as it was in the anchor region. Comparing split guide systems 1 and 2, the capRNA for split gRNA 2 had a slightly higher melting temperature than that for split gRNA 1 and generated a higher relative signal for the same length actRNA targets (Extended Data Fig. 5e). We also observe this trend of increased signal when the capture region length, and associated melting temperature, is increased such that it binds more of a given actRNA (Extended Data Fig. 4d). Even when controlling for actRNA and capture and anchor region identity, we see differences in capRNA performance. We tested three capRNAs with the same anchor and capture regions, but varied whether they had a 3' flank and/or a 5' tail (Extended Data Fig. 2c). All three capRNAs had similarly low background activity, but resulted in significantly different activity when a 10nt actRNA was added. The capRNA with both the 3' flanking sequence and 5' tail resulted in the greatest reaction rate.

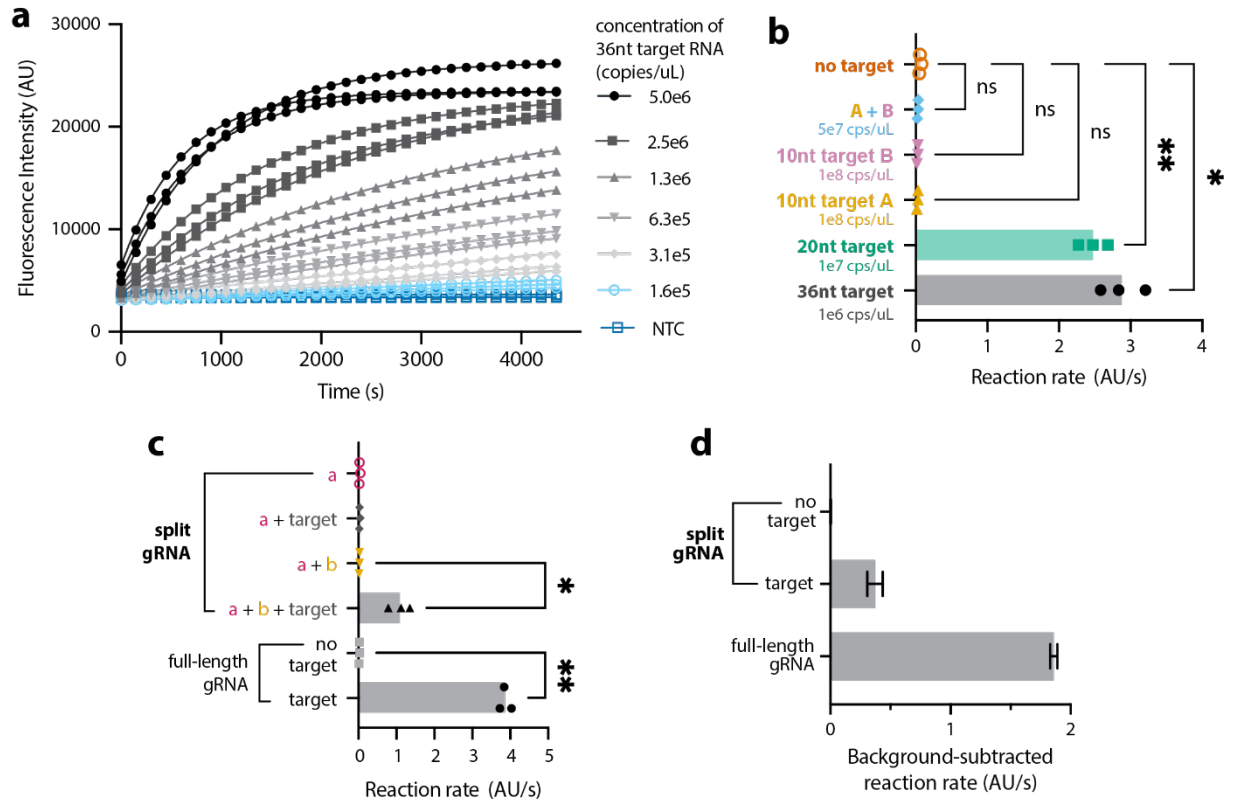
This analysis suggests that capRNA design provides an opportunity for improving actRNA detection with the split guide system. Designs aimed at optimizing signal-to-background ratio should consider sequence, melting temperature of the anchor and capture regions, and secondary and tertiary structure of the complex. The programmability of the capRNA, however, is an asset to the system, as it allows the split guide system to be tuned based on the needs of the assay. For applications that require high specificity, such as detecting isomiRs, designing a shorter capRNA to bind the region of interest will increase specificity to mismatches in the capture region (Fig. 4a, Extended Data Fig. 6a). Alternatively, for applications that require high sensitivity, increasing the length of the capture region will likely enhance detection at low concentrations.

It is interesting to note that the length of the target actRNA is not correlated with activation, at least over the range of 8 to 24nt explored in this study. For example, we show in two separate sequence systems that a 10nt actRNA results in a significantly higher signal than a 21nt actRNA using the same capRNA (Fig. 3c). When probing the size landscape of actRNAs detectable with the split guide system, we observed that the 13, 15, and 20nt actRNAs generated a drastically higher signal than either the 10, 21, or 17nt actRNAs (Fig. 3b). The variation in signal across the size landscape of actRNAs suggests that there might be multiple configurations of the dsRNA region (seed region) that induce the conformational change of the Cas13a RNP required for transcleavage activity. The split of the gRNA may even enable the Cas13a RNP to have a higher transcleavage rate than the conventional full-length gRNA system; the signal generated from a 20nt actRNA in the split guide system is more than 30x that of a 20nt target RNA detected with the full-length guide system (Fig. 3b and 1b). These findings suggest that a split gRNA may be a useful strategy for detecting both short and long RNA targets.

While the Cas13a split guide system is a promising approach to detect miRNAs and other small RNA species, our testing thus far has focused mainly on idealized conditions using synthetic miRNAs. To explore the effect of a complex matrix, we carried out proof-of-concept detection of miRNAs in purified cellular RNA and unpurified cell lysate. Interestingly, the split guide system was able to directly detect endogenous miR_31 in the cell lysate but was not able to detect it in the purified RNA sample (Fig. 5b and 5a). This difference in detection suggests that there could be important loss of miRNA during the extraction and purification steps, further motivating the need to directly detect miRNAs with minimal processing steps.

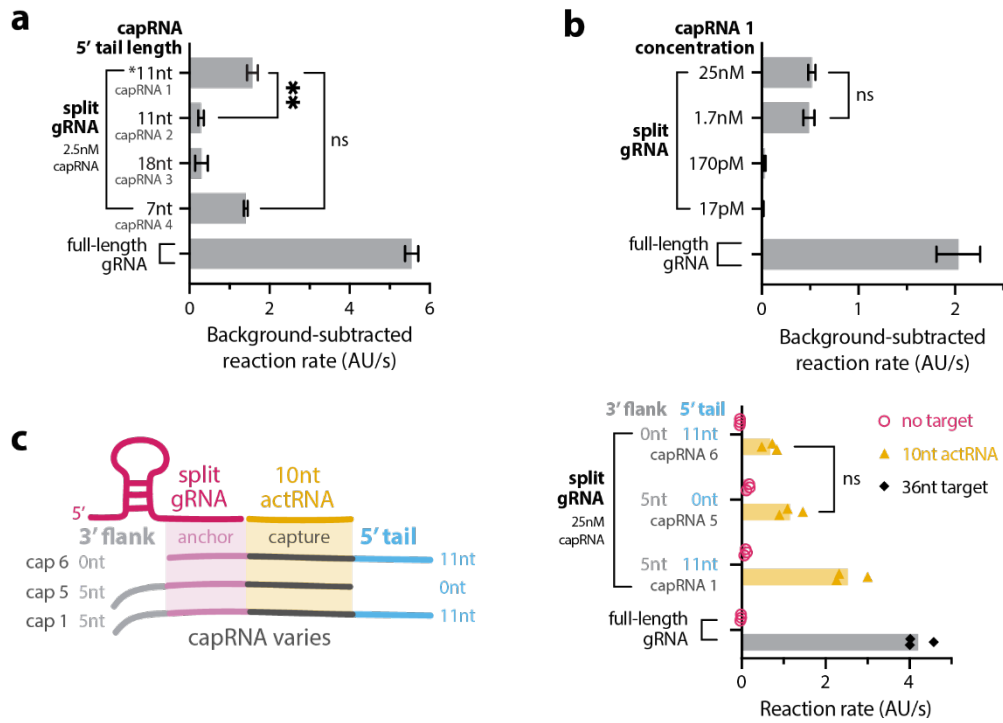
The split guide system we present has been successful in detecting short RNAs with Cas13a and could be applied to other Cas enzymes. For example, split guides could be used with Cas12a to detect DNA targets shorter than the seed region. Alternately, a split guide could be used with Cas12a²⁴³ to detect

miRNAs and induce nonspecific cleavage of both ssRNA and DNA. For short RNAs and beyond, splitting the guide RNA has the potential to open new directions for detection.



Extended Data Figure 1. Supporting data for Fig 1.

A, Fluorescence intensity time course data for full-length gRNA and 36nt target RNA at two-fold dilutions ranging from 5e6 copies/uL to 1.6e5 copies/uL in the final reaction. Each concentration of target RNA has three technical replicates; each replicate is plotted separately. Exponential or linear regression is fit to the data. The slope of the linear region is calculated from the fit and plotted in Fig. 1b. The gRNA is equimolar to Cas13a at 25nM. B, Scatter plot showing reaction rate (AU/s) for Cas13a RNPs activated by target RNA of different lengths and concentrations. Activation of Cas13a RNP is tested with 10nt target A or B at 1e8 copies/uL or both at 5e7 copies/uL each, for a total of 1e8 copies/uL of total RNA (A, yellow, complementary to the first 10nt of the gRNA seed region; B, pink, complementary to the final 10nt of the seed region). Activity with 10nt targets is compared to known activators: 20nt target RNA (green bar, squares, 1e7 copies/uL) and 36nt target RNA (grey bar, circles, 1e6 copies/uL). The reaction rate is calculated as the slope of the linear region of the reaction progress curve (AU vs. time). Cas13a and gRNA are equimolar at 25nM. C, The reaction rates (AU/s) of different combinations of the 'split guide' system are compared to the full-length gRNA and 36nt target RNA. Target RNA is at 1e6 copies/uL in the final reaction. Both parts of the split gRNA are at 25nM in the reaction, equimolar to the Cas13a. The full-length gRNA is also at 25nM. The mean reaction rate (AU/s) with three technical replicates is plotted. D, Supporting data for Fig. 1d. Reaction rate (AU/s) for split gRNA or full-length gRNA system with the mean background activity subtracted (Cas13a RNP and capRNA alone for split gRNA system, or Cas13a RNP alone for the full-length gRNA system). Cas13a, split gRNA, and capRNA 1 are equimolar at 25nM. The 10nt actRNA at 1e6 copies/uL. For the full-length gRNA system, gRNA is equimolar to Cas13a at 25nM and the 36nt target is at 1e6 copies/uL. The mean reaction rate (AU/s) with mean background activity subtracted is plotted with SEM (n=3). Stats: For B and C we used pairwise analysis of the reaction rates between experimental conditions and negative controls using a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for multiple comparisons. We report multiplicity-adjusted p-values. Non-significant interactions ($p > 0.05$), * is $p \leq 0.05$, ** is $p \leq 0.01$.

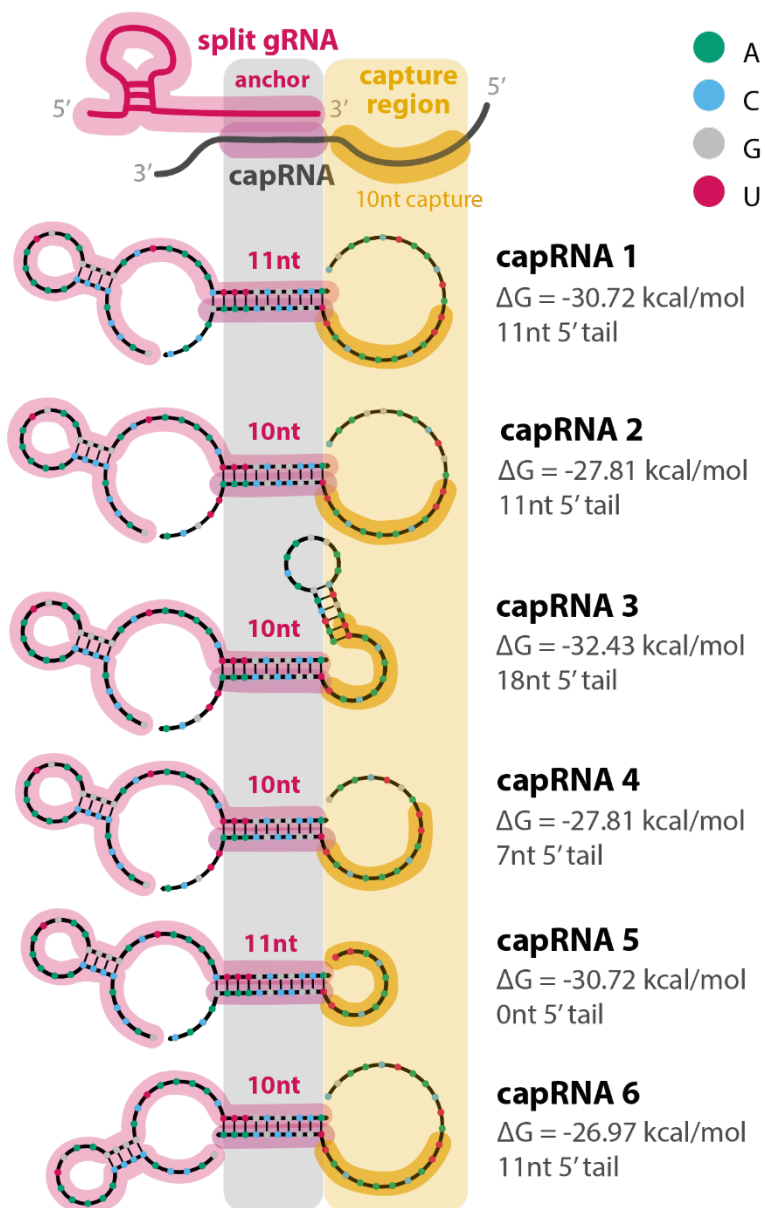


Extended Data Figure 2. Supporting data for Fig 2.

A, Supporting data for Fig. 2b. Reaction rate (AU/s) for split guide or full-length guide system with the mean background activity subtracted. Cas13a and split gRNA are equimolar at 25nM. CapRNA is at 2.5nM, but the capRNA species varies. See Fig 2b for details on capRNAs tested. The 10nt actRNA at 1e6 copies/uL. For the full-length guide system, the gRNA is equimolar to Cas13a at 25nM and the 36nt target is at 1e6 copies/uL. Mean reaction rate (AU/s) with mean background activity subtracted is plotted with SEM (n=3).

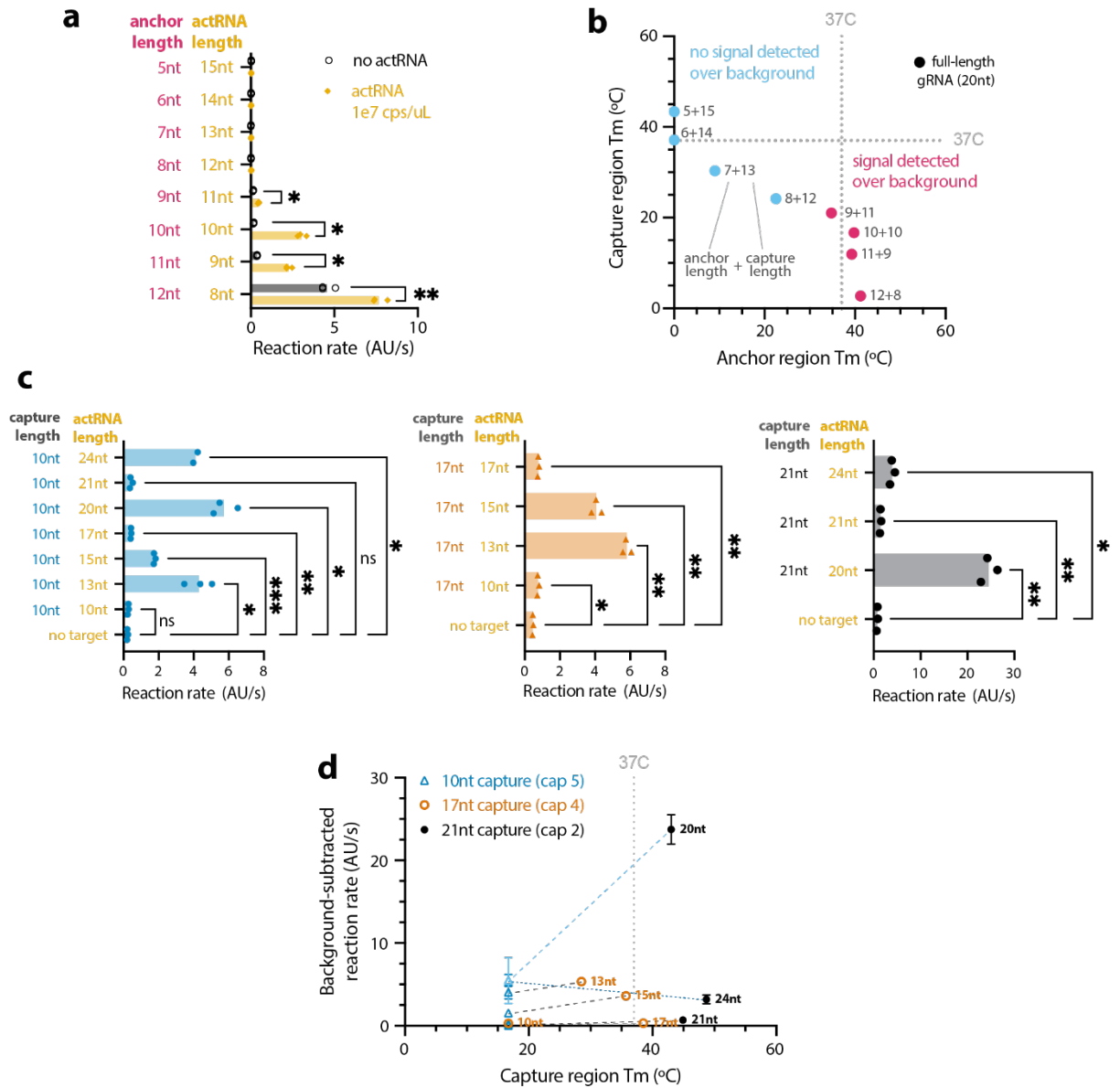
B, Supporting data for Fig. 2c. Reaction rate (AU/s) for split guide or full-length guide system with the mean background activity subtracted. Cas13a and the split gRNA are equimolar at 25nM, while the concentration of capRNA 1 varies from 25nM to 17pM. The 10nt actRNA at 1e6 copies/uL. For the full-length guide system, the gRNA is equimolar to Cas13a at 25nM and the 36nt target is at 1e6 copies/uL. The mean reaction rate (AU/s) with mean background activity subtracted is plotted with SEM (n=3).

D, Left: Schematic illustrating the split guide system with split gRNA (10nt anchor) and 10nt actRNA bound to different capRNAs. The sequence of the anchor and capture regions are held constant, but the presence of the 3' flank or 5' tail varies. Right: For each capRNA species, the reaction was tested with actRNA (yellow bars, triangles) or without actRNA (grey bars, empty pink circles). Cas13a, split gRNA, and capRNAs are present at 25nM. The reaction rate (AU/s) is plotted with and without 10nt actRNA at 1e6 copies/uL. For the full-length guide system, the gRNA is equimolar to Cas13a at 25nM and the 36nt target is at 1e6 copies/uL. The mean reaction rate (AU/s) with three technical replicates is plotted. Stats: For B, C, and D, we performed a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for follow-up testing of key experimental conditions. We report multiplicity-adjusted p-values. Non-significant interactions are $p > 0.05$. * is $p \leq 0.05$, ** is $p \leq 0.01$.



Extended Data Figure 3. Predicted structures for split gRNA-capRNA complexes.

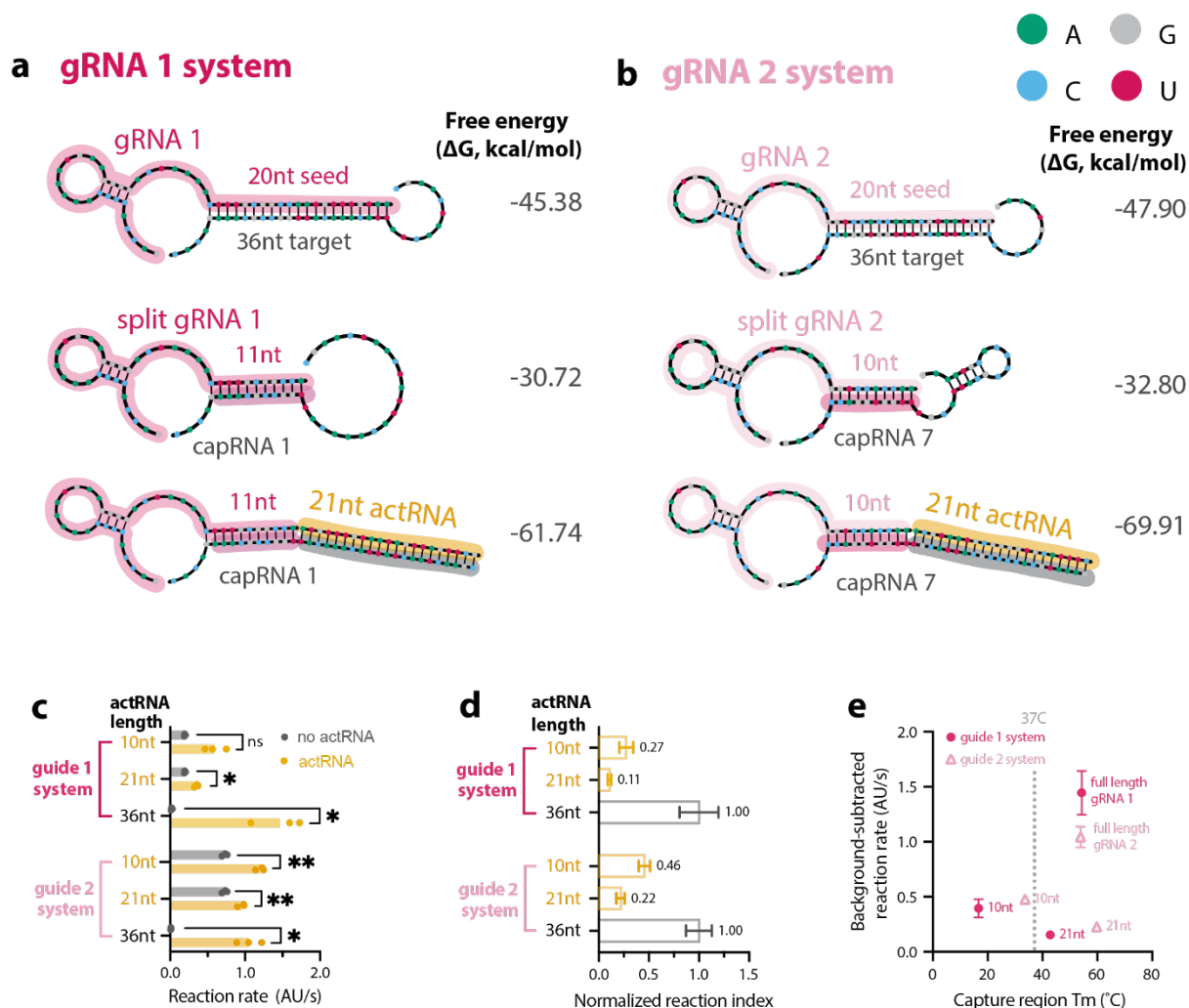
Predicted complex structure with same split gRNA (10nt anchor region) and varying capRNAs. The split gRNA is highlighted in light pink. On the capRNA, the anchor region is highlighted in pink and the capture region is highlighted in yellow. For capRNA 1 and 9, the sequence of the 3' flanking region is such that there is an additional 1nt of complementary base pairing with the split gRNA, making the anchor region effectively 11nt. The complex structures and free energy are predicted with the NUPACK web application.



