Title
Evaluating RNA-based gene regulation across different systems

Author
Kitto, Rebekah Zang

Publication Date
2019

Peer reviewed|Thesis/dissertation

Evaluating RNA-based gene regulation across different systems

by

Rebekah Z. Kitto

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ming C. Hammond, Co-chair
Professor Matthew B. Francis, Co-chair
Professor Evan W. Miller
Professor Donald C. Rio

Spring 2019

Abstract

Evaluating RNA-based gene regulation across different systems
by
Rebekah Z. Kitto

Doctor of Philosophy in Chemistry
University of California, Berkeley

Professor Ming C. Hammond, Co-Chair

Ribonucleic acid (RNA) is an essential biomolecule that is commonly recognized as the template for protein synthesis across different classes of life. Aside from this well-known function, RNA plays many diverse roles within the cell and is directly involved in several processes that control gene expression. Different forms of RNA, spanning a range of lengths and structures, interact with the cellular machinery either alone or in complex with proteins and other nucleic acids to regulate how and when genes are expressed. In the past sixty years, the RNA field has grown at an astounding rate, and our understanding of this pervasive molecule is ever-expanding. However, in addition to its function as a gene regulator, RNA can also be engineered for use as a tool to probe cellular processes. We present here our efforts towards employing RNA in three different roles – targeting device, gene repressor, and sensor – to study fundamental processes in bacteria and eukaryotes.