Extended Data Figure 4. Supporting data for Fig 3.

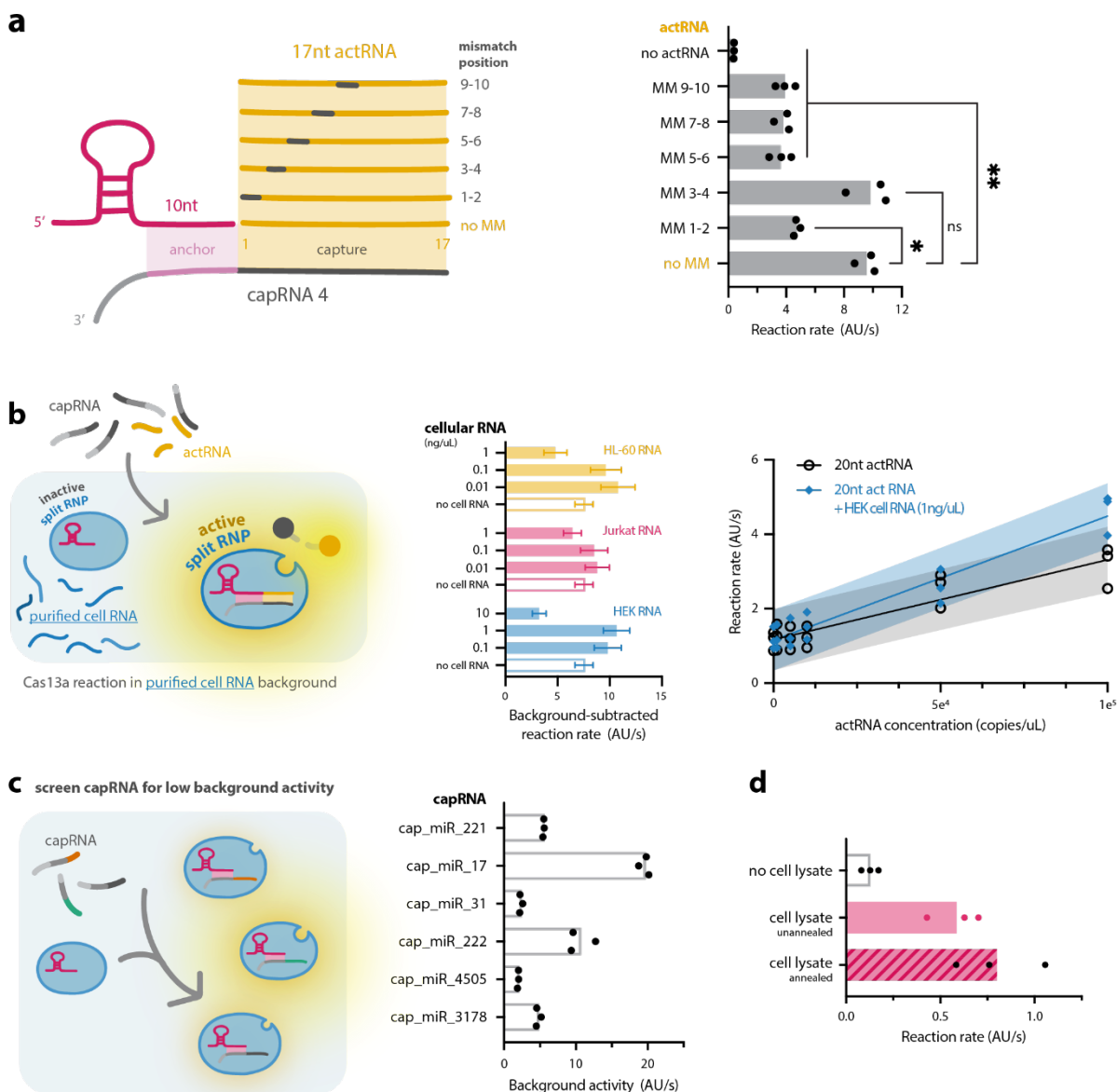
A, Supporting data for Fig. 3a. For each split gRNA species, the reaction was tested with the corresponding actRNA (yellow bars, rhombuses) or without actRNA (grey bars, empty circles). Cas13a, split gRNAs, and capRNA 1 are equimolar at 25nM. All actRNA is at 1e7 copies/uL in the final reaction. The mean reaction rate (AU/s) with three technical replicates is plotted. B, The melting temperature (°C) of the capture region versus anchor region is plotted for each split gRNA and actRNA pair. For a given pair, if the signal was detected over background (see A), the dot is colored pink. If not, the dot is colored blue. Grey dotted lines at 37°C mark the reaction temperature. C, Supporting data for Fig 3b. Left: Reaction rate (AU/s) of actRNA of varying lengths paired with capRNA 5 (10nt capture region). Middle: Reaction rate (AU/s) of a subset of actRNA paired with capRNA 4 (17nt capture region). Right: Reaction rate (AU/s) of a subset of actRNA paired with capRNA 2 (21nt capture region). Cas13a is equimolar to the split gRNA (10nt anchor) at 25nM. All capRNA is at 2.5nM, and all actRNA is at 1e6 copies/uL in the final reaction. The mean reaction rate (AU/s) with three technical replicates is plotted. All reactions are compared to the control with no actRNA. D, For each capRNA-actRNA combo, the background-subtracted reaction rate (see Fig. 3b) is plotted versus the melting temperature of the capture region (°C). Data for the 10nt capture region (capRNA 5) is shown with blue triangles, 17nt capture region (capRNA 4) in orange circle outlines, and the 21nt capture region (capRNA 2) in solid black circles. Dashed lines connect points for the same actRNA paired with different capRNAs to emphasize the change in signal. ActRNA length is labelled next to the data points. A grey dotted line marks the reaction temperature (37°C). Stats: For A and C, we performed a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for follow-

up testing of experimental conditions versus their appropriate no target control. We report multiplicity-adjusted p-values. Non-significant interactions are $p > 0.05$. * is $p \leq 0.05$, ** is $p \leq 0.01$, *** is $p \leq 0.001$.



Extended Data Figure 5. Comparison of two sequence-unique split guide systems.

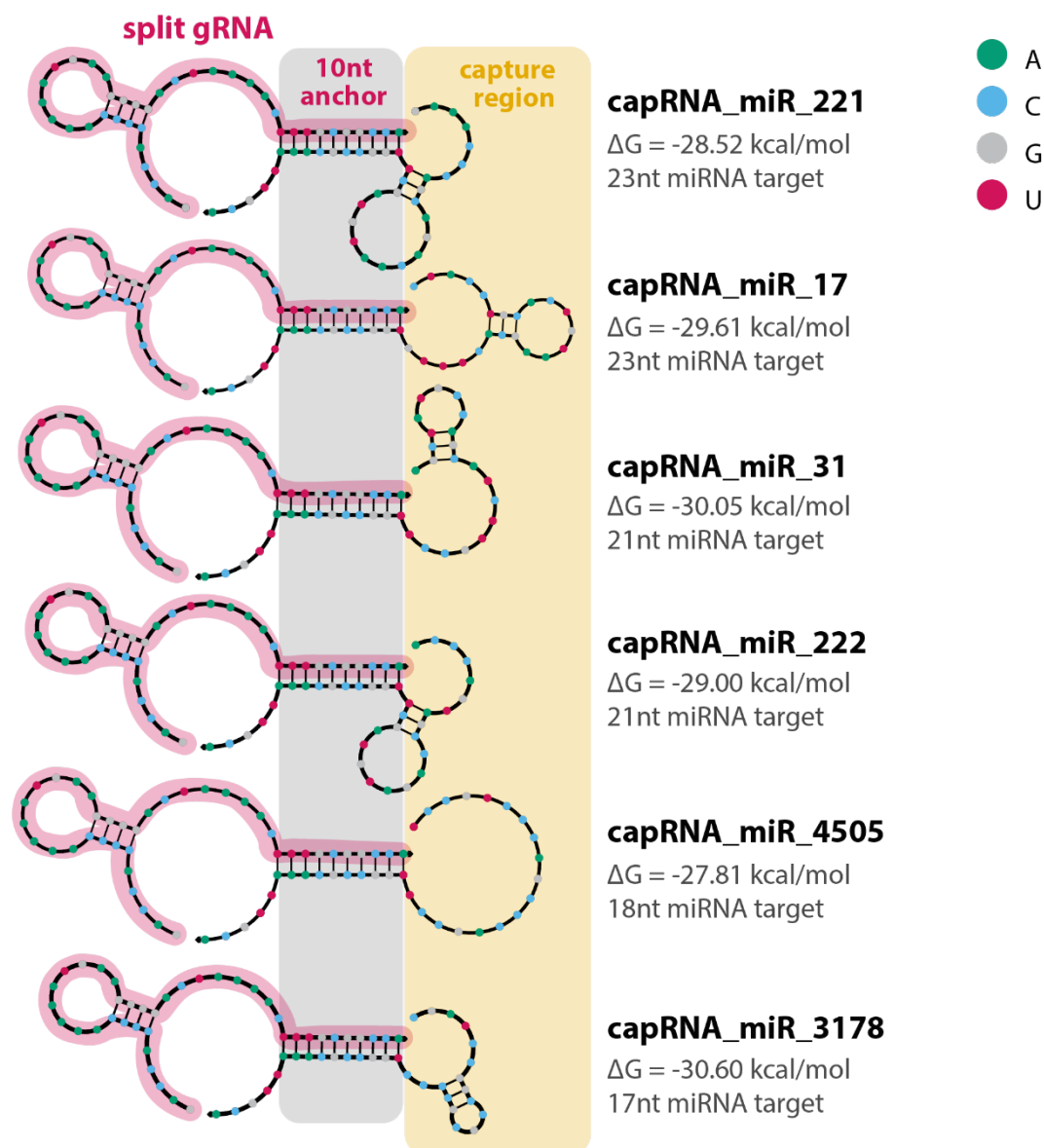
A and B, Predicted complex structure for gRNA system 1 (A) and 2 (B). Top row is full-length gRNA bound to a 36nt target. Middle row is split gRNA bound to capRNA. Bottom row is split gRNA and 21nt actRNA bound to capRNA. The complex stability is reported as free energy (ΔG , kcal/mol). Complex structures and free energy are predicted with the NUPACK web application. C, Supporting data for Fig. 3c. Reaction rates (AU/s) are plotted for gRNA system 1 and 2, either the split gRNA paired with a 21nt or 10nt actRNA or the full-length gRNA paired with a 36nt target. All conditions are compared to their appropriate no actRNA control. For the split guide systems, Cas13a is equimolar to the split gRNA at 25nM, the capRNA is at 2.5nM, and the actRNA is at 1e6 copies/uL. For the full-length guide systems, Cas13a is equimolar to the gRNA at 25nM, and the 36nt target RNA is at 1e6 copies/uL. The mean reaction rate (AU/s) is plotted with SEM (n=3). D, Supporting data for Fig. 3c. The background-subtracted reaction rates are normalized by dividing by the background-subtracted reaction rate of the full-length gRNA with a 36nt target RNA. This normalized reaction index is plotted with SEM (n=3). Normalized values are reported to the right of the bars. E, For each capRNA-actRNA combo, the mean background-subtracted reaction rate and SEM (n=3, see Fig. 3c) is plotted versus the predicted melting temperature of the capture region ($^{\circ}\text{C}$). Data for gRNA 1 system is plotted with hot pink circles, while data for gRNA 2 system is plotted with light pink outlined triangles. The actRNA length is labelled next to the data points. A grey dotted line marks the reaction temperature (37 $^{\circ}\text{C}$). Stats: For C, we performed a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnett's T3 test for follow-up testing of experimental conditions versus their appropriate no actRNA control. We report multiplicity-adjusted p-values. Non-significant interactions are $p > 0.05$. * is $p \leq 0.05$, ** is $p \leq 0.01$.



Extended Data Figure 6. Supporting data for Figs. 4 and 5.

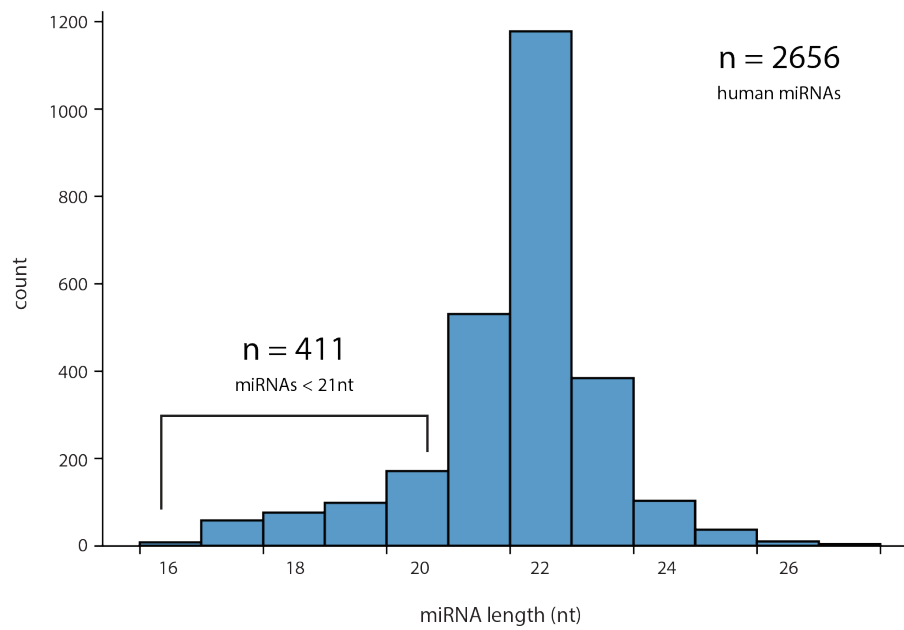
A, Supporting data for Fig. 4a. Left: Schematic illustrating further testing of mismatch sensitivity. The split gRNA (hot pink, 10nt anchor) and 17nt actRNA (yellow) containing a 2nt mismatch tiled across the first 10nt of the actRNA are bound to capRNA 4. CapRNA 4 contains a 10nt anchor region and a 17nt capture region, such that the entire actRNA binds to the capRNA. Right: Reaction rates (AU/s) for all mismatch actRNAs paired with capRNA 4. Cas13a is equimolar to the split gRNA at 25nM, capRNA 4 is at 2.5nM, and all actRNA is at $1e7$ copies/uL. The mean reaction rate (AU/s) is plotted with three technical replicates. All reactions are compared to the actRNA with no mismatches. B, Supporting data for Fig. 5a. Left: Schematic illustrating the detection of synthetic actRNA in the presence of purified cellular RNA. The purified cellular RNA acts as a general RNA background to which the actRNA of interest is spiked in. Any Cas13a RNP activated by the actRNA will cleave both the fluorescent reporter and the cellular RNA. Middle: Test of synthetic 17nt actRNA detection in total cellular RNA background. Cellular RNA, extracted from HL-60 (yellow bars), Jurkat (pink bars), or Lenti-X HEK 293T cells (blue bars) is added at three different concentrations in the final reaction. Cas13a is equimolar to the split gRNA at 25nM, capRNA 4 (17nt capture region) is at 2.5nM, and the 17nt actRNA is at $1e7$ copies/uL. For each concentration and species of cellular RNA, the mean background-subtracted reaction rate (AU/s) is plotted with SEM ($n=3$). Right: Limit of detection analysis of 20nt actRNA in a cellular RNA background. Cellular RNA, extracted from Lenti-X HEK 293T cells, is at 1ng/uL in the final reaction. Cas13a is equimolar to the split gRNA (10nt anchor) at 25nM, capRNA 2 (21nt capture region) is at 2.5nM, and the 20nt actRNA concentration varies. For each concentration of actRNA, the mean reaction rate (AU/s) is plotted for three technical replicates. A linear regression is fit to the reaction rate vs. actRNA concentration data and plotted with 95% prediction bands. The slope of the linear regression is compared for the two conditions with an unpaired t-test; $p = 0.0013$. C, For miRNA detection, screening capRNA

for low background signal is critical. Left: Schematic illustrating addition of capRNA to Cas13a RNPs and resultant fluorescent signal generated from RNPs activated by capRNA in the absence of the activating miRNA. CapRNAs tested contain the same 3' flanking sequence (5nt) and 10nt anchor region. The capture region is unique to each capRNA and perfectly complementary to the miRNA of interest, with no 5' tail. Right: CapRNA for six miRNA targets are tested for background activity. The mean reaction rates (AU/s) of Cas13a RNPs plus capRNA are plotted with three technical replicates. Cas13a is equimolar to the split gRNA (10nt anchor) at 25nM. All capRNA is present at 2.5nM. The three capRNAs with the lowest activity were selected for further testing. D, Supporting data for Fig. 5b. Test of background activity contributed by cell lysate. CapRNA 1 was selected as an off-target capRNA that shouldn't bind any miRNAs in the human genome. Background activity was measured with Cas13a RNP plus capRNA 1 in a no target control (empty grey bar), cell lysate (pink bar), and with capRNA 1 annealed with cell lysate (striped pink bar). Cas13a is equimolar to the split gRNA at 25nM and capRNA 1 is at 2.5nM; unpurified cell lysate from Lenti-X HEK 293T cells is added at a final concentration of 10 cells/uL. The mean reaction rate (AU/s) is plotted with three technical replicates. Stats: For A, we performed a Brown-Forsythe ANOVA to account for variation in standard deviation and Dunnet's T3 test for follow-up testing of experimental conditions versus the no mismatch control. We report multiplicity-adjusted p-values. Non-significant interactions are $p > 0.05$. * is $p \leq 0.05$, ** is $p \leq 0.01$.



Extended Data Figure 7. Predicted structures for split gRNA-capRNA complexes for miRNA targets.

Supporting data for Fig. 5a, 5b, and Extended Data Fig. 6c. Predicted complex structure for split gRNA bound to capRNA targeting six different cellular miRNAs. The split gRNA is highlighted in pink, with the 10nt anchor region highlighted with a grey box. The capture region (length varies) is highlighted with a yellow box. The stability of the complex is reported as free energy (ΔG , kcal/mol). Complex structures and free energy are predicted with the NUPACK web application.



Extended Data Figure 8. Length distribution of validated human miRNAs.

Histogram of lengths of all human-annotated miRNAs in miRbase release 22.1.

Methods and Materials

LbuCas13a fluorescent plate reader assay

For all reactions, Cas13a RNPs were formed by incubating 133nM LbuCas13a (Shanghai ChemPartner) with 133nM gRNA or split gRNA in 1x cleavage buffer (20mM HEPES-Na, pH 7.2, 50mM KCl, and 5% glycerol). These complexes were incubated for 15-30min, then aliquoted into a 384-well plate (flat-bottom, black, polystyrene with non-binding surface treatment; Corning #3820). For full-length guide experiments, target RNA or DEPC-treated water was added to wells. For split guide experiments, actRNA plus capRNA, or capRNA was added to wells.

For experiments with an annealing step, capRNA and target solution (act RNA, total cellular RNA, whole cell lysate, or water) were mixed at 10x their final concentration in 1X annealing buffer (10 mM Tris pH 7.5 - 8.0, 50 mM NaCl, 1 mM EDTA) and annealed by heating at 95°C for 2 minutes and cooling to 4°C at a rate of 1°C/min. The pre-annealed capRNA and target solution was then added to the RNP mixture.

After adding the target solution to the RNP, the mixture was incubated for 2-3min before adding the reaction master mix and pipette-mixing the full reaction. The master mix, composed of 1x cleavage buffer, MgCl₂, and fluorescent cleavage reporter, was preheated to 37°C before adding to the wells. In the final 20uL reaction, LbuCas13a and gRNA were at 25nM in the presence of 1x cleavage buffer (20mM HEPES-Na, pH 7.2, 50mM KCl, and 5% glycerol), 5mM MgCl₂, and 400nM of quenched RNA reporter (5'-FAM-rUrUrUrUrU-IowaBlack FQ-3'). For experiments involving total cellular RNA or whole cell lysate, murine RNase inhibitor (New England Biolabs) was added at a final concentration of 0.1U/uL. For each experiment, the final concentration of target RNA or capRNA and actRNA is listed in the figure caption.

After adding the reaction master mix, the 384-well plate was covered with an optically-clear seal (Thermo Scientific #AB-1170) and immediately inserted into a humidity cassette in the TECAN Spark multimode microplate reader. The plate reader was pre-warmed to 37°C before starting the experiment. Fluorescence was measured every 2.5 minutes for 75-90 minutes at a z-position of 20046 μ m and gain of 70. The excitation wavelength was 485 nm (5 nm bandwidth), and the emission wavelength filter was 535 nm (25 nm bandwidth). We performed three technical replicates for each experimental condition.

Analysis of Cas13a reactions slopes

To compare the reaction rates of the different conditions, we first subtracted the fluorescence value at time $t=0$ for each well. We then calculated the slope of the linear region by performing either an exponential or a simple linear regression. For the exponential regression, which models the data with Michaelis-Menten kinetics, we calculate the equivalent slope of the linear region based on the exponential fit. For reaction progress curves that did not plateau in the timecourse of the experiment and therefore did not yield a good exponential fit, we fit a simple linear regression to the linear portion of the curve.

Oligonucleotides

All gRNA, split gRNA, and capRNA were ordered from Synthego as 5nmol of desalted single-stranded oligonucleotides. Target and activating RNA ≥ 20 nt were also ordered from Synthego. Activating RNA < 20 nt was ordered from IDT as 100nmol or 250nmol of desalted single-stranded

oligonucleotides. The quenched RNA reporter was also ordered from IDT (5'-FAM-rUrUrUrUrU-IowaBlack FQ-3').

Total RNA extraction

Total cellular RNA was prepared using a Monarch Total RNA Miniprep Kit (New England Biolabs). RNA was extracted from 1e6 Lenti-X HEK 293T cells according to the manufacturer's protocol and eluted in 100uL DEPC-water. RNA was not enriched for small species (<200nt). Sample was flash-frozen and stored at -80°C.

Whole cell lysis

Whole cell lysate was prepared with 1e6 Lenti-X HEK 293T cells. Cells were washed twice with PBS before resuspending in 100uL lysis buffer (1% Triton X-100 in 1x cleavage buffer). Sample was incubated at 90°C for 5 minutes to lyse cells and inactivate cellular RNAses. Murine RNase Inhibitor (New England Biolabs) was then added to the cell lysate solution at 0.4U/uL. Sample was flash-frozen and stored at -80°C.

Mammalian cell lines and culture conditions

Lenti-X HEK 293T cells were acquired from the UC Berkeley Cell Culture Facility. Cells were maintained at 37°C, 5% CO2 in DMEM (Corning) supplemented with 10% fetal bovine serum (Gibco) and 100U/mL penicillin-streptomycin (Gibco).

Statistical Analysis

Prism v10 was used to perform statistical analysis. Specific statistical details for each figure are included in the figure caption. In general, pairwise analysis of Cas13a reaction rates was performed between experimental conditions and their appropriate negative control (no activating RNA). To account for variation in standard deviation between conditions, a Brown-Forsythe ANOVA was used. Dunnet's T3 test was used to adjust the p-value for multiple comparisons. Multiplicity-adjusted p-values are used in all figures, unless otherwise stated. * is $p \leq 0.05$, ** is $p \leq 0.01$, *** is $p \leq 0.001$.

Limit of Detection (LOD) calculation

To estimate the LOD for each target RNA, we used the standard method proposed by the Clinical and Laboratory Standards Institute in their guideline EP17⁴⁴:

$$\begin{aligned} \text{Limit of Blank (LoB)} &= \text{mean}_{\text{blank}} + 1.645 \times SD_{\text{blank}} \\ \text{Limit of Detection (LoD)} &= \text{LoB} + 1.645 \times SD_{\text{low concentration}} \end{aligned}$$

For the LoB, we used the mean slope and standard deviation of the RNP only condition for full-length guide experiments or the RNP plus capRNA from the split guide experiments. For the LoD, we used the standard deviation of the lowest target RNA concentration tested. This strategy provides a LoD such that 95% of samples tested at the estimated LOD concentration are at or above the LOD.

Acknowledgements

We would like to thank the members of the Fletcher Laboratory for providing helpful discussions and invaluable feedback on this project, especially Carlos Ng Pitti and Andres Dextre for assay input, Deepak Krishnamurthy for data analysis, and Daniel S.W. Lee for statistical analysis, and María Díaz de León Derby and Caitlin Cornell for moral support. We thank Sungmin Son for contributions to the split guide idea, and we thank Melanie Ott for many helpful interactions. We gratefully acknowledge support from Google, the ANCeR diagnostics consortium, the Pernendu Chatterjee Chair fund, and the Chan Zuckerberg Biohub. D.A.F. is a Chan Zuckerberg Biohub Investigator.

Author contributions

A.I.M., A.L., and D.A.F. conceived and designed the study. A.I.M. and J.P.C. performed experiments and A.I.M. analyzed and interpreted data. D.A.F. supervised the study design and data collection. D.A.F. secured funding. A.I.M. wrote the manuscript with extensive input from D.A.F. and feedback from J.P.C. and A.L.

Declarations of interests

A.I.M., A.L., S.S., and D.A.F. have filed a patent application related to this work.

Chapter 3 – Designing a Cas13a split guide assay

This chapter is a reference guide to aid in the design of new Cas13a split guide assays. I summarize my current understanding of how the Cas13a split guide system works and hypotheses for what constitutes a successful split guide triplet – split gRNA, activating RNA, and capture RNA (Fig. 1). I discuss design principles for the anchor region (split gRNA-capRNA interaction), capture region (actRNA-capRNA interaction), and the capture RNA. Then, I summarize these design rules and recommend next experiments to validate and extend our understanding of the split guide system.

Preliminary rules for designing anchor-capRNA-miRNA triplets

When designing a Cas13a split gRNA triplet for a specific RNA target, there are three key areas to consider: 1) the anchor region, 2) the capture region, and 3) the capRNA (Fig. 1).

While this design space may seem daunting, it will become increasingly smaller and manageable as we learn more about the system. The anchor region, for example, won't need to be modified much once an optimal sequence has been designed and tested. The capture region is mostly dictated by the sequence and length of the activating target, with some room for modifications. The biggest design challenge is the capRNA. Although the anchor and capture regions are mostly defined, the 3' flank and 5' tail must be considered, and all together the effect on secondary and tertiary structure.

I've broken these three design areas into sections and provided my current understanding of the rules guiding the design of each part, supported with data. I've also highlighted key questions to explore to help further our understanding of the Cas13a split guide system.

Anchor Region – Design Principles

The anchor region is dictated by the split gRNA. It is composed of a portion of the seed region from a traditional full-length gRNA. The anchor region can vary in both length and sequence.

In my work, I have probed an anchor ranging from 5nt to 12nt, and I found that an anchor as small as 9nt is sufficient to lead to positive signal detected over background (Fig. 2a and b). That is not to say that any 9nt sequence is sufficient for the anchor region, nor that no 8nt sequence could ever function as the anchor region. Rather, it is more likely that the specific sequence of the anchor region in combination with the length informs the stability of the interaction between the split gRNA and the capRNA.

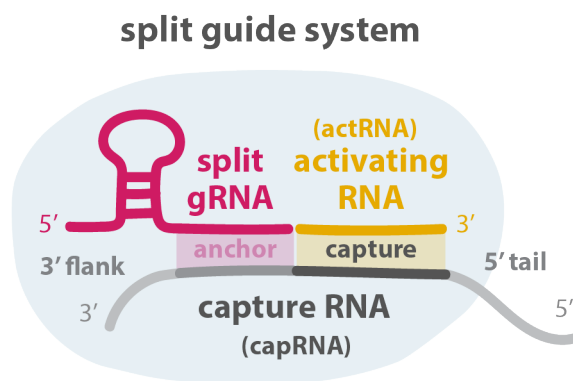


Figure 1. Cas13a split guide system.

Schematic illustrating the RNA components of the Cas13a split guide system: split gRNA (hot pink), activating RNA (yellow), and capture RNA (grey).t

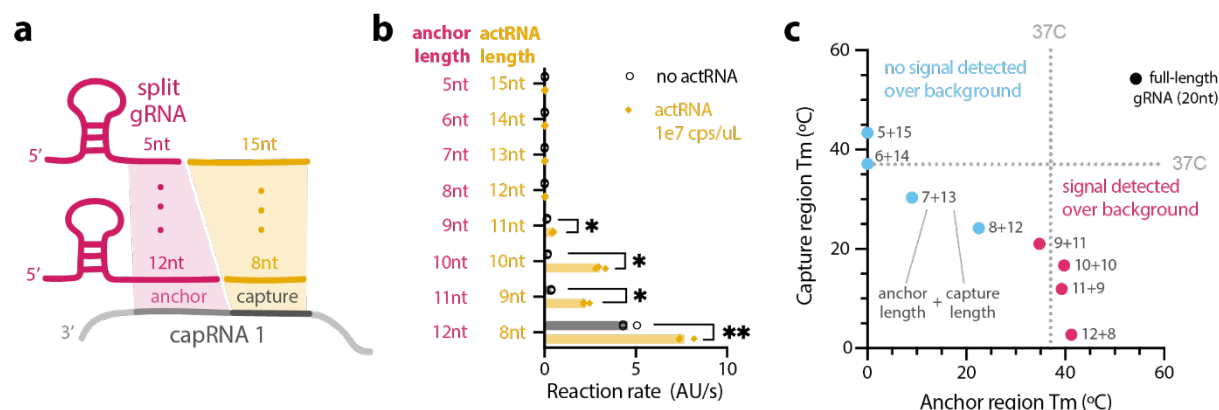


Figure 2. Test of split gRNA with varied anchor region length.

Subfigures are reproduced from Ch. 2, Fig. 3a, Ch. 2, Extended Data Fig. 4a and b, respectively. A, Anchor region of split gRNA varies from 5 to 12nt, paired with an actRNA that maintains a combined length of 20nt. B, Reaction rate of the different combinations with (yellow bars) and without (grey bars) actRNA. C, Melting temperature of the anchor and capture regions for each split gRNA and actRNA combination tested. Points are color-coded to indicate whether a positive signal was detected over background.

The melting temperature of the anchor region duplex can be a useful proxy for predicting the stability of the split gRNA-capRNA interaction. The melting temperature of an oligonucleotide duplex is defined as the temperature at which 50% of oligos are double stranded in a given environment (namely salt concentration). We normally run our Cas13a reactions at 37°C, so we would expect to see good activity when the anchor region has a melting temperature above 37°C. But, even if the melting temperature of a duplex is below the temperature of the reaction, we still expect some portion of the split gRNA-capRNA to be in the duplex form, resulting in some RNP activation. In Fig. 2c, we see that the RNPs with an anchor region melting temperature around or above 37°C result in positive signal detected over the background (see also Table 1).

Table 1. Melting temperature (T_m) analysis for split gRNA anchor regions of varied lengths.

Anchor length (nt)	region	GC content (%)	Melting Temperature (*C)	Signal detected over background?
5		40	0	No
6		50	0	No
7		57.1	9	No
8		62.5	22.5	No
9		66.7	34.8	Yes
10		60	39.8	Yes
11		54.5	39.3	Yes
12		50	41.2	Yes

Note: T_m calculated using IDT oligoanalyzer tool. Settings reflect experimental conditions. Na concentration was 20mM and Mg concentration was 5mM. Oligo concentration was set to the assumption that the capRNA was equimolar to the split gRNA, which is the case in the experiment graphed above.

While we normally run our reactions at 37°C, we have found that the reaction can be run at lower temperatures without a significant decrease in reaction rate (data not shown). Thus, it would be

interesting to test whether lowering the reaction temperature could enable an 8nt anchor region to work.

If the anchor region isn't stable enough, the binding of the capRNA with the actRNA is irrelevant. This is likely why we see in Fig. 2c that RNPs with a high capture region melting temperature do not result in a detectable signal over background: the interaction between the split gRNA and the capRNA via the anchor region is not strong enough to stabilize the binding.

It is important to remember, however, that melting temperature is just a proxy for the stability of the anchor region; Cas13a likely plays an additional role in stabilizing the binding of the capRNA to the split gRNA. If melting temperature was the only predictor for a stable interaction, then choosing an RNA target would be as simple as looking for unique sequences with high melting temperatures. Clearly that is not the case, as many groups are trying to unveil design principles for Cas13a gRNAs.

Another important aspect to consider when designing the anchor region is whether the interaction between the split gRNA and capRNA activates the Cas13a RNP in the absence of the actRNA. A high background activity may reduce the utility of a split gRNA-capRNA pair if the background overwhelms the positive signal. We see this phenomenon happen with the 12nt anchor region. When the background signal is subtracted from the positive signal, the difference is comparable to the performance of the 11nt and 10nt anchor regions (Fig. 3). Thus, a longer anchor region is not better if it results in a high level of background activity.

As a starting point for further optimization of the anchor region, I recommend testing other sequences and validating the performance of different anchor regions paired with the same activating RNA.

All of my testing of the Cas13a split guide system has been with two sets of sequences: guide 1 system (based on crRNA 4²⁷) and guide 2 system (based on crRNA 2²⁸). In both guide 1 and 2, I tested a 10nt anchor region paired with either a 10 or 21nt actRNA (Fig. 4a). Compared to the guide 1 system, the guide 2 system generated a much higher relative signal for both the 10 and 21nt actRNAs (Fig. 4b). The anchor regions for both systems have a GC content of 60% and a comparable melting temperature (39.8°C for split gRNA 1 and 37.1°C for split gRNA 2). This may suggest that the difference in performance of the two systems depends on details other than the anchor region, such as the capture region. The melting temperatures of the two sets of capture regions did vary between the two guide systems, with the capture region for the guide 2 system having a higher melting temperature than the guide 1 system for both the 10 and 21nt capture (Fig. 4c). This difference in melting temperature may explain the difference in performance between the systems. However, to better differentiate the role of the anchor region from that of the capture region, it is best to test the two anchor regions paired with the same length and sequence capture region.

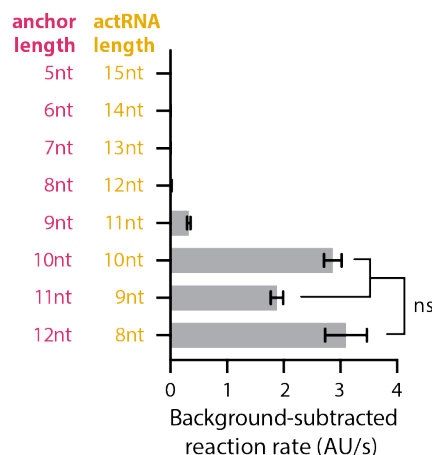


Figure 3. Effect of varied split location on system activation. Background-subtracted reaction rates for different combinations of anchor region and actRNA lengths. Reproduced from Ch. 2, Fig. 3a.

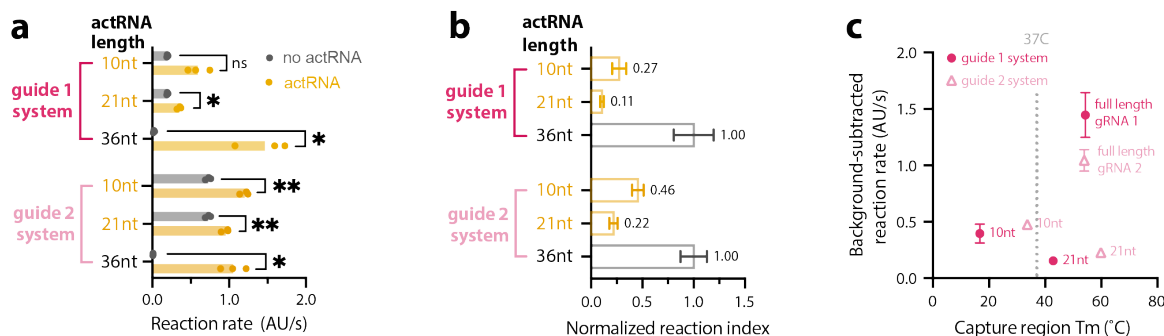


Figure 4. Test of split guide system in two unique sequence spaces.

Subfigures are reproduced from Ch. 2, Extended Data Fig. 5c-e, respectively. A, Reaction rates for two guide sets, comparing signal from the full-length gRNA and a 36nt target to the of the split gRNA paired with either a 10nt or 21nt actRNA. B, Data from A, background subtracted and normalized to the 36nt target control. C, Background-subtracted reaction rates from A plotted with corresponding melting temperature of the capture region.

I am even more curious to test new anchor region sequences based off successful full-length gRNAs. There are other gRNAs that we have recently tested in lab that yield faster reaction rate than either crRNA 2 or 4. I would be interested to know how well those gRNAs perform as split gRNAs with anchor regions ranging from 9-11nt.

Comparing the performance of different anchor region lengths and sequences paired with the same actRNA will get expensive. For every unique anchor region, a new split gRNA and capRNA would need to be ordered. Fortunately, it may be that using a “capture DNA” (capDNA) in lieu of the capRNA is sufficient for testing purposes. A recent preprint demonstrated that single-stranded DNA can activate a Cas13a RNP⁴⁵. This finding suggests that DNA could work as the capture oligonucleotide in the Cas13a split guide system. Ordering DNA instead of RNA for the capture strand is much cheaper and faster, and it could enable a rapid design-test process. Furthermore, unbound capDNA would not be cleaved by an activated RNP, potentially boosting the sensitivity of the assay.

Finally, one last strategy to improve the stability of the anchor region is to incorporate locked nucleic acids (LNAs). LNAs are modified RNA nucleotides that contain a methylene bond bridging the 2' oxygen and 4' carbon of the pentose ring⁴⁶. This modification increases the oligonucleotides melting temperature, which can lead to an enhancement in single nucleotide polymorphism (SNP) and mismatch detection. LNAs in the anchor region, either on the split gRNA or capRNA could increase the stability of the interaction for a given anchor length. The use of LNAs might shift the optimization equation when balancing positive signal over background. LNAs may make shorter anchor regions more viable options, capitalizing on the low background activity, but enhancing binding stability to increase positive signal when a target RNA is bound.

Once an anchor region has been designed that optimizes positive signal over background, no further modifications need be made. However, it would be nice to have a few good options for the anchor region for different applications. If we ever want to have two unique Cas13a RNP species in one well, which could be a strategy for basic multiplexing, it will be necessary to have two split gRNAs with orthogonal anchor regions.

Capture Region

For the most part, both the length and sequence of the capture region is defined by the RNA target. I have demonstrated that RNA targets ranging from 8 to 24nt can be detected with the split guide system, though it is likely that target RNA longer than 24nt can also be detected. My data further suggest that for a given actRNA length, having a higher melting temperature results in a higher signal. This trend was true for two unique 10nt actRNAs and two 21nt actRNAs (see Fig. 4a, c). It is likely that the explanation is more complicated than just melting temperature, but using that as a predictor is a fine place to start when selecting RNA targets and designing a capture region.

In most use cases for the split guide system, there is limited flexibility in choosing the RNA target. A traditional Cas13a assay with a full-length gRNA has lots of flexibility in selecting an RNA sequence to target; for RNA virus detection, 20-28nt sequences tiled along the entire length of the viral genome have the potential to be the RNA target. When detecting short RNA sequences like miRNA, the miRNAs of interest are usually defined by the specific application. So, when given a specific miRNA target, it may be necessary to creatively modify the split guide system.

After selecting an RNA target, it is good to check the melting temperature of the capture region to have a baseline idea of the stability of the capRNA-actRNA duplex. This stability can be improved to some extent by altering the reaction conditions. For a capture region with a low melting temperature, careful annealing of the actRNA and the capRNA and reduction of the assay temperature may help improve detection. Increasing monovalent and divalent salt concentration can also help increase the stability of the duplex, though it may affect the base Cas13 reaction. Additionally, introducing LNAs in the capRNA can help increase the melting temperature of the duplex, and may even assist with specificity against mismatches. Knowing the basic characteristics of the anchor region can help inform modifications to the assay to improve the likelihood of success.

Even beyond target RNAs with low melting temperatures, there are other aspects that make a target RNA difficult to detect. The sequence of the target may bestow undesired secondary and tertiary structure upon the capRNA. This complex structure may result in high background activity, making it difficult to detect a positive signal above background levels. Introducing mismatches in the capture region may help reduce the secondary structure of the capRNA and improve signal detection over background. In the next section, I'll discuss in more detail the potential implications of and more solutions for removing secondary structure in the capRNA.

Once an RNA target has been selected, it is important to consider the purpose of the detection assay. Some applications may value a low limit of detection while others may value high specificity.

To prioritize sensitivity, the capture region should encompass the full length of the RNA target. Binding the full length of the actRNA should result in the highest melting temperature and thus greatest fraction of actRNA-capRNA duplexes in the reaction mixture. Binding the full length also provides an opportunity to modify the 5' tail of the capRNA, which may further increase the positive signal.

To prioritize localized specificity, it may be helpful to use a shorter capture region. In my testing of mismatch detection, I compared two capRNAs with different length capture regions (Fig. 5a). The shorter capture region (10nt) demonstrated high specificity for 2nt mismatches (Fig. 5b); all actRNA with a 2nt mismatch, no matter the location, resulted in signal that was reduced to background. This

is not too surprising, as only 10 of the 17nt in the actRNA bind to the capRNA; making two of those ten nucleotides a mismatch only further destabilizes the interaction. Increasing the length of the capture region to fully complement the 17nt actRNA affected the mismatch discrimination of the system. When 2nt mismatches were present, the positive signal was significantly reduced, but was still greater than background levels (Fig. 5c). Interestingly, when the mismatches were placed at positions 3 and 4 of the actRNA, the signal was no different from the no mismatch condition. This tracks with other work that has shown that mismatches in positions 13-16 in the target RNA (equivalent to positions 3-6 in the actRNA), significantly increases the activity relative to the no mismatch control³⁹. I have not tested specificity of the system for 1nt mismatches, but recommend doing so, as it will enable the use of the Cas13a split guide system for more applications, detailed in Chapter 5. Incorporating LNAs at key positions may help boost specificity for both 1 and 2nt mismatches.

Figure 5. Test of split guide system specificity to mismatches.

Though the design space for the capture region is small, since it is primarily informed by the sequence of the actRNA, there is still some room for manipulation to achieve the desired results for an assay. The length of the capture region can be altered to suit the needs of the assay, prioritizing sensitivity or specificity. The sequence may also be modified to introduce mismatches that reduce unwanted secondary structure or include LNAs to enhance localized stability and augment mismatch discrimination. Further experiments to tease out the key properties of the capture region will help inform future modifications to the assay.

The capture RNA is composed of the anchor and capture regions surrounded by the 3' flank and 5' tail. Once the anchor and capture regions have been selected, the capRNA needs to be carefully characterized and modified to optimize the signal-to-background ratio.

for robust detection. To test this theory in the split guide system, I compared capRNA 1 (5nt 3' flank and an 11nt 5' tail) with capRNA 5 (5nt 3' flank and no 5' tail) and capRNA 6 (no 3' flank and 11nt 5' tail). From this experiment, I saw that the background signal didn't change significantly across the different capRNAs, but that the positive signal did (Fig. 6b). Both capRNA 5 and 6 had a reduced positive signal compared to capRNA 1. There was no significant difference in the positive signal between capRNA 5 and 6. Together, this data suggests that having unbound nucleotides outside of the anchor and capture regions is important for generating a high signal. Perhaps these dangling sequences play a role in the conformational change of the RNP and encourage higher catalytic turnover.

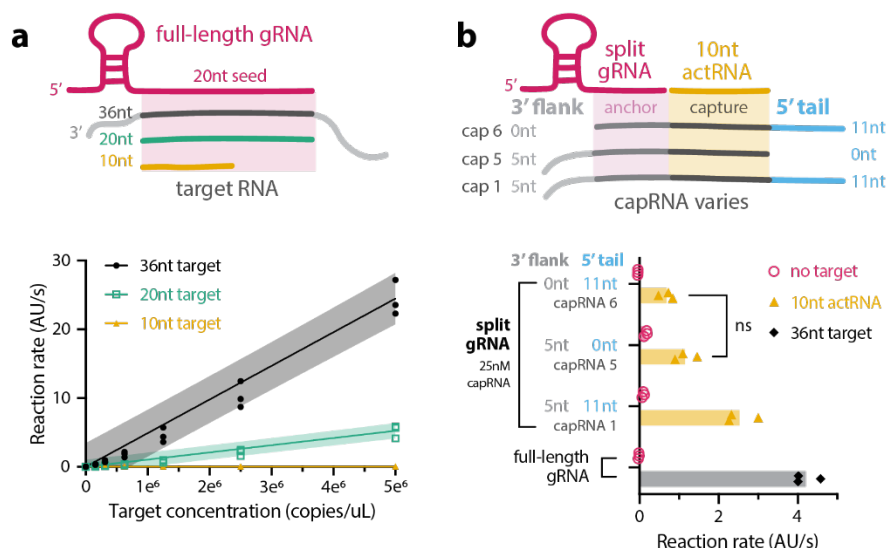


Figure 6. Importance of seed flanking sequences.

Exploration of the role of nucleotides flanking the seed region in the target RNA and capRNA. Subfigures reproduced from Ch. 2, Fig. 1b and Ch. 2, Extended Data Fig. 2c, respectively. **A**, Limit of detection analysis of full-length gRNA detecting 36, 20, and 10nt target RNA. **B**, Comparison of capRNA with and without a 3' flank and/or 5' tail detecting a 10nt actRNA.

More testing is required to understand the optimal length and sequence of the 3' flank. I have only tested two lengths (0 and 5nt) and three sequences of the 3' flank. A particularly intriguing case is the comparison between capRNA 1 and 2. Both have the same anchor and capture region, a 5nt 3' flank, and an 11nt 5' tail. Neither have predicted secondary structure in the capture region or 5' tail. The main differences are the sequences of the 3' flank and the 5' tail and the accidental base pairing of 1nt 3' to the anchor region of capRNA 1. This base pairing results in an effective 3' flank of 4nt for capRNA 1. The difference in the performance of the two capRNAs is striking: both generate a similar positive signal when 10nt actRNA is added, but capRNA 2 has such a high background signal that the difference between the background and positive signal is not statistically significant. Thus it seems that the only possible contributors to this difference is the sequence of the 3' flank and 5' tail. It would be interesting to use molecular dynamics to simulate the structure of the Cas13a split guide system, comparing the two capRNAs, to see if there are any predicted structural differences between the two complexes.