For my family –
past, present, and future

Table of Contents

1 RNA in gene regulation 1

1.1 Introduction 2

1.2 Outlook 4

1.3 Figures 6

2 Alternative mRNA splicing for gene regulation in eukaryotes 11

2.1 Introduction 12

2.2 Results 13

2.2.1 Efforts towards validation of a splicing-regulated inducible protein expression system 13

2.2.2 Efforts towards targeted delivery of dCas9 fusion proteins to influence alternative splicing 14

2.3 Discussion 17

2.3.1 Splicing-regulated protein expression 17

2.3.2 Targeted control of alternative splicing 18

2.4 Materials and Methods 18

2.5 Figures 21

2.6 Tables 35

3 Inducible RNAi systems in microgravity 46

3.1 Introduction 47

3.2 Results 48

3.3 Discussion 52

3.4 Materials and Methods 53

3.4.1 Materials 53

3.4.2 Hardware Design 54

3.4.3 Experimental Setup 54

3.5 Figures 57

3.6 Tables 69

4 Development of genetically encodable mRNA and sRNA sensors 70

4.1 Introduction 71

4.1.1 Biosensor engineering 71

4.2 Results 74

4.3 Discussion 79

4.4 Materials and Methods 80

4.5 Figures 83
4.6 Tables 101
5 References 110

Table of Figures

Figure 1.1 Central dogma of molecular biology. 6
Figure 1.2 miRNA and siRNA biogenesis and RNAi mechanism. 7
Figure 1.3 Mechanisms for riboswitch gene regulation 8
Figure 1.4 *Trans*-acting bacterial sRNAs. 9
Figure 1.5 CRISPR/Cas response pathway. 10
Figure 2.1 Key regulators and general mechanism of splicing. 21
Figure 2.2 *Hy*P5SM Schematic. 22
Figure 2.3 Illustration of leaky protein expression. 23
Figure 2.4 Cas9 RNA-guided cleavage of target DNA. 24
Figure 2.5 Two-component regulation of Cas9 expression. 25
Figure 2.6 Loque lab Cas9 activity visualization system. 26
Figure 2.7 Cas9 vs. Cas9-*Hy*P5SM performance 27
Figure 2.8 siRNA-mediated regulation of alternative splicing 28
Figure 2.9 dCas9-HMP fusions for gene regulation 29
Figure 2.10 Targeted dCas9 fusions for histone modification 30
Figure 2.11 Identifying dCas9-HMP targets through RNA-seq 31
Figure 2.12 Characterizing dCas9-HMP targets from RNA-seq library 32
Figure 2.13 Optimizing transfection conditions for dCas9-HMP testing 33
Figure 2.14 Testing dCas9-HMP fusions targeted to the UAP1 alternative exon 34
Figure 3.1 Development of the CubeSat plant growth system 57
Figure 3.2 Verifying pX7-AtPDSi inducible protein expression. 58
Figure 3.3 Liquid media accelerates growth but requires sterilization. 59
Figure 3.4 Plant growth optimization in Stage 2 design. 60
Figure 3.5 Plant growth optimization in Stage 3 design. 62
Figure 3.6 Timeline for Hydra1 experiment. 64
Figure 3.7 Plant growth in the ground control and flight units. 65
Figure 3.8 Verifying chemical induction through PCR and image analysis. 66
Figure 3.9 Hydra1 CubeSat unit on the ISS. 67
Figure 3.10 Pictures of plant growth hardware. 68
Figure 4.1 Spinach2 aptamer and RBF-biosensor general schematic. 83
Figure 4.2 Riboswitch-based fluorescent biosensor structures. 84
Figure 4.3 tRNA scaffold schematic. 85
Figure 4.4 Testing tRNA scaffolds with RBF sensors and aptamers. 86
Figure 4.5 tRNA scaffold effects on RNA levels. 87
Figure 4.6 tRNA scaffold performance *in vitro*. 88
Figure 4.7 Imaging mRNA splicing patterns with RBF sensors 89

Figure 4.8 FASTmiR and SR-21 miRNA sensors. 90
Figure 4.9 Design criteria for FAST sensors 92
Figure 4.10 Initial screening of FASTmRNA designs. 93
Figure 4.11 Characterizing FASTmRNA sensors *in vitro*. 94
Figure 4.12 FASTmRNA sensor 21IU *in vivo* testing. 95
Figure 4.13 FASTsRNA sensor application. 96
Figure 4.14 Screening FASTsRNA designs *in vitro*. 97
Figure 4.15 Characterizing FASTsRNA sensor hits. 98
Figure 4.16 FASTsRNA sensor performance *in vivo* under different expression systems. 99
Figure 4.17 Alternate strategies for improved FASTsRNA signal. 100

Table of Tables

Table 2.1: Construct sequences for Cas9-*Hy*P5SM project 35
Table 2.2: Construct sequences for dCas9-HMP project 39
Table 2.3: UAP1 target exon /flanking intron sequence and sgRNAs 45
Table 3.1: Materials tested for growth apparatus. 69
Table 3.2: Filter paper and aluminum mesh dimensions 69
Table 4.1: Promoter, tRNA Scaffold, and RBF biosensor sequences. 101
Table 4.2: FASTmRNA Sequences. 102
Table 4.3: FASTsRNA sequences. 103
Table 4.4: Free energy and predicted folding frequency of FASTsRNA sensors 105
Table 4.5: Oligonucleotide library for FASTmRNA and SRB-2 sensor designs 106
Table 4.6: Oligonucleotide library for FASTsRNA designs 108

Acknowledgments

I never thought that I would end up in graduate school – not even after growing up with two PhD-wielding parents and enrolling in three years of undergraduate chemistry research. Perhaps this was because I was in denial, or maybe it was the fact that my mother still, to this day, insists I try taking the MCAT to "see how I would do," but either way it was one of the best decisions of my life. My PhD may have been laden with countless failed experiments and a number of exceedingly embarrassing research presentations, but this experience has allowed me to grow and mature in ways I could not have foreseen as a fresh-faced college grad. Of course, that is not to say that this five-year journey has been an easy one – and it certainly would not have been an enjoyable one without the help of my family and friends.