I would recommend focusing initial efforts on finding an optimal sequence and length of the 3' flank. Since the 3' flank doesn't interact with binding the actRNA, it becomes a constant in the design of the system. I would test different lengths, ranging from 0 to 5nt and different sequences, based on successful full-length gRNAs. I think the understanding the rules guiding design of the 3' flank is less critical since, like the anchor region, once an optimal 3' flank has been selected, no further modifications need to be made from assay to assay. That being said, the curious researcher may want

to investigate more to satisfy their desire to understand the system. I have already designed and ordered a set of capRNAs to begin the exploration of the 3' flank.

The 5' tail of the capRNA, on the other hand, will likely vary between applications, motivating a need to better understand rules for what makes a good 5' tail. While it seems like it is useful to have both a 3' flank and 5' tail, it is unclear how long a 5' tail is needed. I tested actRNA of varied lengths paired with capRNAs with different length capture regions (Fig. 7a). From this testing, I saw that binding the whole actRNA and having a 5' tail (unbound nucleotides 5' of the capture region) boosted the positive signal (Fig. 7b and c). For example, with the 10nt actRNA there was a significant increase in the background-subtracted reaction rate going from a capRNA with a 10nt capture region and no 5' tail to one with a 7nt tail. A similar result occurred for the 13nt actRNA (3nt unbound actRNA to 4nt 5' tail), 15nt actRNA (5nt unbound actRNA to 2nt 5' tail), and 20nt actRNA (10nt unbound actRNA to 1nt 5' tail). The actRNA that had a perfect match with their capRNA and no 5' tail (21nt and 17nt actRNA) or were still longer than the capRNA (24nt actRNA) did not see the same gain in signal. This data supports the theory that having an unbound 5' tail on the capRNA, even just 1nt, helps boost the positive signal. Boosting signal would help increase assay sensitivity and improve detection of small RNA targets.

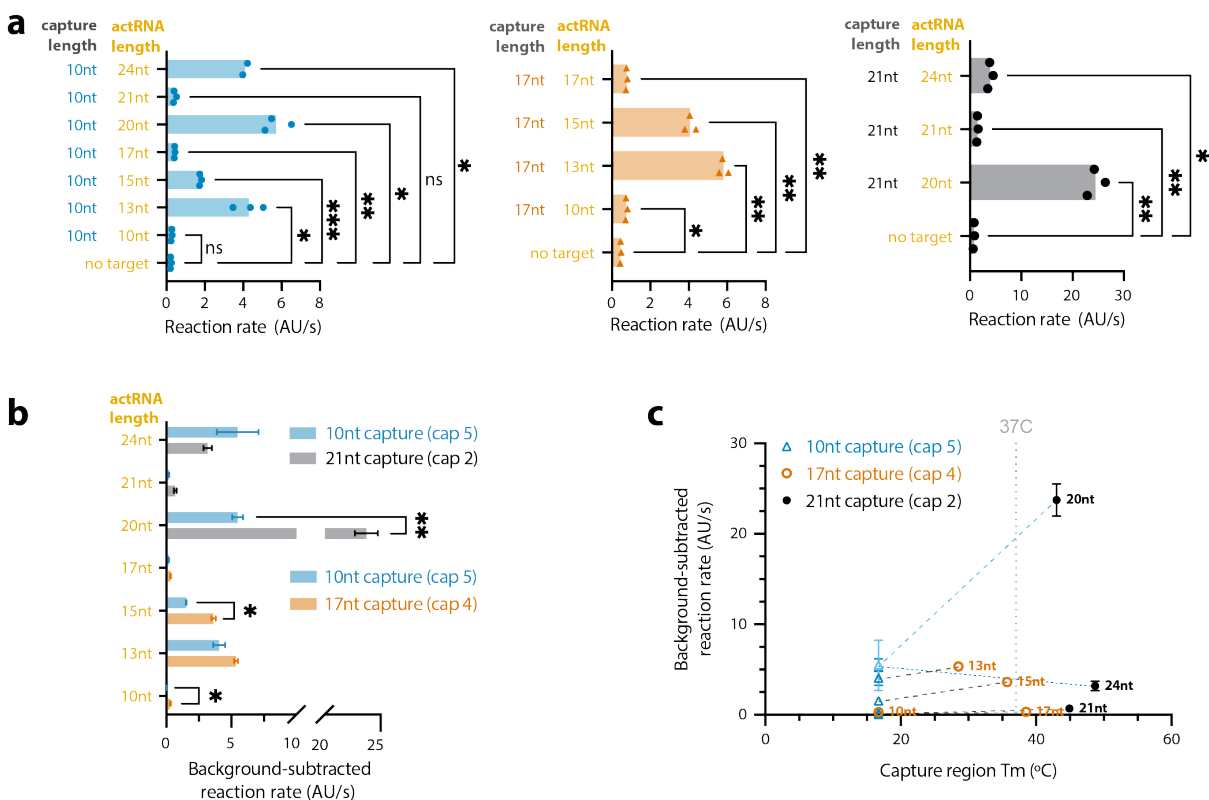


Figure 7. Test of actRNA detection with capRNAs of varied length.

Subfigures reproduced from Ch. 2, Extended Data Fig. 4c and d for subfigures A and C, respectively. Subfigure B is derived from subfigure A. A, ActRNA detection with capRNA 5 (left panel, 10nt capture region), capRNA 4 (middle panel, 17nt capture region), and capRNA 2 (right panel, 21nt capture region). B, Background-subtracted reaction rates from A. C, Analysis of change of signal as capture region melting temperature increases.

However, depending upon the purpose of the assay, having an unbound 5' tail may not be necessary and may add undesired complications. For example, if the actRNA already leads to a robust signal, it may not be necessary to further boost that signal. Adding a 5' tail may introduce the opportunity for other off-target RNA species to bind, depending on the sequence and length of the 5' tail. This may be especially important in the detection of miRNA isoforms, which can have untemplated nucleotide additions on the 3' end³⁸. Thus, it is critical to query what other sequences are similar to the target RNA in a given sample.

When modifying the 3' and 5' tail, it is critical to understand how those changes affect the background signal generated by the capRNA in complex with the Cas13a RNP. Before ordering and testing a capRNA, it is useful to check for predicted secondary and tertiary structure. To predict these structures, I use NUPACK. In NUPACK, the user can specify the sequences present and the number of strands permissible in a complex. Unfortunately, there is no option to specify the salt concentration of the environment.

Having a predicted stem loop structure in the capture region may explain high background activity, though this is not always the case. For example, when comparing the split guide 1 system with the split guide 2 system, the background signal is much greater from split guide 2 system (Fig. 8b). The capture region and 5' tail of capRNA 7 (split gRNA 2 system) has a predicted stem-loop, whereas capRNA 1 (split gRNA 1 system) has no predicted secondary structure in the capture region and 5' tail (Fig. 8a). While the difference in secondary structure could explain the difference in background activity, it is also possible that the difference in sequence contributes to the varied background activity. It would be interesting to see if mutating the 5' tail of capRNA 7 to remove the secondary structure causes a reduction in the background signal. I have designed and ordered a few alternate capRNAs for the split gRNA 2 that don't have stem-loop structures. Testing those capRNAs would help to build our understanding of the role of secondary structure in capRNA performance.

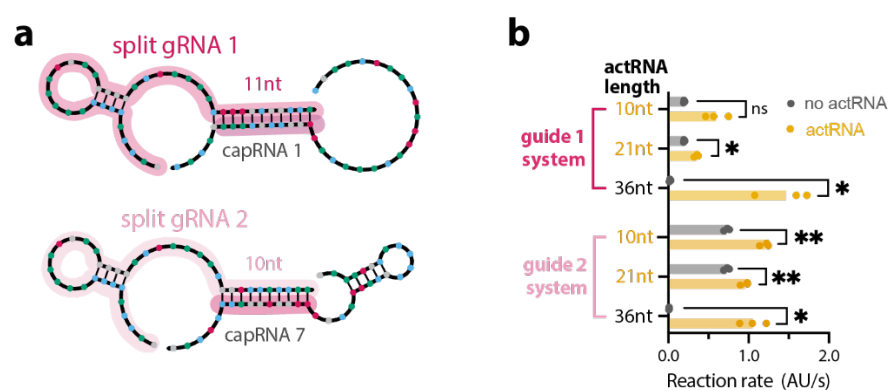


Figure 8. Stem loop structures in the capRNA affect background signal.

Subfigures reproduced from Ch. 2, Extended Data Fig. 4a-c. **A**, predicted structure of the split gRNA bound to capRNA in NUPACK. **B**, Reaction rates for two sequence-unique guide systems. Both systems tested with a full-length gRNA paired with a 36nt target RNA, or a split gRNA paired with a 10 or 21nt actRNA. Background signal from capRNA shown with grey bars.

Another example supporting the secondary structure theory is capRNA 3 versus capRNA 1. CapRNA 3 has a 10nt capture region, an 18nt 5' tail, and a large predicted stem-loop; capRNA 1 has the same 10nt capture region and an 11nt 5' tail with a different sequence, and no predicted secondary structure (Fig. 9a). CapRNA 3 has such a high background signal that the positive signal is no different, whereas capRNA 1 has a low background signal and the positive signal can be clearly differentiated from it (Fig. 9b). However, in this same experiment, I test capRNA 2, which has the same 10nt capture region,

a different 11nt 5' tail that is sequence matched to capRNA 3's 5' tail, and no predicted secondary structure. CapRNA 2 has a similarly high background as capRNA 3, and its positive signal is not significantly greater than the background. Together, this data suggests that having a stem-loop structure in the capture region and 5' tail may contribute to a high background signal, but the lack of a stem-loop structure doesn't ensure that the capRNA will yield a low background signal.

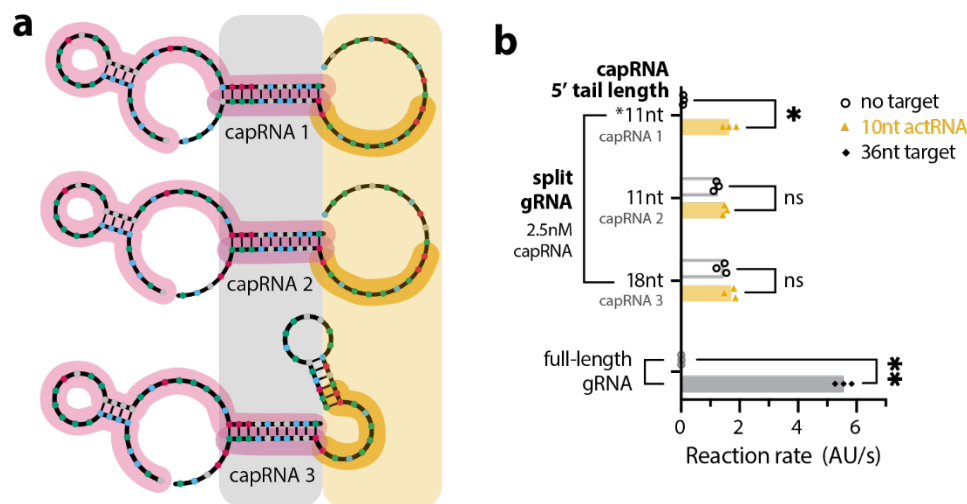


Figure 9. Analysis of secondary structure in split gRNA-capRNA duplexes.

Subfigures adapted from Ch. 2, Extended Data Fig. 3 and Ch. 2, Fig. 2b. **A**, Predicted structures of split gRNA bound to capRNA in NUPACK. **B**, Reaction rates for the three capRNAs detecting a 10nt actRNA. Background signal from capRNA plotted with grey-outlined bars.

Finally, I want to discuss the data for the capRNA I designed for six cellular miRNA targets (Fig. 10). This data weakly supports the theory that a lack of stem-loop structure correlates with low background signal and that the presence of a stem-loop structure is correlated with high background signal, depending on the size and location of the stem loop. The capRNA for miR_4505 had no predicted stem loop and the lowest background activity. The next smallest background activity was from the capRNA for miR_31, which had a small stem loop structure spaced 10nt from the anchor region. The next two capRNAs had similar background activities (for miR_3178 and miR_221) and stem loops with a loop of increasing size and proximity to the anchor region. The capRNA with the highest background (miR_17) had a stem loop structure that was spaced 5nt from the anchor region, but with an additional 5nt dangling beyond the stem loop, perfectly fitting (though not complementary base pairing) the space between the anchor region and the start of the stem loop.

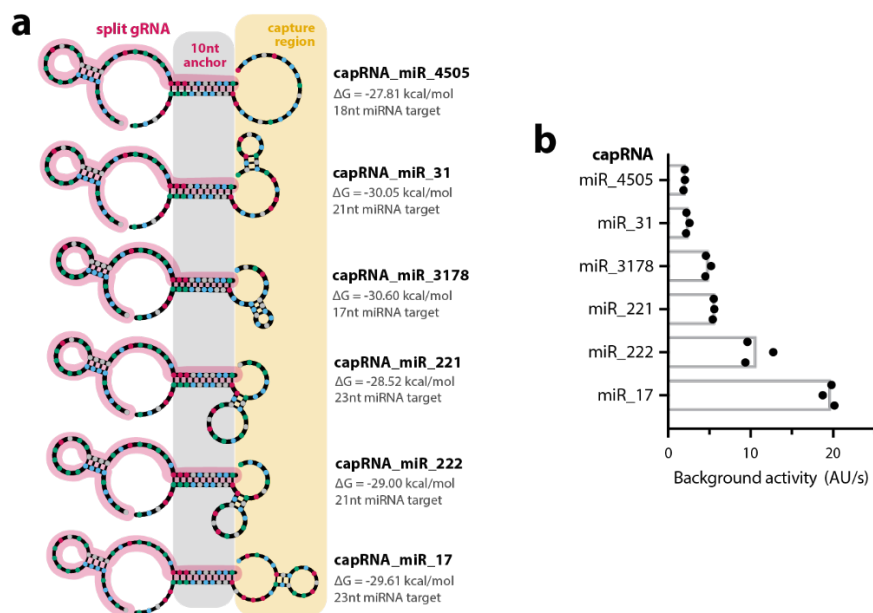


Figure 10. Structural analysis of capRNAa for cellular miRNA targets.

Subfigures adapted from Ch. 2, Extended Data Fig. 7 and 6c. **A**, NUPACK predicted structures of split gRNA in complex with capRNA, ordered from top to bottom by increasing background activity. **B**, Background activity of the different capRNAs.

Two strategies to remove predicted stem-loop structures in the capRNA is to change the sequence of the 5' tail or introduce key mutations in the capture region. Modifying the sequence of the 5' tail should be tried first, as it shouldn't affect the interaction between the capRNA and the actRNA. If the stem-loop structure can't be removed through modifications of the 5' tail, or if there is no 5' tail, it may be necessary to introduce mutations in the capture region. In the work of Tambe et al, they see that four nucleotide mismatches in positions 1-4 or 13-16 in the seed region of the gRNA results in an increased cleavage rate³⁹. This increase is particularly dramatic when the mismatches are in position 13-16: the apparent cleavage rate increase by two-fold. Perhaps introducing mismatches in positions 13-16 on the capRNA would result in a similar increase in signal, or at worse, wouldn't decrease the reaction rate. From my experiments, I see that a 2nt mismatch between the capRNA and actRNA in positions 3-4 of the capture region does not significantly impact the signal relative to the no mismatch control, whereas a 2nt mismatch at positions 5-6 does reduce the signal (Fig. 5c). I do not know how 1nt mismatches affect the reaction rate. To introduce mismatches for purposes of reducing stem-loop structure, I would recommend a 2nt mutation at positions 3-4 or a 1nt mutation within positions 3-6.

Another idea that I'm curious about is the potential of a fold-back capRNA to improve specificity. This idea is inspired by the strategy of using a foldback crRNA to detect small RNA targets³⁶. The 5' end of the capRNA would fold back on itself, creating a 5' cap. This cap would limit the size of RNA that could bind to the capRNA, preventing longer RNA sequences that are partially complementary from binding. However, the 5' cap is essentially a stem-loop structure. I wonder if, with careful design to the size of the stem-loop and distance from the seed region, it wouldn't result in high background signal, like the capRNA for miR_31 (Fig. 10).

All the design modifications I have suggested would be expensive and time-intensive to test; but what if DNA could be used in lieu of the capRNA? A pre-print recently demonstrated that Cas13a can also be activated by single stranded DNA⁴⁵. Thus, I wonder whether capRNA could be replaced with capDNA. Ordering DNA is much cheaper and faster than RNA and would make rapid iterations of the capture strand much faster. I propose testing whether capDNA mimics the behaviors of capRNA, both desirable and undesirable. If it does, then I would recommend doing much of the design and troubleshooting with capDNA and then ordering the final optimized sequence as capRNA.

Alternatively, capDNA could be kept in the system. Since Cas13a doesn't cleave DNA, the capDNA would not be at risk of degradation by the activated Cas13a. This could help boost the sensitivity of the assay, as more capDNA would be available to bind to the target RNA. However, it might also contribute to a higher background signal, negating any potential gains. It would also be interesting to test whether DNA could replace the actRNA. Replacing the actRNA with DNA would similarly help with the cost and speed of assay design and testing.

Design Principles Summary & Next Steps

A summary of my hypotheses of the principles for designing a split guide system – split gRNA, capRNA, and actRNA – is listed below (Fig. 11), along with recommended next experiments to validate these theories. Following the summary, I include a schematic of all the RNA components we currently possess for the split guide system to aid with experimental design (Fig. 12).

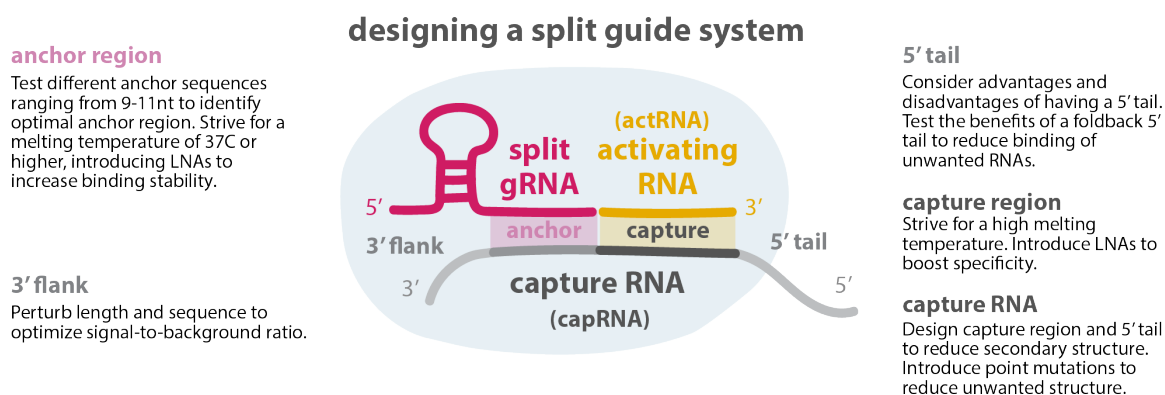


Figure 11. Summary of the principles for designing a Cas13a split guide assay.

1. Anchor region

- a. Strive for an anchor region ranging from 9-11nt with a melting temperature around 37°C.
- b. Verify the performance of the anchor region with a constant actRNA and capture region.
- c. Next steps
 - i. Test whether DNA can replace RNA as the capture strand.
 - ii. Compare split gRNAs 1 and 2 with a 10nt anchor region and the same actRNA.
 - iii. Compare split gRNA 1 with a 10 and 11nt anchor region and the same actRNA.
 - iv. Test new sequences for the anchor region based on successful full-length gRNAs.
 - v. Test whether LNAs in the anchor region increase binding stability, boosting signal-to-background ratio.

2. Capture region

- a. Check the melting temperature of the actRNA; higher is likely better.
- b. Modify the length of the capture region to prioritize sensitivity of specificity.

- c. Introduce mismatches to remove any unintended secondary or tertiary structure in the capRNA.
- d. Introduce LNAs in the capture region of the capRNA to enhance duplex stability and mismatch discrimination.
- e. Next steps
 - i. Test whether introduction of mismatches in the capRNA to reduce secondary structure disrupts positive signal.
 - ii. Consider mismatches in positions 13-16 (number based on seed region of full-length gRNA).
 - iii. Test whether LNAs in the capture region increases positive signal and helps with 1 and 2nt mismatch discrimination.

3. CapRNA

- a. Optimize the 3' flank length (1-5nt) and sequence. Take inspiration from successful target RNAs from the full-length gRNA system.
- b. Decide whether to have 5' tail, and if so, what length and sequence.
- c. Using NUPACK, check resulting capRNA designs for secondary structure when bound to the split gRNA.
- d. Modify the 5' tail or introduce point mutations in the capture region to reduce secondary structure. If introducing point mutations, consider 1nt mutations at positions 13-16, counting from the start of the anchor region, or 2nt mutation at position 13-14.
- e. Consider molecular dynamics simulations of the capRNA bound with the Cas13a and split gRNA to aid the design process. Compare newly designed complexes to predicted structures of experimentally-optimized split guide structures.
- f. Next steps
 - i. Test whether DNA can replace RNA as the capture strand. Test replacing a capRNA with low background and one with high background. If capDNA replicates behaviour of capRNA, conduct design and optimization testing with capDNA. If capDNA performs better than capRNA, consider switching to capDNA.
 - ii. Test whether mutating away secondary structure helps to reduce background signal. Test modified capRNAs for split guide system 2.
 - iii. Optimize 3' flank. Test capRNAs with varied length and sequence 3' flanks.

Split guide system components list

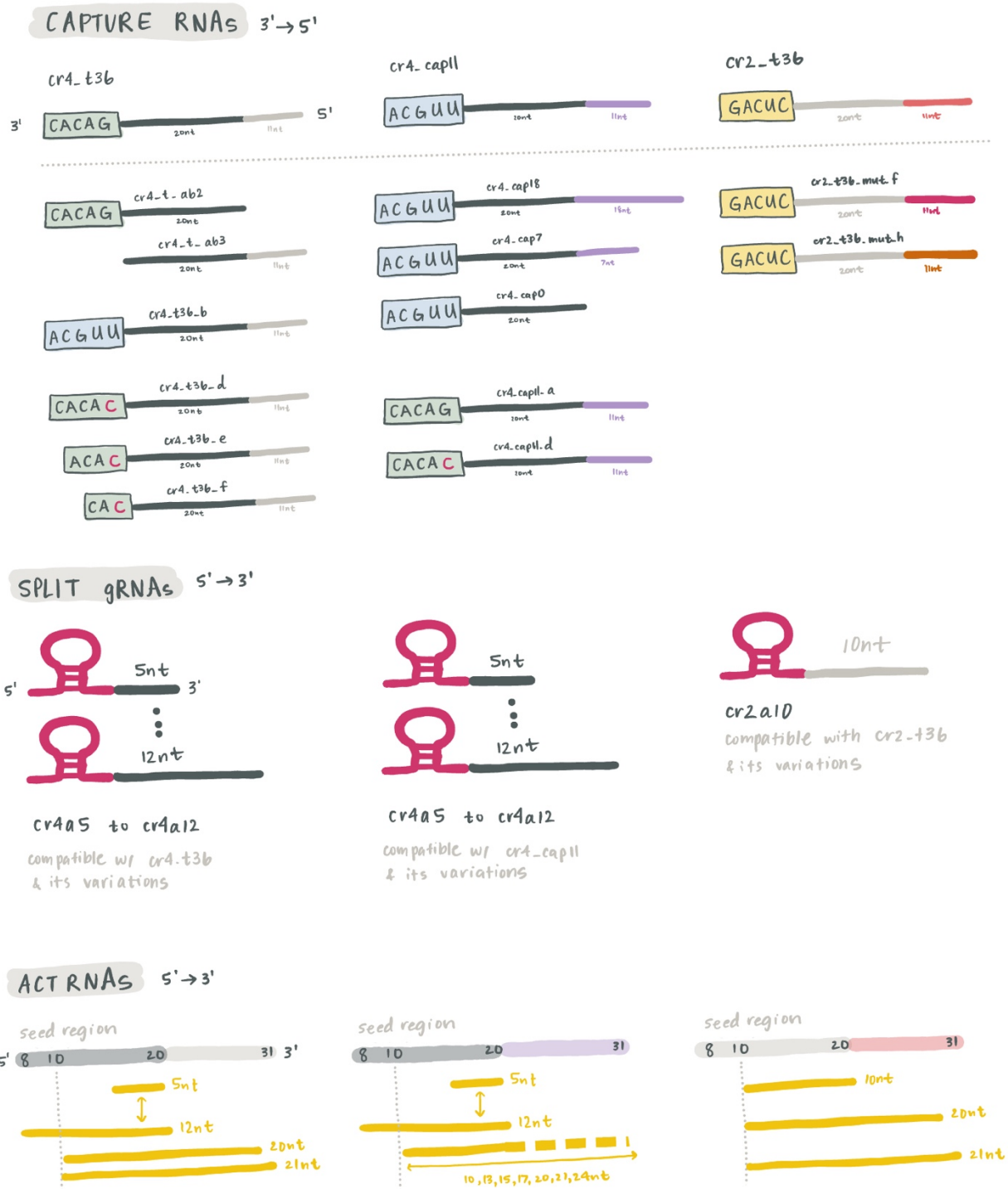


Figure 12. Split guide system components list.