First and foremost, I would like to thank my parents for continuously supporting me mentally, emotionally, and (thank you, Bay Area housing bubble) sometimes even financially throughout graduate school. I love you very much and I appreciate that you listen to me rant about my research even if you don't know what I'm talking about half the time. If it makes you feel any better, I tried to read your theses and realized I am no longer mentally capable of doing higher-level math. You have taught me to be optimistic and try to make the best out of any situation – and, apparently, that I should have been an M.D. like Alex. As for Alex, I know that you are busy sawing off people's limbs and stitching them back on again so I understand that we haven't been able to talk much during grad school. However, I still enjoy our yearly visits and am looking forward to living near you again this summer. We both know that you're not going to stop playing video games any time in the near future. I should also thank my Aunt Sherry for listening to me rant about both science and academic politics in the department – and for teaching me that women still need to fight for what they want in today's society. I love you all very much, Kittos.

I would also like to thank the friends that I have made in Berkeley over the years. Nothing makes a long day of cloning more enjoyable than a quick coffee break and snek and I will always remember the excellent movie nights Rachel, Helen, and I had as a trio (triangle?). I hope that we can all stay in touch after school and reunite for future movie nights one day.

As much fun as I have had outside of work, I have also made some good friends within the lab and have to thank the many group alumni – Tania, Cindy, Zach, and Yichi – for making the Hammond lab such an enjoyable place to do science. And, of course, three members of the fearsome foursome: Todd, Johnny, and Andrew. It's been a fun and interesting five years and I hope we can reunite again in the future (also, I'll never forget that one game of HoTS in lab). I also have to thank my PI, Ming, for being an excellent boss and mentor throughout my time in grad school. You stood by me at some of my lowest points and still managed to keep me excited about my projects – no easy task given how many I've cycled through! It has been inspiring to watch your scientific star keep rising and I hope that you have a long and illustrious career in research.

Finally, I would like to thank my fiancé, Alec. You are my rock and I could not have done any of this without you. Who would have thought that when we met four and a half years ago, we'd be together today and headed to the east coast to pursue our dreams? Certainly not me (no offense). Nothing makes a long day at work more worthwhile than coming home to see you and Migos on the couch waiting for me. You are the love of my life and I look forward to the future that you envisioned for us – full of food and animals.

1 RNA in gene regulation

1.1 Introduction

Sixty years ago, the central dogma of molecular biology upheld the belief that "DNA makes RNA, RNA makes proteins, and proteins make us."[1] To the average person, this statement still holds true; in a living organism, DNA is the master code, protein the workhorse, and RNA some ambiguous element that links the two together. In reality, RNA plays a number of roles within the cell, not only as a messenger to transform DNA into functional protein, but also as a prominent regulator of gene expression. Over the years, more and more non-coding RNAs have been uncovered that expand the role of RNA in the long-standing central dogma. These sequences are diverse in length, fold, and function and are essential to normal cell processes in all domains of life. This work focuses on eukaryotic and bacterial RNAs involved in gene regulation, either studied in their native context or adapted for synthetic biology applications.

Eukaryotic regulatory RNAs

Following the development of the central dogma, there were a number of studies that suggested large portions of cellular RNA were being transcribed but not translated into protein.[2–4] In eukaryotes, some of these unexplained sequences were eventually found to be involved in splicing, a process in which non-coding (intronic) sequences are removed from pre-mRNA (preliminary form of messenger RNA), leaving behind protein-coding (exonic) regions that make up the mature mRNA transcript (Figure 1.1). Briefly, splicing occurs at defined sequences in the pre-mRNA through the recruitment of specific protein factors that make up the complex known as the spliceosome. Splicing patterns can be more complicated than simple intron removal, and often involve what are known as alternatively spliced exons. These genetic elements can be included or excluded from the mRNA sequence to produce different protein products, or isoforms, and add a great deal of diversity to the cellular proteome. Alternative splicing patterns can change depending on environmental stresses or developmental lifestages, allowing an organism to adjust protein levels and function in response. Although pieces of the core spliceosomal machinery are highly conserved across eukaryotes, certain features do vary between species. For example, the plant *Arabidopsis thaliana* contains nearly two times the number of protein splice factors found in the human genome, a shocking number given its diminutive 25 centimeter height. Splicing is an essential process in eukaryotes, and malfunctions in the spliceosomal machinery have been directly linked to human disease. For further information, we refer readers to relevant reviews on alternative splicing ,[5–7] as well as a more detailed explanation in Chapter 2 of this work.