Schematic of all split guide system RNA components, except actRNA for mismatch and gap testing.

Chapter 4 – Turning pathogens into partners: Converting Influenza A Virus into a delivery platform for CRISPR-Cas gene therapy

Amanda Meriwether, Roni Weissman, Abinaya Anbuchelvan, and Daniel A. Fletcher

Unpublished work.

Abstract

Gene editing with CRISPR-Cas9 offers the potential to cure genetic diseases and engineer cell therapies against cancer. However, its therapeutic promise is limited by challenges of *in vivo* delivery that lead to inefficient, unreliable, and off-target gene editing. New strategies focused on delivering Cas9 protein hope to solve many of the issues associated with delivery of Cas9, guide RNA, and template RNA as gene cassettes. Here we present a new delivery platform based on Influenza A Virus (IAV) that uses its unique biophysical features to overcome existing barriers to *in vivo* CRISPR-Cas9 gene editing. Our approach leverages the IAV genome packaging machinery to introduce guide RNA (gRNA) complexed with Cas9 within a non-replicating Influenza A virus-like particle to form a Cas9 Influenza-based Delivery Vehicle (CasIDV). We demonstrate proof-of-concept that these CasIDVs mediate gene editing in cells. This work expands the options for delivery of gene editing tools to cells and suggests a potential solution for editing hard-to-reach cells protected by mucosal layers.

Introduction

To fully realize the therapeutic potential of CRISPR-Cas gene editing tools, it is necessary to develop efficient and specific methods to deliver these tools to cell in situ. Current strategies rely on the use of viral vectors, like AAV, to deliver DNA cassettes encoding Cas9, gRNA, and template DNA^{47,48}. These methods are limited by the genomic carrying capacity of the viral vector, tropism of the viral vector, and the need to express the Cas9 protein and form the RNP in the cell. Furthermore, the long-term expression of Cas9 from plasmid DNA delivered to cells may result in off-target editing events, which can be toxic to the cell. An alternative is to deliver Cas9 protein already complexed with gRNA, reducing the likelihood of off-target editing events due to the transient nature of the Cas9 RNP. Strategies to deliver Cas9 RNPs include virus-derived enveloped particles. Hamilton et al have developed Cas9-packaging enveloped delivery vehicles (Cas9-EDVs) that are derived from HIV^{49,50}. They fuse Cas9 to the HIV Gag protein to package Cas9 protein inside a budding viral-like particle. With this strategy, they demonstrate targeted deliver of Cas9 RNPs to T-cells *in vivo* and *in situ*.

We propose to develop a complementary system derived from a different enveloped virus – Influenza A Virus (IAV) – that leverages IAV's unique genetic and biophysical characteristics to deliver Cas9 protein, guide RNA, and eventually template DNA within non-replicating virus-like particles. Like many viruses, IAV has evolved to enter cells and deliver cargo to their nuclei. What distinguishes IAV as a promising viral vector is its (1) large carrying capacity due to its pleiomorphic filamentous shape^{51,52}, (2) segmented genome that can be modified to package genes and proteins of interest while preventing the virus from replicating^{53,54}, and as we recently discovered, (3) spatially organized envelope that allows IAV to ratchet itself through the mucous layer in the airway to the lung epithelium⁵⁵.

Specifically, Influenza A virions are spherical to filamentous with a diameter approaching 100nm⁵¹. The filamentous particles can extend over 1µm in length⁵¹. The interior volume of the virion is large and has the potential to accommodate a substantial protein and plasmid cargo. At one end of the virion is the viral genome, composed of eight unique negative sense viral RNA segments complexed with viral nucleoprotein and polymerase complex to form a viral ribonucleoprotein (vRNP) (Fig. 1a)⁵⁶. The eight vRNPs are selectively packaged inside the budding virion in a ‘7+1’ configuration, where one vRNP is surrounded by the other 7 vRNPs (Fig. 1b)⁵⁷. To synthesize IAV from scratch, a plasmid reverse genetics system was developed encoding one vRNA segment per plasmid⁵⁸. With this reverse genetics system, it is relatively easy to synthesize minimal virus-like particles (VLPs) that maintain the biophysical properties of IAV but are non-replicative. Furthermore, there is the potential to pseudotype the enveloped virus-like particles to alter the particle tropism⁵⁹.

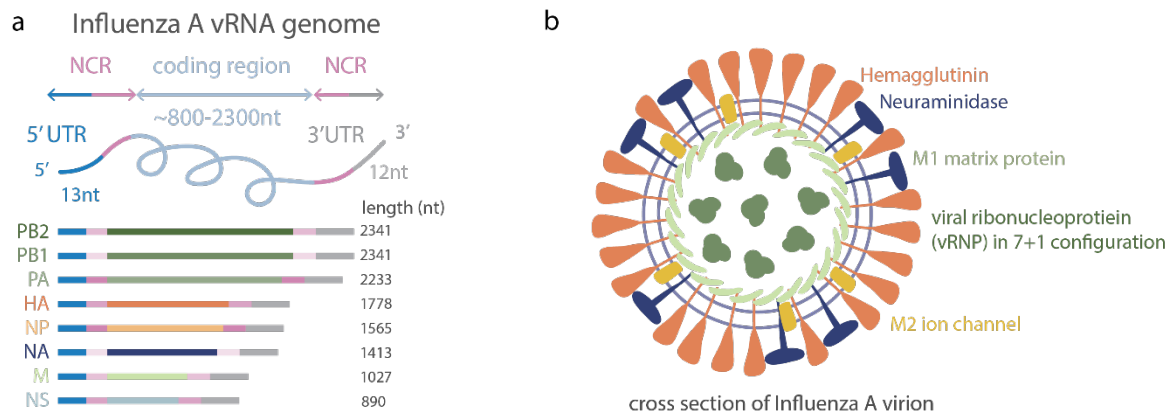


Figure 1. Influenza A Virus genomic architecture.

A, Schematic of eight-segmented IAV genome, adapted from Dou et al 2018⁵⁶. B, Schematic of a cross-section of an Influenza A virion, highlighting the ‘7+1’ configuration of the vRNPs and IAV structural proteins.

IAV virus-like particles can be synthesized that maintain the desirable characteristics of a delivery vehicle and reduce the unwanted side effects of a full virion. The key properties that we want to maintain in an IAV-derived delivery vehicle are its ability to access mucous protected cells, enter cells via endocytosis, escape the endosome and release its genetic cargo, and transport that genetic cargo to the nucleus. To achieve this, VLPs should be synthesized with all the viral proteins, but ultimately, none of the viral RNA. Omitting the viral RNA will prevent the VLPs from replicating inside the host cell and generating new VLPs.

Our approach to engineering Influenza-derived delivery vehicles for delivery of gene editing tools is to leverage the native IAV genome packaging machinery. We propose to package a guide RNA (gRNA) complexed with Cas9 within a non-replicating IAV VLP by disguising Cas9 gRNA as viral RNA (Fig. 2a). We term this particle a Cas9 Influenza-based Delivery Vehicle (CasIDV). By tricking the IAV genome packaging machinery into loading Cas9 RNP cargo inside the CasIDV based on the chimeric vRNA-gRNA, we control a packaging plug-and-play system where guide RNA can be swapped out for different gene editing applications. To develop the CasIDV platform, we think of the problem in two main steps: 1) packaging Cas9 RNPs inside the IDVs (Fig.2a) and 2) delivering the Cas9 RNPs to and editing cells (Fig. 2b).

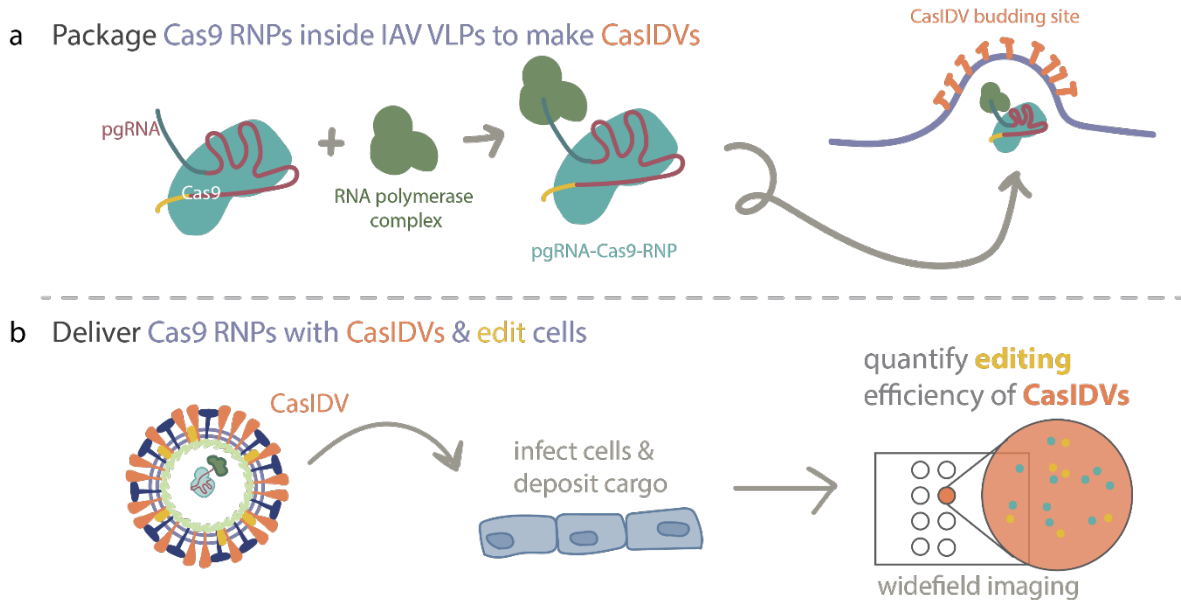


Figure 2. Graphical overview of CasIDV project.

A, Schematic of process of packaging Cas9 RNPs inside budding CasIDVs. B, Schematic of delivery of CasIDVs to cells and analysis of editing.

Results

Designing the Cas9 Influenza-based Delivery Vehicle (CasIDV) system

Influenza A Virus can be synthesized from scratch using a plasmid-based reverse genetics approach. In this system, a set of eight plasmids, one for each of the vRNA segments, is transfected into factory cells. Each plasmid contains a bidirectional construct that generates positive sense mRNA from RNA Polymerase II transcription and negative sense viral RNA from RNA polymerase I transcription (Fig. 3a)⁵⁸. Thus, each plasmid results in the expression of both the viral protein and vRNA segment. To create fully infectious IAV, it is common to transfect this set of eight plasmids into a co-culture of HEK 293T and MDCK cells^{55,60}. The HEK cells have high transfection efficiency and are generally thought to generate the primary IAV virions. These virions can then infect MDCK cells in the culture, which are hardy to infection, budding off many virions before they die. This co-culture system enables the generation of a high viral titer from a set of eight plasmids (Fig. 3b).

The reverse genetics approach for generating IAV is flexible, since the IAV genome is spread across eight segments on eight different plasmids. To make non-replicative particles, or virus-like particles (VLPs), one can replace a bidirectional plasmid with one that only synthesizes viral mRNA (Fig. 3b). This approach can be taken to the extreme to make IAV VLPs that contain no viral RNA, or VLPs that are missing some of the IAV proteins⁵².

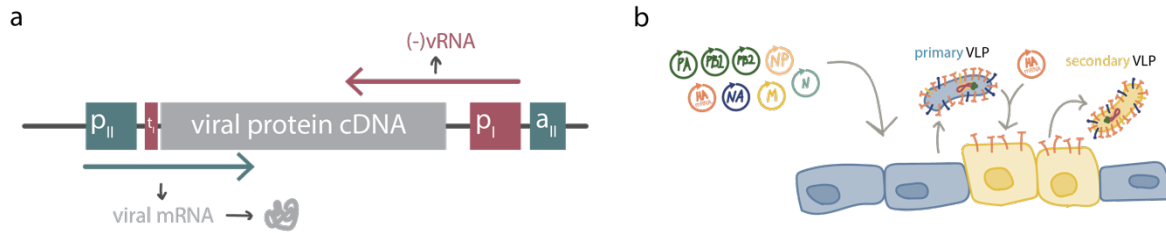


Figure 3. Overview of the IAV plasmid reverse genetics system.

A, Schematic of the bidirectional construct resulting in the synthesis of viral mRNA for viral protein expression and viral RNA for packaging inside virions. B, Schematic of transfection process to make Influenza A virus-like particles using the reverse genetics system. Briefly, eight plasmids are transfected into a culture of HEK 293T and MDCK II cells. One plasmid (for HA) only synthesizes viral mRNA, resulting in the generation of VLPs that are not fully infectious. When primary VLPs from the HEK cells infect MDCK cells that also express the HA protein, secondary VLPs are generated.

For the CasIDV application, we wanted to make VLPs that maintained the advantages of IAV, namely its long, filamentous shape, ability to directionally ratchet through mucous layers, and entry into cells and delivery of genetic payloads into the host cell's nucleus. However, for safety reasons, we did not want the virus to replicate and cause secondary infection of surrounding cells. The safest way to achieve both those goals would be to make IAV VLPs containing no viral RNA. From a practical experimental perspective, synthesizing empty VLPs is an inefficient process. A co-culture system cannot be used, as any VLP that infects a MDCK cell would not result in the generation of new viruses. Thus, for every experiment, the eight viral plasmids would need to be transfected into HEK cells, along with the plasmids for the cargo to be packaged inside the VLPs. Assuming that we added one additional plasmid encoding the Cas9 gRNA and transfected HEK cells already expressing Cas9, we would need each cell to receive nine plasmids to have a chance at producing a CasIDV. HEK cells do have a high transfection efficiency, but even assuming an efficiency of 80%, the probability that one cell received all nine plasmids is 0.89, or roughly 0.13.

To increase the CasIDV titer, we opted instead to use a leave-one-out approach, where we replaced one segment of viral RNA with the Cas9 gRNA. This strategy would enable us to utilize a modified co-culture system to synthesize more VLPs (Fig. 3b). With the co-culture, we would make VLPs missing one vRNA segment, supplementing expression of that protein by exogenously expressing the missing protein in the MDCK cells. Then, we could transfect a new culture of cells with plasmids encoding the Cas9 RNP cargo (gRNA and Cas9), the missing viral proteins, and infect those same cells with the partially empty VLPs (Fig 4). A subset of particles that then bud off would be CasIDVs.

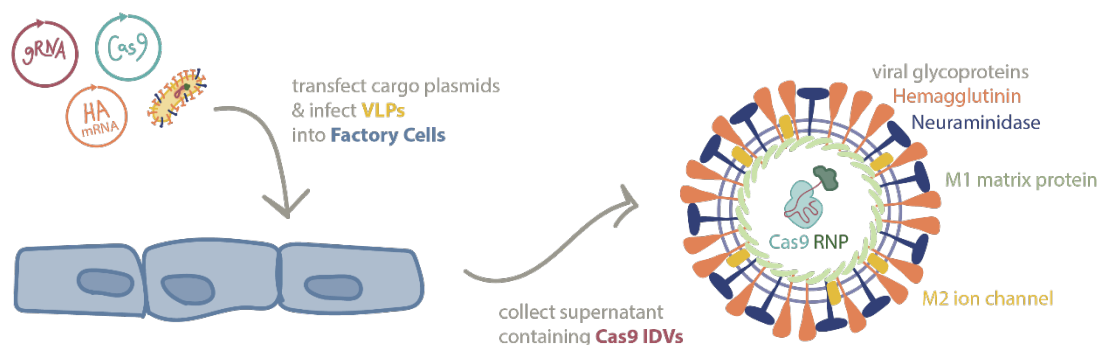


Figure 4. Proposed strategy for CasIDV synthesis.

Plasmids encoding the gRNA, Cas9, and HA mRNA are transfected into cells that are simultaneously infected with IAV VLPs lacking the HA vRNA. The resultant particles generated by the factory cells would then be a mixed population of more IAV VLPs and some CasIDVs.

When choosing which vRNA segment to omit, I prioritized safety, which led me to consider the IAV glycoproteins: hemagglutinin (HA) and neuraminidase (NA). Hemagglutinin is critical for binding to sialic acid on the cell's surface and fusion with the endosomal membrane⁵⁶; without it, VLPs are unlikely to infect a cell. Neuraminidase cleaves sialic acid to release the virion, enabling the virus to move through thick mucous layers⁵⁵. Without NA, an IAV VLP could still infect a culture cell, so long as it wasn't producing a mucous layer. Thus, I decided to replace the HA vRNA segment with a Cas9 gRNA.

Designing Cas9 viral packaging guide RNAs (pgRNAs)

The CasIDV system leverages the native packaging mechanism of IAV to package non-native guide RNA and Cas9 protein into the VLPs. To package the Cas9 gRNA inside the VLP, we disguise gRNA as viral RNA (vRNA) by appending the vRNA sequences that are critical for packaging native vRNA into budding virions. The vRNA-gRNA chimera then serves as bait for Cas9, which will be transported as a complex with gRNA into the assembling VLP. We term this vRNA-gRNA chimera “packaging guide RNA” or pgRNA.

The native vRNA packaging mechanism of IAV is impressive – it must selectively package eight unique vRNA segments from among all the cell and viral RNA expressed in the cell. Without one copy of each of the eight segments, a virion is not fully infectious. This specific packaging is mediated in part by the vRNA sequences. Each vRNA segment is composed of a coding region flanked on either side by a non-coding region (NCR) (Fig. 5). Each NCR is further subdivided into the proximal untranslated region (UTR), which is mostly conserved across all eight vRNA segments. Variable amounts of the NCR and coding region on the 5' and 3' ends of the vRNA are considered its packaging signal⁵⁷. The packaging signal can be further differentiated into the incorporation signal and the bundling signal. The incorporation signal, composed of the NCR, is the sequence required for a given vRNA to be packaged inside a

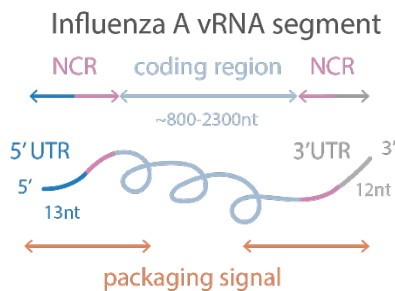


Figure 5. Annotated Influenza A viral RNA segment.

Non-coding region (NCR), composed of conserved untranslated regions (UTRs) and segment specific NCR flank the coding region. The packaging signal encompasses the NCR and part of the coding region.

virion. The bundling signal, extending into the coding region, is the additional sequence required for packaging of other vRNAs into the virion⁵⁷. This explanation is an oversimplification, as there are other portions of the vRNA throughout the coding sequence that likely also play a role in mediating specific packaging of all eight vRNA segments through RNA-RNA interactions. Furthermore, different segments appear to be more or less important for packaging of all eight segments, suggesting a hierarchy in packaging. Studies have demonstrated that PB2, PA, NP, and M all play critical roles in packaging^{61,62}.

Sufficient packaging signals of the HA segment were validated by replacing the HA coding sequence with the sequence for GFP, GFP-GFP, or RFP and maintaining the 45nt NCR plus 80nt of the coding region on the 5' end the 32nt NCR plus 45nt of the coding region on the 3' end⁵³. The resultant VLPs produced fluorescent protein in infected cells, reinforcing the idea that a vRNA segment can be replaced with non-native RNA so long as sufficient vRNA packaging signal is included. While it seems likely that nestling a Cas9 gRNA sequence between the demonstrated packaging signals would result in packaging of the gRNA inside an IAV VLP, it is less obvious whether a Cas9 protein would bind to the gRNA, and whether the resultant Cas9 RNP would be functional. Part of this doubt stems from the fact the 5' and 3' UTRs partially complement with each other, forming a corkscrew structure that complexes with the three-protein viral polymerase⁶³. This folding of the native vRNA may sterically inhibit the Cas9 from binding, or the Cas9 may prevent the corkscrew structure from forming, impeding packaging of the Cas9-pgRNA complex.

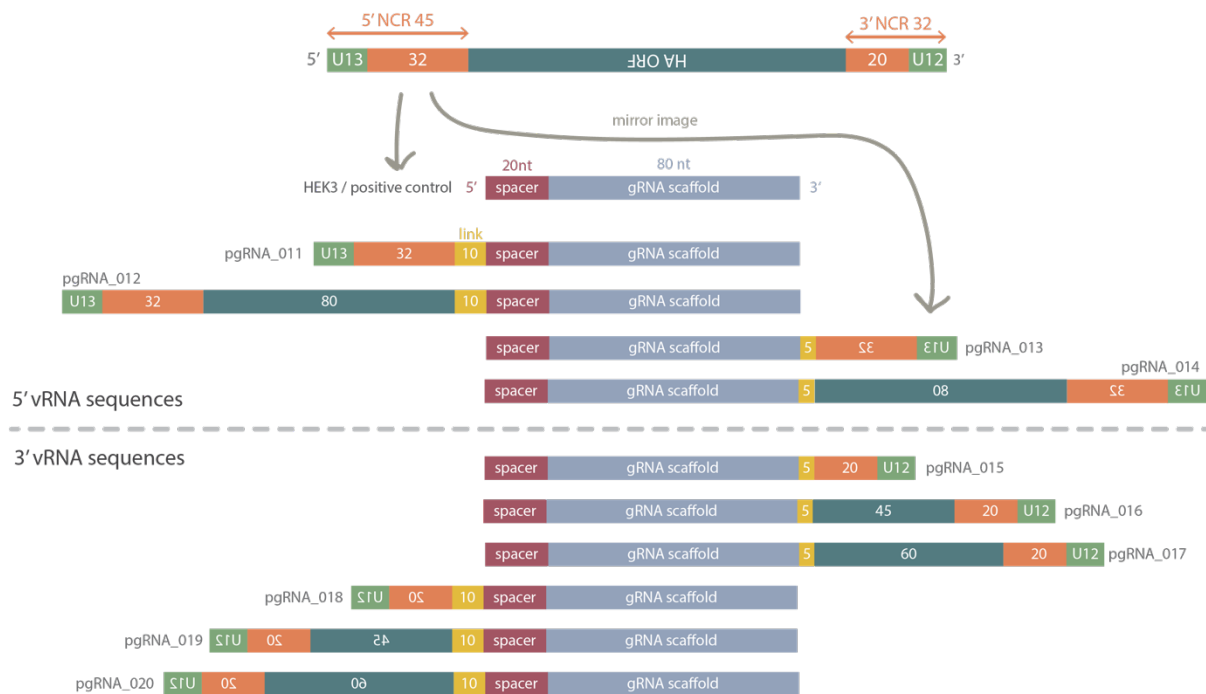


Figure 6. Packaging guide RNA (pgRNA) designs.

Variable amounts of the 5' and 3' packaging signals of the HA vRNA segment are appended onto the 5' and 3' end of the gRNA scaffold.

In attempts to balance the need of including sufficient packaging signal to package the gRNA inside the VLP without disrupting Cas9 RNP formation and function, I added either the 5' or the 3' packaging signal to either the 5' or 3' end of the gRNA scaffold (Fig. 6). Previous work had shown that the 5' end of the vRNA is sufficient for binding to the viral RNA polymerase complex⁶³, suggesting that I may be able to achieve successful packaging with just the 5' packaging signal. Based on the structure of Cas9 bound to gRNA⁶⁴ and the development of prime editing gRNAs⁶⁵, I hypothesized that appending the vRNA packaging signal to the 3' end of the gRNA scaffold would be the least disruptive to Cas9 RNP function. With these design considerations, I synthesized a set of pgRNAs to test my hypotheses.

Packaging gRNAs mediate dsDNA cleavage *in vitro* and *in vivo*

We first tested the pgRNAs in an *in vitro* cleavage assay to validate that the Cas9 RNP formed and was functional. Briefly, we incubated Cas9 and pgRNA to form Cas9 RNPs before adding target DNA for cleavage. We then analyzed dsDNA cleavage on an agarose gel (Fig. 7a). We show that Cas9 RNPs formed with pgRNAs permit dsDNA cleavage (Fig. 7b). Packaging gRNAs 13, 15, 16, and 17 all demonstrate efficient cleavage, while pgRNA_018 demonstrates partial cleavage of the target DNA. The pgRNAs that seem to perform the best all have vRNA packaging signal appended to the 3' end of the gRNA scaffold. Packaging gRNA_018 has the vRNA packaging signal appended to the 5' end of the gRNA scaffold, but the vRNA sequence appended is only 32nt, compared to the 45nt and 125nt appended on the 5' end of pgRNA_011 and 12. This difference suggests that there is some tolerance to additional nucleotides on the 5' end of the gRNA, but that there is a length limit.

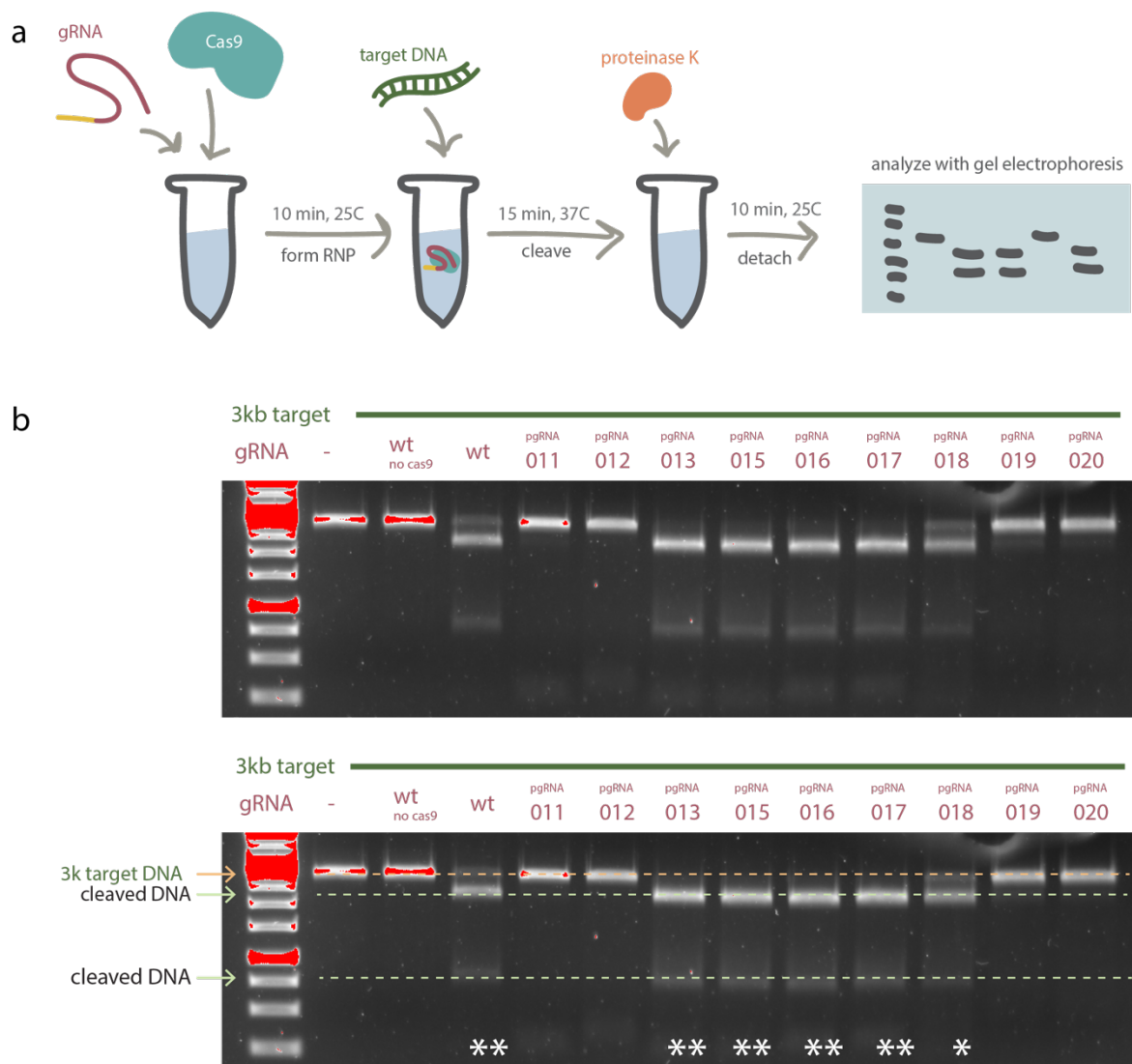


Figure 7. *In vitro* Cas9 cleavage assay.

A, Schematic illustrating *in vitro* cleavage assay. Cas9 RNPs are formed by incubating Cas9 protein with gRNA or pgRNA. Target DNA is then added to the reaction and incubated for 15 minutes. Then, proteinase K is added to detach the Cas9 RNP from the target DNA. Finally, the dsDNA products are analyzed via gel electrophoresis. B, Agarose gels showing cleavage of dsDNA containing the target sequence. Bottom gel is annotated version of top gel, highlighting locations of uncleaved and cleaved dsDNA bands. Asterisks indicate rough estimate of cleavage efficiency.

Next, we sought to validate whether these Cas9 RNPs were functional in cells. To test dsDNA cleavage and repair in cells, we wanted to develop a cleavage assay that was relatively fast and high throughput, enabling us to analyze and iterate through many pgRNA designs. We decided to use fluorescent proteins as a readout, which would allow to leverage the microscopy expertise in lab complemented with bulk data from flow cytometry. We designed a fluorescent cleavage reporter encoding mCherry linked to an out-of-frame GFP (Fig. 8a), inspired by a similar reporter designed for bacteria⁶⁵. The linker region is the target of the gRNA and requires either a 1nt deletion (+1 frameshift, or fs+1), a 1nt insertion (-1 frameshift, or fs-1) or a 1nt substitution (stop) to remove a premature stop codon and bring the GFP in frame. We stably expressed the three different fluorescent cleavage reporters in MDCK cells, then electroporated Cas9 RNPs to validate the performance of the

reporter (Fig. 8b, left panel). We found that all three cleavage reporters worked, but that the reporter requiring a 1nt insertion yielded the highest editing and repair rate (Fig. 8b, right panel). For all future experiments, we used the fs-1 cleavage reporter.

With the validated cleavage reporter cell line, we tested the functionality of the Cas9 RNPs formed with the best pgRNAs from the *in vitro* testing (Fig. 8c, left panel). We electroporated Cas9 RNPs into MDCK fs-1 cleavage reporter cells and compared cleavage and repair events identified via flow cytometry and amplicon sequencing. We found that the flow cytometry and sequencing results tracked well, reinforcing the use of the fluorescent cleavage reporter as a proxy for tracking dsDNA cleavage events (Fig. 8c, right panel). The data also shows that some pgRNAs performed better than the control (HEK3 gRNA scaffold). This difference could be due to variations in pgRNA concentration when forming the RNPs, since the in-house synthesized pgRNAs were quantified on a nanodrop. However, the Cas9 protein was the limiting reagent during formulation, so even if the pgRNA concentration was lower than expected, it shouldn't have had a large impact on the number of formed RNPs. Regardless of these nuances, the data demonstrate that all pgRNAs tested mediated dsDNA cleavage *in vivo* when electroporated as Cas9 RNPs into cells.

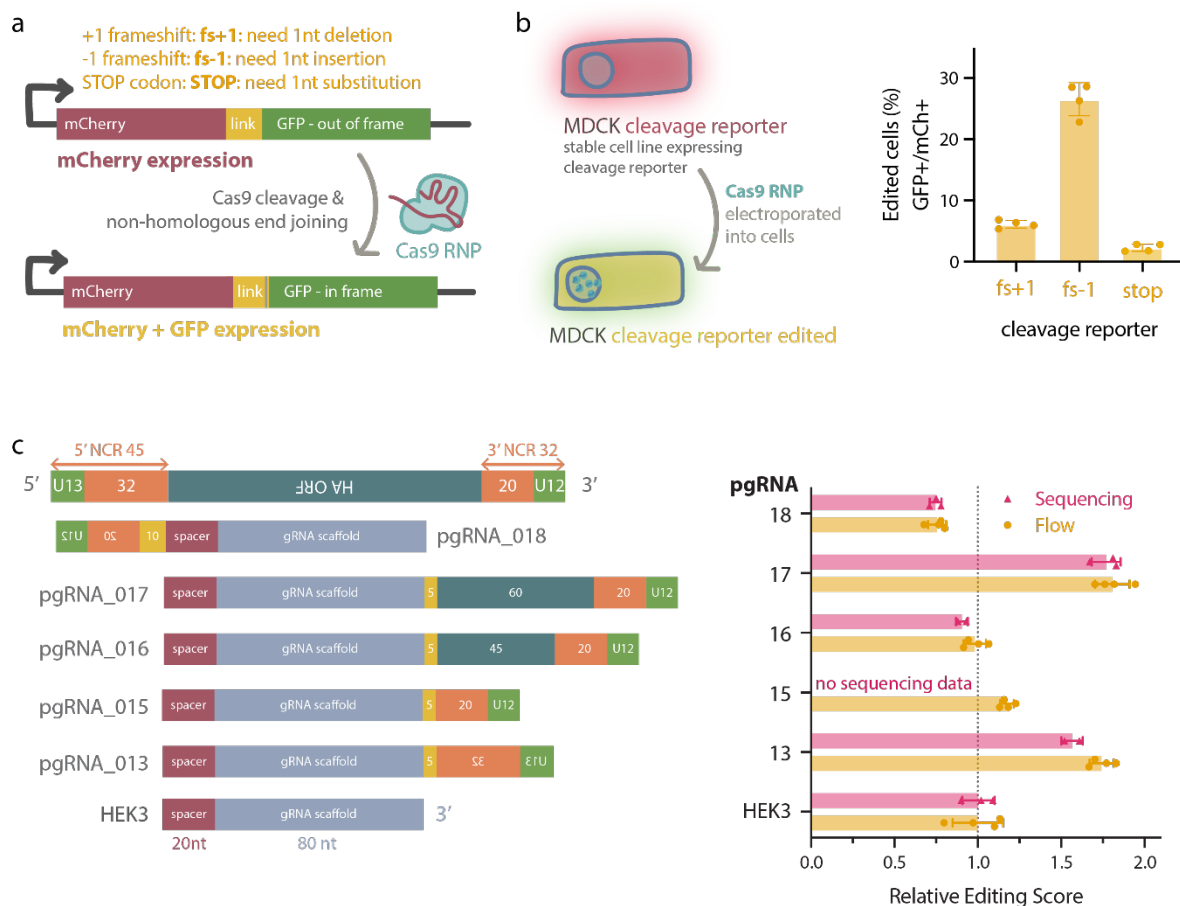


Figure 8. Design and validation of an *in vivo* reporter of dsDNA cleavage and repair.

A, Schematic detailing fluorescent cleavage reporter containing mCherry linked to a downstream GFP that is out of frame. The linker region (yellow) contains the target sequence for the Cas9 RNP. When the linker is cleaved and repaired via non-homologous end joining, the GFP may come into frame, depending on the repair made. A reporter construct is designed to reflect a 1nt insertion, deletion, or substitution. B, The fluorescent cleavage reporter is stably expressed in MDCK II cells. Cas9 RNPs are then electroporated into the

cells to validate the reporter. Four biological replicates are plotted for each cleavage reporter with mean and standard deviation. C, Schematic of pgRNA constructs tested for *in vivo* editing of cells. Cas9 RNPs formed with pgRNAs are electroporated into MDCK II cells expressing the fs-1 reporter. Samples are then processed on flow cytometry (yellow bars) and with amplicon sequencing (pink bars) to validate editing events. Data is normalized to control (HEK3, wt gRNA) and plotted as a “relative editing score”. Each condition contains four biological replicates. Mean and standard error is plotted for each condition.

CasIDVs mediate gene editing *in vivo*

Finally, we sought to demonstrate that the pgRNA-Cas9 complex could be packaged into IAV VLPs, forming CasIDVs, and mediate editing in cells. To synthesize CasIDVs, we transfected a cocktail of ten plasmids into HEK 293T cells; eight plasmids for the IAV proteins, one for the pgRNA, and one for the Cas9 protein. We then collected the CasIDVs and IAV VLPs in the supernatant and dosed that mixture onto MDCK cleavage reporter cells (Fig. 9, left panel). Here we show representative images on MDCK cleavage reporter cells expressing both mCherry and GFP, suggesting that a cleavage and repair event occurred (Fig. 9, right panel). All pgRNAs tested resulted in cleavage and repair events, though rare. This low rate of editing is likely due in part to the low titer of CasIDVs. Future work will include immunofluorescence staining for IAV proteins to identify which subset of cells were infected with a CasIDV or IAV VLP. This staining will allow us to normalize the number of edited cells relative to the number of infected cells, controlling for variability in CasIDV production.

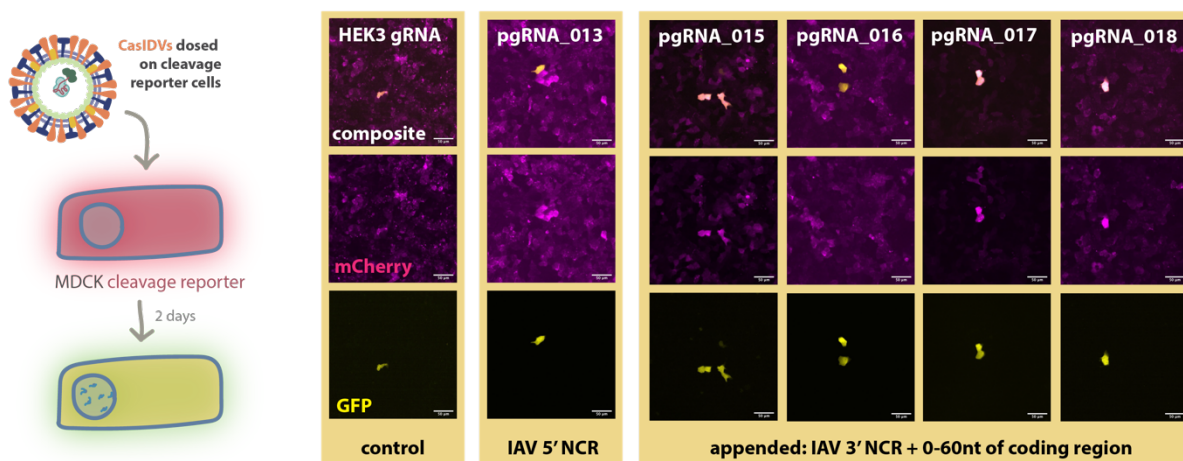


Figure 9. CasIDVs edit cells.

CasIDVs are dosed on MDCK II cells expressing the fs-1 cleavage reporter. 48 hours after dosing, cells were analyzed with confocal microscopy. Representative images are shown for each pgRNA tested. Scale bars are 50um.

Discussion

Our preliminary *in vitro* and *in vivo* live cell studies have validated key principles underlying the CasIDV project. First, we have shown that appending Influenza A vRNA packaging signals to Cas9 gRNA does not inhibit specific dsDNA cleavage *in vitro* (Fig. 7). Second, we have shown that these chimeric packaging gRNAs mediate cleavage *in vivo* when electroporated as Cas9 RNPs into cells (Fig. 8c). Finally, we have produced initial evidence that Cas9 RNPs packaged in Influenza-based Delivery Vehicles (CasIDVs) are delivered to cells and mediate cleavage (Fig. 9). Optimization and improvements in efficiency are needed for all steps, but the preliminary work to date is promising.

The CasIDV project is complex; many components must come together for the system to work, which requires lots of iterations. I have developed a pipeline to guide the process of rapid pgRNA design and validation (Fig. 10). One step that I haven't mentioned yet is quantifying packaging of pgRNA inside VLPs. I have developed RT-qPCR primers for this step but have been limited by the low VLP yield. It might be useful to consider alternative methods for detection, such as Cas13.

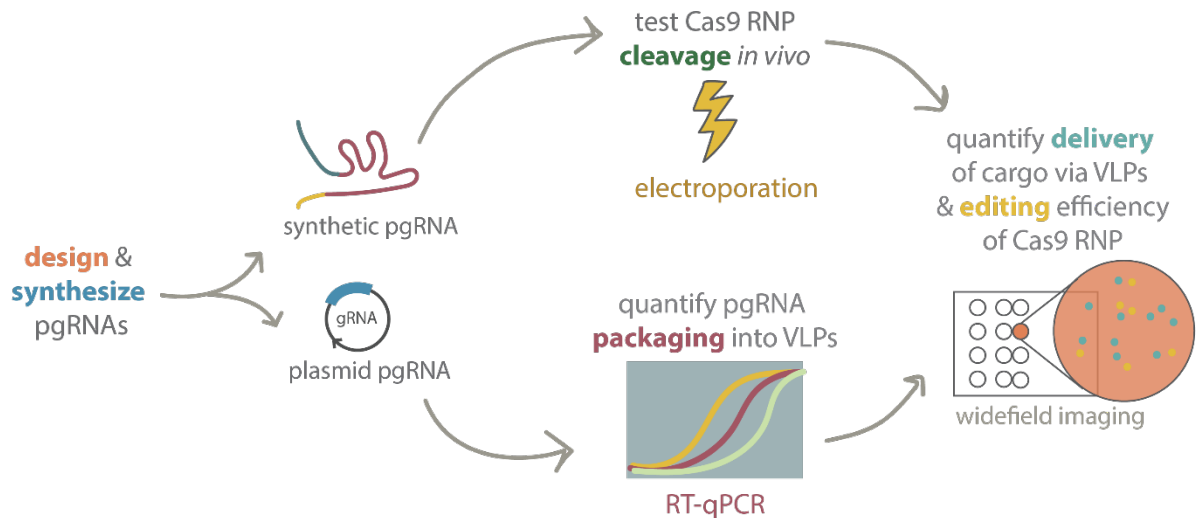


Figure 10. Proposed pipeline for quickly iterating through pgRNA designs.

Once pgRNAs are designed, they should be *in vitro* transcribed for testing via electroporation. They should also be cloned into a plasmid to validate packaging inside CasIDVs. Finally, CasIDVs should be dosed on cells and editing quantified with widefield imaging, or potentially flow cytometry if the editing events are more frequent.

One key step to improve the development of the CasIDV technology is to optimize the synthesis of CasIDVs. The current synthesis method is to transfect all ten plasmids (eight IAV, one gRNA, and one Cas9 plasmid) into HEK 293T cells. This is inefficient and highly variable from well-to-well. One improvement would be to transfect HEK cells that stably express Cas9. I did attempt this approach, but I found that the HEK-Cas9 cells generated an abysmally low viral titer, even when making wild type IAV with or without an MDCK co-culture (data not shown). Cas9 plasmids are notoriously tricky to transfect, so troubleshooting IAV synthesis with HEK-Cas9 cells would be a worthwhile endeavor. Alternatively, it may be helpful to consider other Cas9 homologs that have shorter sequences. The next big help would be to infect cells with IAV VLPs that are missing one vRNA segment. Having a validated stock of IAV VLPs with a known titer would eliminate seven plasmids from the transfection cocktail. The plasmid encoding the viral protein whose vRNA is missing would still need to be included. Ideally the CasIDV factory cells would stably express this missing protein; my attempts to achieve that failed. I made an MDCK cell line expressing HA but found that the cells quickly silenced expression of the viral protein. It could be useful to make another attempt at exogenous expression using a lower-expressing promoter.

For the next generation of testing, I have developed a new cleavage reporter that should increase the ease of the assay. I swapped out mCherry and GFP for mScarlett and mEmerald, which both have a higher quantum yield and better spectral separation. I appended them to an H2B motif, which traffics the reporter to the nucleus, effectively concentrating the signal. Finally, I switched to a UbC promoter to prevent the cleavage cell line from silencing expression of the cleavage reporter. In addition to this new reporter cell line, which I have stably expressed in HEK 293T and MDCK cells, I recommend

staining cells for Hoechst to identify the nucleus and an IAV nucleoprotein (NP) antibody to identify which cells have been infected. All the signals will be localized in the nucleus, making automated image analysis easy with tools like CellPose.

Despite the challenges, I still believe that developing CasIDVs is a worthwhile goal. Using a delivery vehicle that retains the biophysical properties of Influenza A Virus will enable us to target cells that are hard to access, like mucous-coated lung epithelial cells: CasIDVs could help edit the lung epithelium of patients suffering from cystic fibrosis. They may also be able to access cells in the digestive tract, which are also protected by a mucosal layer, suggesting delivery of cargo to mucosal immune cells, for example. Ultimately, I hope the CasIDV system is developed a flexible delivery platform that can help shuttle therapeutic cargo to hard-to-reach cells in need.

Methods and Materials

Mammalian cell lines and culture conditions

Lenti-X HEK 293T, HEK 293T, and MDCK II cells were acquired from the UC Berkeley Cell Culture Facility. Cells were maintained at 37°C, 5% CO₂ in DMEM (Corning) supplemented with 10% fetal bovine serum (Gibco) and 100U/mL penicillin-streptomycin (Gibco). Cells were confirmed mycoplasma free with a MycoAlert mycoplasma detection kit (Lonza).

Synthesis of Influenza A virus, VLPs, and CasIDVs

To generate new Influenza A virions, VLPs, and CasIDVs, we used a set of reverse genetics plasmids: ambisense pIAV plasmids for expression of vRNA and viral mRNA and pcDNA plasmids for expression of viral mRNA only, as previously described. We then transfected either a 70% confluent monoculture of HEK 293T cells or a co-culture of near-confluent HEK 293T and MDCK cells in a 6-, 12-, or 24-well plate 18 to 24 hours after plating. For a 12-well plate, we used 400ng of each plasmid complexed with 3uL of TransIT 293T (Mirus) per 1ug of total plasmid in 200uL OptiMEM. This ratio was scaled according to the surface area of the well plate. To make wild type virus, we used all eight pIAV plasmids. To make VLPs, we replaced select pIAV plasmids with their equivalent pcDNA plasmid. To make CasIDVs, we also included plasmids for Cas9 and pgRNA. 24 hours post transfection, we exchanged the media for virus growth media (MEM, 0.25% BSA, 1% penicillin/streptomycin, and 1mg/mL TPCK-treated trypsin). For purposes where we didn't want to activate the virions or VLPs, we omitted the TPCK trypsin. 48 to 72 hours post transfection, we collected supernatant containing virions/VLPs/CasIDVs and spun the supernatant for 10 minutes at 2000G to remove cellular debris. We saved the top 95% of the supernatant, supplemented with TPCK trypsin (0.5ug/uL final concentration) to activate all virions/VLPs, and then dosed on cells. We allow virus to bind for one hour, with gentle rocking every 15 minutes, before removing unbound virus and adding virus growth media supplemented with TPCK trypsin.