With the advent of next-generation sequencing techniques, the amount of known, untranslated RNA in the cell has continued to grow.[8,9] Although alternative splicing accounted for a small portion of this RNA, studies soon showed that the remainder of this sequence space is dominated by ncRNAs (non-coding RNAs). There are a number of ncRNAs involved in gene regulation and these elements are classified as either *cis*- or *trans*-acting. *Cis*-acting RNAs influence the expression of genes sharing the same locus and can originate from upstream elements or even intron sequences in an RNA transcript.[10] In contrast, *trans*-acting RNAs govern expression in distal regions of the genome. Here, we will focus on two particular types of *trans*-acting RNAs in

eukaryotes, miRNA (microRNA) and siRNA (small interfering RNA), as both are relevant to this work.

Both miRNAs and siRNAs serve to repress gene expression through direct interaction with a target mRNA sequence (Figure 1.2). miRNAs are 21-25 nucleotide sequences that are prepared within the cell through several cleavage steps by the Drosha and Dicer RNase-III enzymes. First, pri-miRNAs (nascent miRNA transcripts) are cut into individual pre-miRNA stem loops by Drosha in the nucleus. Next, these pre-miRNAs are exported to the cytoplasm and processed by Dicer to produce miRNA duplexes. These duplexes are unwound to form mature, single-stranded miRNA which is incorporated into the RISC (RNA-induced silencing complex) protein machinery of the cell. This RISC-miRNA complex associates with mRNA targets that have partial complementarity to the miRNA sequence, typically leading to either mRNA degradation or recruitment of protein factors to repress translation.[11]

Although siRNAs are also 21-25 nucleotide sequences that repress gene expression, their biogenesis and function are slightly different within the cell. These ncRNAs originate from double-stranded RNA sequences in the cytoplasm, which associate with Dicer to produce siRNA duplexes. A single strand of the duplex preferentially associates with the RISC complex, targeting it to mRNA with perfect complementarity to the siRNA sequence. Unlike miRNA (with a few rare examples), siRNA generally bind to only a single site and direct cleavage of the target mRNA at the region of complementarity.[11,12] Despite these minor distinctions, both miRNAs and siRNAs are important mechanisms that control gene expression within the cell. These ncRNAs are essential for development and are found in within the genomes of nearly all metazoans.[12]

Bacterial RNAs

Although bacteria lack the machinery for alternative RNA splicing, they contain a diverse set of RNAs to provide tight regulation of gene expression. Like eukaryotes, bacteria require cellular components that can sense and respond to stimuli from the outside environment. Although some of these reactive pathways are dominated by proteins,[13] there are also several types of RNA that control bacterial responses. Three major classes of RNAs that regulate native or pathogenic gene expression are riboswitches, sRNAs (small RNAs), and CRISPR (clustered regularly interspaced palindromic repeat) RNAs.

Riboswitches are structured RNAs typically found in the 5' UTR (untranslated region) of mRNA that provide post-transcriptional control of gene expression.[14] These RNA elements are predominantly found in prokaryotes, although there are some known examples in eukaryotes.[15] Briefly, riboswitches recognize and bind to specific small molecule targets in the cell and undergo a conformational "switch" to produce changes in gene expression through several different pathways (Figure 1.3). There are a number different riboswitch classes, and these RNAs are involved in many different processes within the cell, such as metabolic feedback loops, intra- and intercellular signaling, and ion sensing. For further detail on riboswitch function and applications, we refer readers to relevant reviews[16–18] as well as Chapter 4 of this work.