- To expand viral populations, we dosed fully infectious virions on confluent monolayers of MDCK II cells. We let virions grow for 24-48 hours before collecting and quantifying viral titer.
- To assay editing with CasIDVs, we dosed CasIDV supernatant on 96-well plates containing confluent monolayers of MDCK cells expressing a fluorescent cleavage reporter. Unbound CasIDVs were not washed off cells. We quantified editing 48 hours post infection by microscopy, flow cytometry, and/or amplicon sequencing.
- To quantify viral titer, we performed a plaque assay by dosing serial dilutions of viral supernatant on a 6- or 12-well plate. After washing off unbound virions, we added virus growth media containing microcrystalline cellulose to stall intercellular viral movement. After 48 hours, we washed the wells and stained with crystal violet to quantify viral titer.

In vitro cleavage assay

To assay the functionality of Cas9 RNPs formulated with pgRNAs, we tested cleavage *in vitro*. Briefly, Cas9 (QB3 MacroLabs) was incubated with *in vitro* transcribed gRNA or pgRNA (NEB HiScribe T7 RNA synthesis kit) at a ratio of roughly 1:1 for 10 minutes at 25°C (30nM final concentration of RNP). Then, 3nM target DNA was added to the RNP mixture and incubated for 15 minutes at 37°C. Finally, Proteinase K was added to dislodge Cas9 bound to the target DNA, then the cleaved DNA was run through an agarose gel to analyze cleavage.

***In vivo* cleavage assay – electroporation**

To assay the functionality of Cas9 RNPs formulated with pgRNAs *in vivo*, we electroporated Cas9 RNPs into MDCK II cells expressing the fluorescent cleavage reporter. The cleavage reporter contains an mCherry linked to an out-of-frame GFP under a CMV promoter. The linker region contains the DNA the pgRNA targets. Depending on how the DNA is cleaved and repaired, the GFP cassette will come into frame, resulting in a cell that expresses both mCherry and GFP. A new version of this reporter contains mScarlett instead of mCherry, mEmerald instead of GFP, a UbC instead of CMV promoter, and an H2B motif to traffic the fused fluorescent proteins to the nucleus.

We used a Neon transfection system (Invitrogen) to electroporate Cas9 RNPs into MDCK cleavage reporter cells according to the manufacturer's recommended protocol. Briefly, we formulated the RNPs by incubating Cas9-NLS (QB3 MacroLabs) with *in vitro* transcribed pgRNA (NEB HiScribe T7 RNA synthesis kit) at a 1:1 ratio. Cells were combined with the RNP (1 pmol of RNP for 2e5 cells) and electroporated in a 10uL tip at 1650 V for 20 ms. Cells were then plated in a 24-well plate and incubated for 48 hours before analysis by microscopy, flow cytometry, and/or amplicon sequencing.

Acknowledgements

We thank members of the Fletcher Laboratory, past and present, for support throughout this project. We especially thank Ari Joffe for his care in the early stages of the project, Emily Suter for her eagerness to learn about the project, and Caitlin Cornell for being bold enough to embark on her own Influenza project, sharing her hard-learned knowledge along the way. We thank Mike Vahey for laying the groundwork for IAV manipulation in the lab, along with Eva Schmid and Sungmin Son for initial brainstorming sessions. We thank María Díaz de León Derby and Liya Oster for their constant support. Finally, we thank our financial supporters for believing in this project: the NSF GRFP, Laboratory for Genomics Research (LGR) Innovation Award Program, and the Chan Zuckerberg Biohub. D.A.F. is a Chan Zuckerberg Biohub Investigator.

Author contributions

A.I.M and D.A.F. conceived and designed the study. A.I.M. performed experiments and analyzed and interpreted data, with help from R.W. and A.A.. D.A.F. supervised the study design and data collection. D.A.F. secured funding. A.I.M. wrote the preliminary draft with input from D.A.F.

Chapter 5 – Future directions

I have presented two distinct projects centered around the theme of creatively engineering biology. With the Cas13 split guide system, I started with a module evolved in bacteria as a defense against invading viruses. I repurposed that system into a strategy to detect small RNAs, which has implications for the molecular diagnostics and RNA biology fields. With the CasIDV project, I started with a viral pathogen – Influenza A Virus – and re-engineered it into a delivery vehicle to transport therapeutic cargo to cells. Nature is truly an awesome source of inspiration and materials for engineers to build new technologies.

While both technologies I present are promising, their future and utility relies on further optimization. Once the systems have been optimized, the door opens to new and exciting applications.

Cas13a split guide system

The Cas13a split guide system is a promising strategy for detecting small RNAs, but with further technological development, the approach can become an even more useful tool for both the molecular diagnostics and RNA biology fields. I will first discuss proposed technical improvements to the assay, then I will focus on improvements for different use cases.

Improving and validating sensitivity of the Cas13a split guide assay

To increase the utility of the Cas13a split guide system for the molecular diagnostics and RNA biology fields, the baseline sensitivity of the assay needs to be improved.

Thus far, we have demonstrated that the Cas13a split guide system has a sensitivity on the femtomolar level for a synthetic 10nt actRNA. Ideally, we would be able to robustly detect target RNA at even lower concentrations and in more complex matrices.

First, I would focus on characterizing and improving the limit of detection (LOD) of synthetic actRNA across a range of relevant lengths. Small RNAs ranging from 10 to 26nt are generally of interest, though it may be prudent to focus on the smaller end of that scale (<21nt) as other research groups have developed strategies to detect longer miRNAs (~23nt)^{32–35}. It would be interesting, however, to compare the detection of RNAs ranging from 21 to 26nt in the split guide system vs the full-length guide system; it may be that the split guide system generates a better signal than the full-length guide system.

I have suggested strategies to improve the assay sensitivity in Chapter 3, which I will briefly recap here. The main challenge is to maximize the signal-to-background ratio, which I believe can be achieved by careful design of the 3' flank and anchor regions of the capture RNA. I have already designed and ordered a few capRNAs to explore the importance of the length and sequence of the 3' flank (Fig. 1). To first explore the importance of sequence, cr4_t36 and cr4_cap11 should be compared with cr4_t36_b and cr4_cap11_a, which have swapped 3' flanks. Then, to begin perturbing the 3' flank length, cr4_t36 should be compared with cr4_t36_d, e, and f, which have decreasing length of the 3' flank. All these capRNAs should be tested with a constant 10nt actRNA and capRNA concentration

of 25nM to highlight any changes in the background signal. Learnings from those experiments may be applied to other capRNAs with stronger anchor regions, such as the cr2 anchor region.

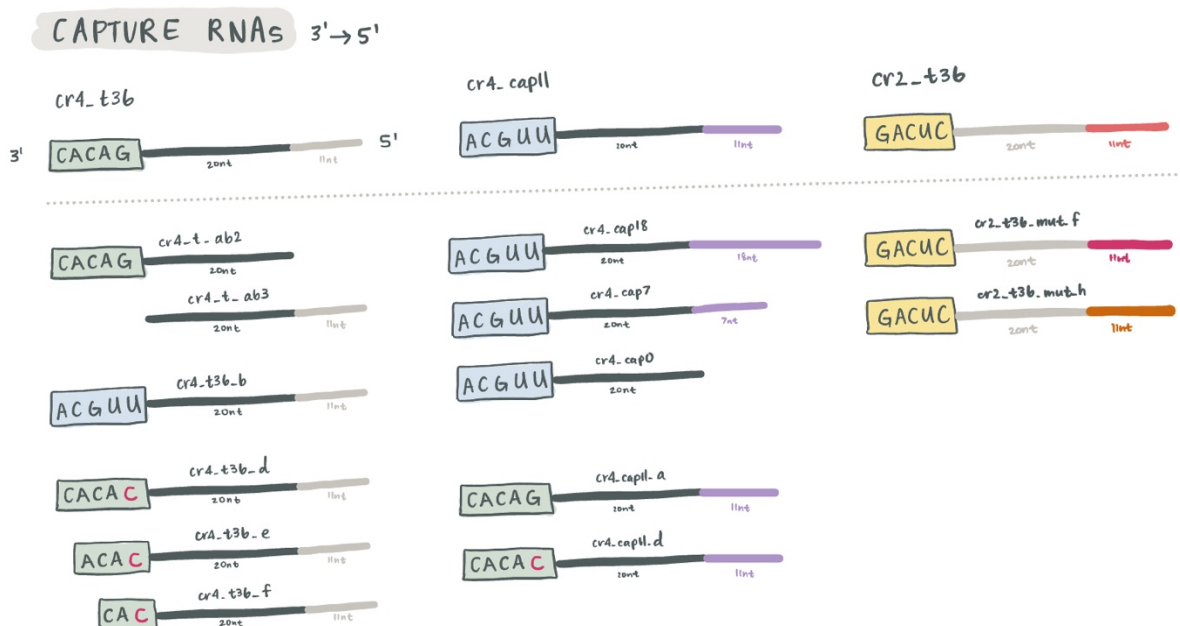


Figure 1. Schematic of synthesize capRNAs currently available in lab.

See Chapter 3 for more details about strategies to optimize background-to-signal. Beyond sequence-specific modifications to the assay, some structural modifications may help. To increase the assay speed and sensitivity, we could switch from a linear to a hairpin ssRNA reporter molecule, like the DNA hairpin reporters used for Cas12a DNA-detection assays⁶⁶. To further increase sensitivity, we could adopt a droplet-based approach, which we have previously demonstrated in our lab for conventional Cas13a assays²⁸.

In parallel, I would optimize sample preparation for the Cas13a assay. This work, while not that exciting, will benefit most of the Cas13 projects. I would focus on blood serum and saliva samples for miRNA diagnostics, and cell lysate and cell culture supernatant (a source of exosomes) for RNA biology applications.

Finally, I would benchmark the Cas13a split guide assay against RT-qPCR, the field standard apart from RNA sequencing, for a validated miRNA target with published primer and probe designs. Ideally, this benchmarking would demonstrate an increased sensitivity of the Cas13a split guide assay compared to RT-qPCR. Demonstrating the LOD of the Cas13a split guide system across a range of RNA lengths, in a range of samples, and benchmarked against RT-qPCR would make for a short and sweet methods paper.

Improving specificity of the Cas13a split guide system

The utility of the Cas13a split guide system also hinges on its specificity for target vs off-target RNA. In Chapter 2, I demonstrate a specificity to 2nt consecutive mismatches that varies with the length of the capture region of the capRNA (Fig. 2a). I also demonstrate moderate specificity for target RNA

that is misaligned with the capture region, resulting in gaps or overlaps between the split gRNA and the actRNA (Fig. 2b and c).

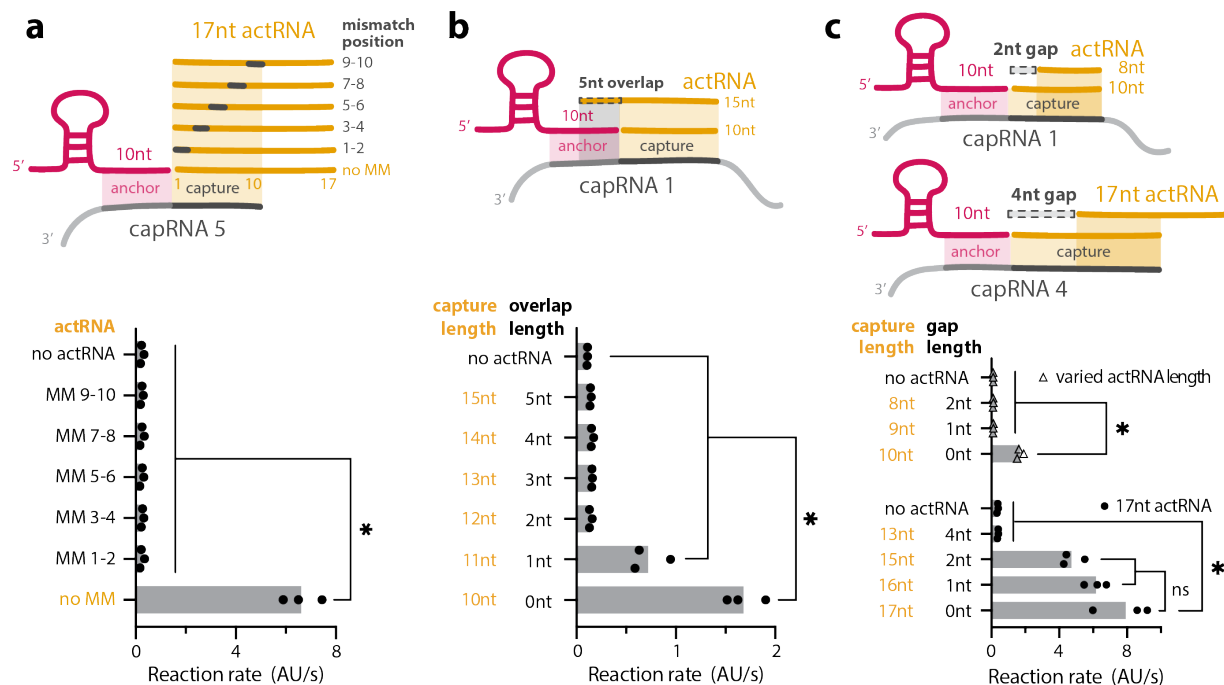


Figure 2. Specificity of Cas13a split guide system.

Reproduction of Ch. 2, Fig. 3, highlighting specificity of the Cas13a split guide system.

This high-to-moderate specificity, which is comparable to the Cas13a full-length gRNA system³⁹, needs to be improved for robust miRNA detection in molecular diagnostics or RNA biology applications. MicroRNAs, while generally referred to in their canonical state, exist in a variety of isoforms³⁸. Distinguishing these isoforms is critical, as miRNA functionality varies depending on whether and how the canonical miRNA is modified.

MiRNAs exhibit a variety of modifications and mutations. These modified miRNAs are generally referred to as miRNA isoforms, or isomiRs³⁸. Canonical miRNAs can be altered on the 5' or 3' end with added or removed nucleotides. These changes are considered templated if nucleotides are removed or non-templated if nucleotides are added. These non-templated modifications can be more difficult to detect, since the number and identity of the added nucleotides is unknown. MiRNA isoforms can also contain nucleotide mutations within the canonical sequences; this class of isomiRs is called polymorphic. The nucleotide mutations may be single nucleotide polymorphisms (SNPs) that are a product of mutations to the DNA or error in transcription to RNA. The mutations may also be caused by RNA editing events (A-to-I or C-to-U edits)³⁸. IsomiRdb is a good starting point for exploring the depth of miRNA isoforms⁶⁷. Finally, any miRNA may be subject to nucleotide modifications, which don't change the sequence of nucleotides, but rather change the structure of individual nucleotides⁹. Examples of RNA modifications include N6-methyladenosine (M6A), 2'-O-methylation (Nm), and 5-methylcytidine (M5C). These small RNA modifications and their functional impact are reviewed in a 2023 article⁹.

MiRNA modifications and mutations affect their function, which makes it important to distinguish the relative levels of different isomiRs for both diagnostic and RNA biology applications. SNPs in the seed region of a miRNA, for example, will affect what mRNA it targets for degradation. Thus, upregulation of the mutated miRNA has a different affect than upregulation of the canonical miRNA. When identifying unique miRNA signatures for different diagnostic applications, it is critical to know which specific isomiRs, rather than the canonical miRNA, are upregulated or downregulated. Using relative isomiR expression levels from small RNA sequencing data, researchers have been able to differentiate healthy from cancerous tissues and even identify different types of cancer⁶⁸. Easily and reliably detecting unique isomiRs will help improve the use of miRNAs for molecular diagnostics. RNA biologists will also benefit from reliable detection of isomiRs, especially ones containing nucleotide modifications, as RNA sequencing methods are currently unable to propagate the nucleotide modifications. In fact, the National Academies of Science, Engineering, and Medicine recently released a report calling for an international effort to develop standardized ways to sequence RNA modifications⁶⁹.

Improving the specificity of the split guide system will enable RNA biologists to ask and answer more questions. One area of interest is miRNA biogenesis. Mature miRNAs start out as primary miRNAs (pri-miRNAs) transcribed by RNA polymerase II into a series of stem-loop structures¹⁵. These stem-loops are then processed into individual stem-loop segments called precursor miRNAs (pre-miRNAs). The pre-miRNA is then cleaved into a miRNA duplex, with each piece referred to as the 5' or 3' strand and functionally as the guide or passenger strand¹⁵. Identifying the RNA species throughout the stages of miRNA biogenesis is critical to elucidating the regulation and dysregulation of miRNAs in healthy and diseased tissues. Differentiating the stages of miRNA biogenesis, however, is tricky, as the same sequences are present for the duration of the process. RT-qPCR primers and probes designed to amplify a mature miRNA would also amplify pre-miRNA and pri-miRNA. Likely more sets of primers and probes could be designed for pre-miRNA, excluding mature miRNA, and pri-miRNA, excluding pre- and mature miRNA, but such a system, if possible, becomes increasingly complicated. One can instead imagine a Cas13a-based approach to differentiate the stages of miRNA biogenesis that relies upon specificity to target length. Such an assay should first be developed to differentiate stages of miRNA in unpurified cell lysate before attempting to adapt it to work in cells. Translating the split guide assay into one that works in cells, like Liu et al have done for Cas12⁷⁰, would make the Cas13a split guide tool even more useful for RNA biologists.

Other applications of the Cas13a split guide system

The Cas13a split guide system is useful beyond the detection of human miRNAs and other small non-coding RNAs. For one, it can be applied to detecting small RNAs in plants. RNA segments as small as 10nt have been identified in plant exosomes⁷¹. In fact, these 10-17nt tiny RNAs (tyRNAs) were enriched in *Arabidopsis* exosomes⁷¹. The split guide system has already demonstrated sensitivity to target RNA as small as 8nt, making it an ideal new tool to detect the tyRNAs found in plant exosomes. Easier tools to analyze tyRNAs may help reveal the roles of these tyRNAs and suggest interesting applications in plant diagnostics.

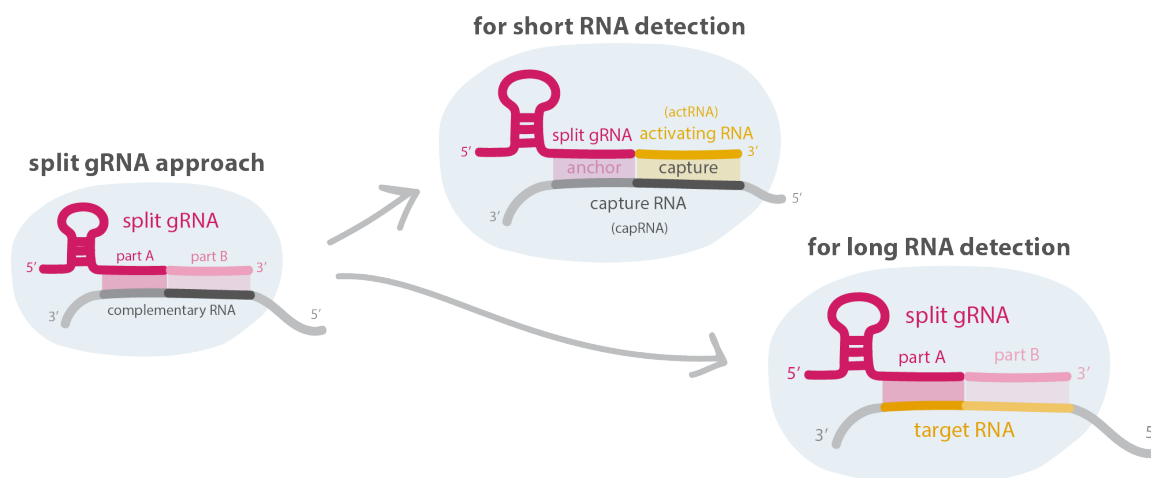


Figure 3. Use of the Cas13a split gRNA both for short and long RNA detection.

For short RNA detection, the activating RNA (yellow) is the limiting reagent. For the long RNA detection, the target RNA (yellow) is the limiting reagent.

Another use for the split guide system is as an alternative to the full-length guide system for detecting longer RNA targets (Fig. 3). Generally, we observe that the split guide system generates a lower signal than the full-length guide system, regardless of whether the actRNA or capRNA is the limiting reagent (Fig. 4a and b). For a fixed system with a split gRNA containing a 10nt anchor region, a 36nt capRNA, and 10nt actRNA, the signal is roughly 20% of the control for a limiting actRNA (1e6 copies/uL), and roughly 28% of the control for a limiting capRNA (1e6 copies/uL) (Fig. 4a and b, respectively). However, with a 20nt actRNA (1e6 copies/uL), we find that the split guide system generates a signal that is nearly 15x greater than the control (full-length gRNA activated by 36nt target) (Fig. 3c). If the capRNA was the limiting reagent, it is likely that the signal generated would be even greater, as it was with the 10nt actRNA. This behavior of the split guide system suggests that if the split gRNA and 20nt actRNA were added as reagents to the system, and the capRNA returns to being the target RNA (limiting reagent), the split guide approach would result in increased sensitivity to long RNA targets. This approach to long target RNA detection would help reduce background signal of the RNP. However, as with the conventional full-length guide system, a new gRNA and RNP would have to be made for each target RNA.

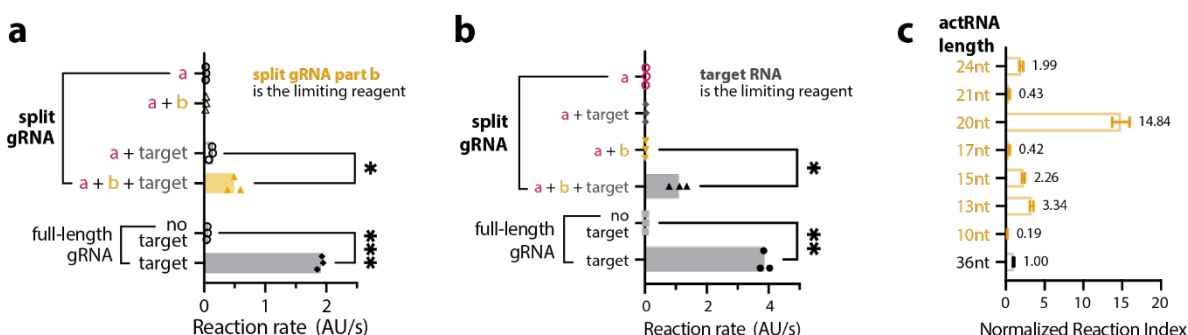


Figure 4. Data suggest the potential of a split gRNA for detection of long RNA targets.

A, Reproduced from Ch. 2, Fig. 1d, test of split gRNA approach, where split gRNA part b is the limiting reagent. B, Reproduced from Ch. 2, Extended Data Fig. 1c, test of split gRNA approach where target RNA is the limiting reagent. C, Demonstration of the split guide system detecting actRNA ranging from 10 to 24nt normalized relative to the full-length gRNA control, detecting a 36nt target.

Finally, I suggest the use of the split guide system as a strategy to amplify signal in an approach I term Cas26 (Fig. 5). The Cas26 system couples together two Cas13a reactions, where the cleavage product of the first reaction becomes the activating RNA of the second reaction. The first reaction can either be with a full-length guide system or a split guide system; the main requirement is that the fluorescent reporter, once cleaved, must be able to activate the second Cas13 reaction. This will likely best be achieved with a hairpin reporter, as the fragments of cleaved reporter should be at least 8nt long, though ideally 10nt or longer, to activate the split guide system. The second reaction is a split guide reaction, with a cleaved hairpin reporter from the first reaction serving as the actRNA for the signal-amplifying secondary reaction. The secondary Cas13a reaction will also cleave the same hairpin reporter, continuing the cascading amplification of signal. Alternatively, to differentiate the original hairpin reporter cleavage events from the amplified signal, the secondary Cas13a reaction can be performed using a Cas13a homolog that preferentially cleaves poly-A sequences instead of poly-U sequences, like LbaCas13a²⁵. The fluorescent reporter for the secondary reaction should have a different fluorophore from the hairpin reporter to separate the original from amplified signal. Using a Cas13a homolog would also prevent a cascading amplification of the signal, which may be useful for reducing false positive signals.

The success of Cas26 will depend upon the careful design of a hairpin reporter. The hairpin reporter must be easily cleaved and result in a consistent population of fragments. Those fragments must generate a robust signal in the secondary split guide reaction.

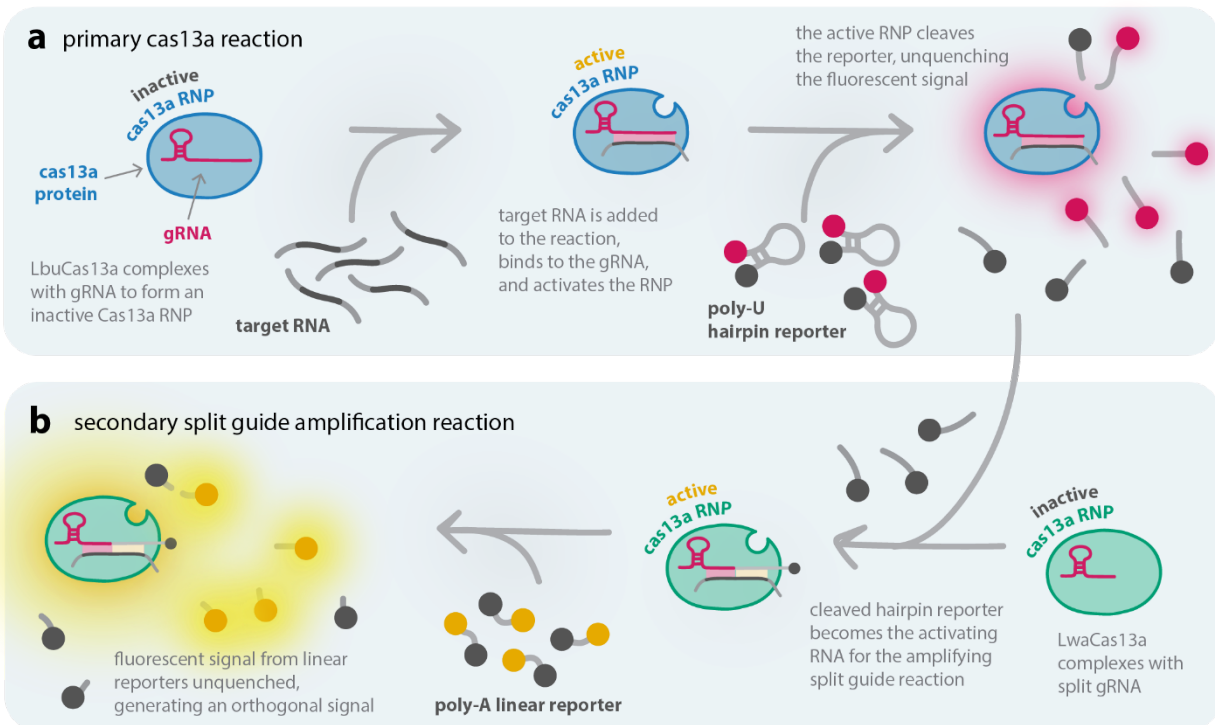


Figure 5. Schematic of Cas26 proposal.

In the Cas26 system, a Cas13a split guide system is coupled to another Cas13a reaction to amplify the signal. The product of the first Cas13a reaction, the cleaved hairpin reporter, becomes the reagent of the Cas13a split guide reaction.

CasIDVs

The CasIDV project is an ambitious and frustrating project. Perhaps I should've aborted the project early on, or pivoted it to something more feasible, but ultimately, I kept working on it for so long because I believed, and still do believe, that it can work. The CasIDV project embodies the style of science I like to pursue – creatively repurposing and re-engineering technologies that have evolved in the natural world for the benefit of human health. For this reason, I hope the foundation I have laid for this project will be built upon by the lab and that CasIDVs will be developed to robustly edit mucous-producing cells.

Improving CasIDV synthesis

One of the biggest struggles of the CasIDV project is consistently generating high titers of IAV VLPs. Part of that difficulty stems from the nature of the IAV reverse genetics system. The eight-plasmid system to synthesize fully infectious Influenza A virions is effective at what it was originally designed to do. For the researcher that needs to make a few different variants of the virion, they can synthesize those virions using a co-culture system, then expand and freeze the IAV stock. The CasIDV project does not have that luxury since it depends on the use of non-replicating IAV VLPs. Thus, for every condition of every experiment, 10+ plasmids need to be transfected into the well of cells. I have attempted to design clever strategies to overcome these limitations, such as exogenously expressing HA protein in MDCK cells for a co-culture approach. But I have failed to get such a system to work due to the cells silencing expression of the viral protein. One other approach that I would suggest is to replace a different vRNA segment that is less critical to the virus. Neuraminidase (NA), for example, could be replaced without sacrificing VLP formation or function. An IAV VLP without NA is still capable of infecting a cell; the neuraminidase can instead be supplemented in the media. The only other feasible option would be to replace the N segment, which encodes for accessory proteins that help suppress the cell's antiviral response. I have some pgRNAs designed to replace the NA segment that I've done preliminary testing with. To take that line of work further, I would recommend modifying those pgRNAs to include both vRNA UTRs on the 5' and 3' end of the gRNA. These UTRs are relatively short (13nt and 12nt, respectively) and shouldn't inhibit the cleavage function of the Cas9 RNP. By including both UTRs, the pgRNA can be replicated by the viral polymerase complex. This way, when a VLP packaging a pgRNA infects another cell, the pgRNA will be replicated, along with the rest of the viral genome, resulting in a next generation of VLPs packaging the pgRNA. With this strategy, it would be possible to make high titers of pgRNA-VLPs, then infect those into MDCK cells expressing Cas9. The particles that bud off the MDCK cells would have the opportunity to package both the Cas9 and the pgRNA. I have pgRNAs for the HA segment that contain both 5' and 3' UTRs, and have done some preliminary testing with them (data not shown). Solving the issue of robust CasIDV synthesis will make the rest of the testing that much easier, as editing events will be more frequent due to the higher titer.

CasIDV cargo

Once the issue of synthesis has been resolved, it opens opportunities for modifying what is packaged inside. We are currently attempting to replace one vRNA segment with a pgRNA, but ultimately, we may want to replace all eight vRNA segments with pgRNA to maximize the number of Cas9 RNPs per CasIDV. The first step toward achieving that would be to attempt to replace both the HA and NA segments, both of which we have multiple pgRNA options. To further extend the loading capacity, we could consider loading multiple Cas proteins per pgRNA by extending the pgRNA to

contain multiple gRNA scaffolds. This strategy would look like a “string” of RNA decorated with Cas9 “pearls”, with the end of the RNA string containing the critical IAV packaging sequence. In addition we may want to reserve some slots to package template DNA. A key characteristic of CRISPR-Cas9 gene therapy is the ability to excise existing DNA and introduce new DNA. We could design CasIDVs packaged with Cas9 RNPs and template DNA to insert a reporter gene into cells. The template DNA could be disguised as viral RNA or tethered to packaging RNA sequence. Introducing a template encoding a fluorescent protein would make quantification easy. Fluorescent protein expression could then be used as a proxy to compare delivery efficiency between Cas9 VLPs, electroporation, and AAV.

The next step would be to excise a segment of DNA and then knock-in a new segment of DNA. This could be achieved by building off the previous demonstration, but including RNPs that target both ends of the DNA segment to be excised. To make quantification easy with the systems we have set up in lab (microscopy and flow cytometry), we could excise a fluorescent protein reporter sequence and insert a different fluorescent reporter sequence.

CasIDVs don’t need to be limited to Cas9. In principle, the IDV can be loaded with any protein that binds RNA. We should consider packaging other CRISPR-Cas technologies, like base editors, to fully demonstrate the platform potential of the CasIDV system.

Packaging cargo via RNA-binding is not the only way to load the CasIDV. Other physical loading mechanisms should be considered. One strategy, similar to that used in the HIV-derived Cas9-EDVs developed by Hamilton et al^{49,50}, would be to make fusion proteins linking Cas9 to the M1 matrix protein of IAV. The M1 matrix protein coats the interior of an Influenza A virion and is critical for virion formation, especially formation of filamentous particles⁷². We have demonstrated that the M1 protein can be modified by adding a small tag for a fluorophore to bind, enabling labelling of virions⁶⁰. However, in this work, we found that the M1 protein is sensitive to modifications. Thus, it would be critical to find the optimal ratio between wildtype M1 and an M1-Cas9 fusion protein for successful synthesis of CasIDV particles.

Applications of CasIDV system

The CasIDV project is so exciting not only for its unique packaging strategy, but also for its potential applications. The natural targeting of Influenza A to airway cells and clearance by immune cells suggest two initial targets for the platform: developing CasIDVs that (1) correct genetic mutations in lung epithelial cells that cause cystic fibrosis (CF) and (2) introduce DNA encoding specific chimeric antigen receptors (CARs) into T-cells. CF is caused by a variety of mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene^{73,74}. To demonstrate the utility of the CasIDV system for treating CF, I would attempt to modify the CFTR gene in mucous-secreting cells, like Calu-3. I would synthesize CasIDVs that package Cas9 RNPs flanking a portion of the CFTR gene and a template DNA encoding a restricting enzyme. The introduction of a gene segment containing a restriction enzyme serves as a proxy for replacing mutated sequences and allows us to analyze knock-in efficiency by digesting CFTR PCR products. The editing efficiency could be benchmarked against that of AAV.

Another potential application is editing immune cells, a subpopulation of which reside in the mucosa. A motivating preliminary demonstration would be to introduce a new CAR to a T-cell. We could design CasIDVs with template RNA encoding a clinically useful CAR, such as CTL019 for B-cell

precursor acute lymphoblastic leukemia (ALL) and dose the CasIDVs on isolated T-cells. CAR expression could be benchmarked against electroporation.

Safety considerations for CasIDV system

The CasIDV project brings up questions of safety – rightfully so. Fortunately, there is an extensive literature around influenza vaccine development that can help with this problem. We expect the immune response to CasIDV delivery to be comparable to the annual IAV vaccine. However, to further limit the immune response, we can pseudotype the CasIDVs to express HA and NA subtypes that have already been well tolerated as part of previous seasonal influenza vaccinations. Another strategy to reduce the immune response is to reduce the dosage, which itself relies upon how well-loaded the CasIDV are with Cas9 RNPs.

References

1. Magalhães, F. *et al.* Regeneration of starfish radial nerve cord restores animal mobility and unveils a new coelomocyte population. *Cell Tissue Res* **394**, 293–308 (2023).
2. Algrain, M. *et al.* In the footsteps of sea stars: deciphering the catalogue of proteins involved in underwater temporary adhesion. *Open Biol* **12**, (2022).
3. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J Bacteriol* **127**, 1550–1557 (1976).
4. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (1979)* **337**, 816–821 (2012).
5. Dull, T. *et al.* A Third-Generation Lentivirus Vector with a Conditional Packaging System. *J Virol* **72**, 8463–8471 (1998).
6. Hung, S. S. C. *et al.* AAV-Mediated CRISPR/Cas Gene Editing of Retinal Cells In Vivo. *Investigative Ophthalmology & Visual Science* **57**, 3470 (2016).
7. Koonin, E. V, Makarova, K. S. & Koonin, E. V. Origins and evolution of CRISPR-Cas systems. (2019).
8. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (1979)* **353**, (2016).
9. Xiong, Q. & Zhang, Y. Small RNA modifications: regulatory molecules and potential applications. *J Hematol Oncol* **16**, (2023).
10. Li, Z. *et al.* Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. *Nucleic Acids Res* **40**, 6787–6799 (2012).
11. Park, M. S., Sim, G. Y., Kehling, A. C. & Nakanishi, K. Human Argonaute2 and Argonaute3 are catalytically activated by different lengths of guide RNA. *Proc Natl Acad Sci U S A* **117**, 28576–28578 (2020).
12. Nakanishi, K. Are Argonaute-Associated Tiny RNAs Junk, Inferior miRNAs, or a New Type of Functional RNAs? *Front Mol Biosci* **8**, 795356 (2021).
13. Sim, G. Y. *et al.* Manganese-dependent microRNA trimming by 3'→5' exonucleases generates 14-nucleotide or shorter tiny RNAs. *Proc Natl Acad Sci U S A* **119**, e2214335119 (2022).
14. Ho, P. T. B., Clark, I. M. & Le, L. T. T. MicroRNA-Based Diagnosis and Therapy. *Int J Mol Sci* **23**, 7167 (2022).
15. Saliminejad, K., Khorram Khorshid, H. R., Soleymani Fard, S. & Ghaffari, S. H. An overview of microRNAs: Biology, functions, therapeutics, and analysis methods. *J Cell Physiol* **234**, 5451–5465 (2019).
16. Coleman, D. & Kuwada, S. miRNA as a Biomarker for the Early Detection of Colorectal Cancer. *Genes (Basel)* **15**, 338 (2024).
17. Hrach, H. C. & Mangone, M. miRNA profiling for early detection and treatment of duchenne muscular dystrophy. *Int J Mol Sci* **20**, 4638 (2019).
18. Daoud, A. Z., Mulholland, E. J., Cole, G. & McCarthy, H. O. MicroRNAs in Pancreatic Cancer: Biomarkers, prognostic, and therapeutic modulators. *BMC Cancer* **19**, 1130 (2019).
19. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* **105**, 10513–10518 (2008).
20. Krepelkova, I. *et al.* Evaluation of miRNA detection methods for the analytical characteristics necessary for clinical utilization. *Biotechniques* **66**, 277–284 (2019).
21. Hindson, C. M. *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* **10**, 1003–1005 (2013).

22. Benesova, S., Kubista, M. & Valihrach, L. Small RNA-Sequencing: Approaches and Considerations for miRNA Analysis. *Diagnostics* **11**, 964 (2021).
23. Zhang, J. *et al.* Direct microRNA Sequencing Using Nanopore-Induced Phase-Shift Sequencing. *iScience* **23**, 100916 (2020).
24. Chandrasekaran, S. S. *et al.* Rapid detection of SARS-CoV-2 RNA in saliva via Cas13. *Nat Biomed Eng* **6**, 944–956 (2022).
25. East-Seletsky, A., O'connell, M. R., Burstein, D., Knott, G. J. & Doudna, J. A. RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. *Mol Cell* **66**, 373–383 (2017).
26. Fozouni, P. *et al.* Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell* **184**, 323–333.e9 (2021).
27. Fozouni, P. *et al.* Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. *Cell* **184**, 323–333 (2021).
28. Son, S. *et al.* Sensitive and multiplexed RNA detection with Cas13 droplets and kinetic barcoding. *medRxiv* (2021) doi:10.1101/2021.08.02.21261509.
29. Shinoda, H. *et al.* Amplification-free RNA detection with CRISPR–Cas13. *Commun Biol* **4**, 476 (2021).
30. Arizti-Sanz, J. *et al.* Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2. *Nat Commun* **11**, 5921 (2020).
31. Myhrvold, C. *et al.* Field-deployable viral diagnostics using CRISPR-Cas13. *Science (1979)* **360**, 444–448 (2018).
32. Shan, Y., Zhou, X., Huang, R. & Xing, D. High-Fidelity and Rapid Quantification of miRNA Combining crRNA Programmability and CRISPR/Cas13a trans-Cleavage Activity. *Anal Chem* **91**, 5278–5285 (2019).
33. Zhou, T. *et al.* CRISPR/Cas13a Powered Portable Electrochemiluminescence Chip for Ultrasensitive and Specific MiRNA Detection. *Advanced Science* **7**, 1903661 (2020).
34. Bruch, R. *et al.* CRISPR/Cas13a-Powered Electrochemical Microfluidic Biosensor for Nucleic Acid Amplification-Free miRNA Diagnostics. *Advanced Materials* **31**, 1905311 (2019).
35. Granados-Riveron, J. T. & Aquino-Jarquin, G. CRISPR/Cas13-Based Approaches for Ultrasensitive and Specific Detection of microRNAs. *Cells* **10**, 1655 (2021).
36. Chen, Y. *et al.* Foldback-crRNA-Enhanced CRISPR/Cas13a System (FCECas13a) Enables Direct Detection of Ultrashort sncRNA. *Anal Chem* **95**, 15606–15613 (2023).
37. Rananaware, S. R. *et al.* Programmable RNA detection with CRISPR-Cas12a. *Nat Commun* **14**, 5409 (2023).
38. Tomasello, L., Distefano, R., Nigita, G. & Croce, C. M. The MicroRNA Family Gets Wider: The IsomiRs Classification and Role. *Front Cell Dev Biol* **9**, 1–15 (2021).
39. Tambe, A., East-Seletsky, A., Knott, G. J., Doudna, J. A. & O'Connell, M. R. RNA-binding and HEPN-nuclease activation are decoupled in CRISPR-Cas13a. *Cell Rep* **24**, 1025–1036 (2018).
40. Jiao, H. *et al.* MicroRNA expression profiles from HEK293 cells expressing H5N1 avian influenza virus non-structural protein 1. *Innate Immun* **25**, 110–117 (2019).
41. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* **47**, D155–D162 (2019).
42. Fornace, M. E. *et al.* NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems. *ChemRxiv* (2022) doi:10.26434/CHEMRXIV-2022-XV98L.
43. Bravo, J. P. K. *et al.* RNA targeting unleashes indiscriminate nuclease activity of CRISPR–Cas12a2. *Nature* **613**, 582–587 (2023).
44. Armbruster, D. A. & Pry, T. Limit of Blank, Limit of Detection and Limit of Quantitation. *Clin Biochem Rev* **29**, S49–S52 (2008).

45. Liu, Y. *et al.* From RNA to DNA: CRISPR/LbuCas13a Demonstrates Exceptional Single-Nucleotide Specificity. (2024) doi:10.21203/RS.3.RS-3920513/V1.
46. Kurreck, J., Wyszko, E., Gillen, C. & Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res* **30**, 1911–1918 (2002).
47. Yang, W. *et al.* Progress of delivery methods for CRISPR-Cas9. *Expert Opin Drug Deliv* **19**, 913–926 (2022).
48. Mout, R., Ray, M., Lee, Y.-W., Scaletti, F. & Rotello, V. M. In Vivo Delivery of CRISPR/Cas9 for Therapeutic Gene Editing: Progress and Challenges. *Bioconjug Chem* **28**, 880–884 (2017).
49. Hamilton, J. R. *et al.* Targeted delivery of CRISPR-Cas9 and transgenes enables complex immune cell engineering. *Cell Rep* **35**, 109207 (2021).
50. Hamilton, J. R. *et al.* In vivo human T cell engineering with enveloped delivery vehicles. *Nat Biotechnol* (2024) doi:10.1038/s41587-023-02085-z.
51. Calder, L. J., Wasilewski, S., Berriman, J. A. & Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proc Natl Acad Sci U S A* **107**, 10685–10690 (2010).
52. Chlanda, P. *et al.* Structural Analysis of the Roles of Influenza A Virus Membrane-Associated Proteins in Assembly and Morphology. *J Virol* **89**, 8957–66 (2015).
53. Marsh, G. A., Hatami, R. & Palese, P. Specific Residues of the Influenza A Virus Hemagglutinin Viral RNA Are Important for Efficient Packaging into Budding Virions. *J Virol* **81**, 9727–9736 (2007).
54. Liang, Y., Hong, Y. & Parslow, T. G. cis-Acting Packaging Signals in the Influenza Virus PB1, PB2, and PA Genomic RNA Segments. *J Virol* **79**, 10348–10355 (2005).
55. Vahey, M. D. & Fletcher, D. A. Influenza A virus surface proteins are organized to help penetrate host mucus. *Elife* **8**, (2019).
56. Dou, D., Revol, R., Østbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Front Immunol* **9**, 1581 (2018).
57. Noda, T. Selective genome packaging mechanisms of influenza A viruses. *Cold Spring Harb Perspect Med* **11**, 1–14 (2021).
58. Hoffmann, E., Neumann, G., Hobom, G., Webster, R. G. & Kawaoka, Y. “Ambisense” Approach for the Generation of Influenza A Virus: vRNA and mRNA Synthesis from One Template. *Virology* **267**, 310–317 (2000).
59. Cronin, J., Zhang, X.-Y. & Reiser, J. Altering the Tropism of Lentiviral Vectors through Pseudotyping. *Curr Gene Ther* **5**, 387–398 (2005).
60. Vahey, M. D. & Fletcher, D. A. Low-Fidelity Assembly of Influenza A Virus Promotes Escape from Host Cells. *Cell* **176**, 281-294.e19 (2019).
61. Muramoto, Y. *et al.* Hierarchy among Viral RNA (vRNA) Segments in Their Role in vRNA Incorporation into Influenza A Virions. *J Virol* **80**, 2318–2325 (2006).
62. Gao, Q. *et al.* The Influenza A Virus PB2, PA, NP, and M Segments Play a Pivotal Role during Genome Packaging. *J Virol* **86**, 7043–7051 (2012).
63. Tomescu, A. I., Robb, N. C., Hengrung, N., Fodor, E. & Kapanidis, A. N. Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proceedings of the National Academy of Sciences* **111**, (2014).
64. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**, 935–949 (2014).
65. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
66. Rossetti, M. *et al.* Enhancement of CRISPR/Cas12a trans-cleavage activity using hairpin DNA reporters. *Nucleic Acids Res* **50**, 8377–8391 (2022).

67. Aparicio-Puerta, E. *et al.* isomiRdb: microRNA expression at isoform resolution. *Nucleic Acids Res* **51**, D179–D185 (2023).
68. Telonis, A. G. *et al.* Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res* **45**, 2973–2985 (2017).
69. Medicine, N. A. of S. E. and. *Charting a Future for Sequencing RNA and Its Modifications : A New Era for Biology and Medicine*. (The National Academies Press, Washington, DC, 2024). doi:10.17226/27165.
70. Liu, H. *et al.* Light-activated CRISPR-Cas12a for amplified imaging of microRNA in cell cycle phases at single-cell levels. *Sci Adv* **10**, eadp6166 (2024).
71. Baldrich, P. *et al.* Plant extracellular vesicles contain diverse small RNA species and are enriched in 10- to 17-nucleotide “Tiny” RNAs. *Plant Cell* **31**, 315–324 (2019).
72. Elleman, C. J. & Barclay, W. S. The M1 matrix protein controls the filamentous phenotype of influenza A virus. *Virology* **321**, 144–153 (2004).
73. Rafeeq, M. M. & Murad, H. A. S. Cystic fibrosis: current therapeutic targets and future approaches. *J Transl Med* **15**, 84 (2017).
74. Marangi, M. & Pistritto, G. Innovative Therapeutic Strategies for Cystic Fibrosis: Moving Forward to CRISPR Technique. *Front Pharmacol* **9**, (2018).