Bacterial sRNAs are 50-500 nucleotide sequences that are broadly classified as *cis*- or *trans*-acting depending on whether they have perfect complementarity with their target sequence[14]. *Cis*-acting sRNAs are transcribed from and directly bind to their target mRNA sequence, forming duplexes to create changes in gene expression. *Trans*-acting sRNAs bind to their target with partial complementarity and interact with RNA chaperones, such as the Hfq protein, to influence the fate of mRNA transcripts (Figure 1.4)[14]. We will focus on *trans*-acting elements here, since they are relevant to this work.

Depending on the target region, sRNAs can either up- or downregulate gene expression. Known examples of gene repression include sRNAs that obscure the ribosomal binding site or reveal an RNase E cleavage target (Figure 1.4). Gene activation has also been observed when sRNAs alter mRNA secondary structure upon binding, releasing terminator hairpins or blocking similar programmed cleavage sites. These ncRNAs can even influence transcription by recruiting protein factors to alter to rate of RNA polymerase.[19,20] sRNAs are often tightly regulated depending on environmental context, allowing bacteria to respond rapidly to changes in their surroundings. For example, certain sRNAs are only expressed during specific stages of infection, allowing pathogens to produce the proteins necessary for host cell invasion.[21] Other sRNAs are essential to certain metabolic processes, and levels spike upon exposure to a potential food source.[22]

CRISPR RNAs and Cas (CRISPR-associated) proteins are found in roughly half of all bacteria[23,24] and act as a primitive immune system, responding to invading pathogens and repressing foreign gene expression.[25] Unlike the other two RNAs described in this section, they are not strictly classified as regulatory RNAs, but they still control gene expression. The mechanism is slightly different, however, since CRISPR RNAs exclusively target foreign genetic material. Although the exact mode of action varies between organisms, the general mechanism remains the same (Figure 1.5). Foreign DNA is recognized by bacteria and integrated into the CRISPR locus of the host genome, then transcribed and processed into mature crRNAs (CRISPR RNAs) that are approximately 23-47 nucleotides in length.[26] These RNAs bind to Cas proteins and target them to specific regions of foreign nucleic acids, leading to single- or double-stranded cleavage and effectively repressing gene expression of any invading pathogens. The CRISPR / Cas9 system from *Streptococcus pyogenes* is the most well-studied member of this family and can produce double-stranded DNA breaks when programmed for recognition of a 20-nucleotide protospacer (target region). This system also requires a particular 3-nucleotide PAM (protospacer adjacent motif) in target DNA[27] and has been shown to tolerate up to 5 base-pair mismatches when sampling DNA sequences for cleavage.[28]

1.2 Outlook

The numerous classes of regulatory RNA described above provide a small glimpse into the cellular processes that are controlled by this versatile biomolecule. However, in addition to these native functions, RNA has been adapted for use in a variety of synthetic and chemical biological applications. miRNAs and siRNAs can be constructed artificially and expressed in cells to suppress gene function, even acting as a form of gene therapy for humans in the near future.[11] Alternative

splicing machinery has been engineered and repurposed for selective control of pre-mRNA splicing, allowing scientists to develop tools for future treatment of diseases such as cancer.[29,30] Riboswitch-based tools have been developed to study bacterial signaling and lifestyle changes, facilitating the development of more effective treatments for infection. Sensors to study bacterial sRNA have also shed light on the processes of bacterial infection, allowing us to understand how pathogens evade our immune system. Even CRISPR / Cas systems, once a simple bacterial defense mechanism, have been altered for use as tools that control gene editing and expression, revolutionizing the field of genome engineering.

All of these systems showcase the potential for RNA in the future of chemical and synthetic biology. We present here our efforts towards understanding and influencing native cellular functions through the use of these RNA-based tools. This work illustrates the versatility of RNA not only for the study of organisms here on Earth, but also in the environment of outer space. We hope that these studies convince you that the future for RNA is bright – the sky is not the limit for this biomolecule.

1.3 Figures

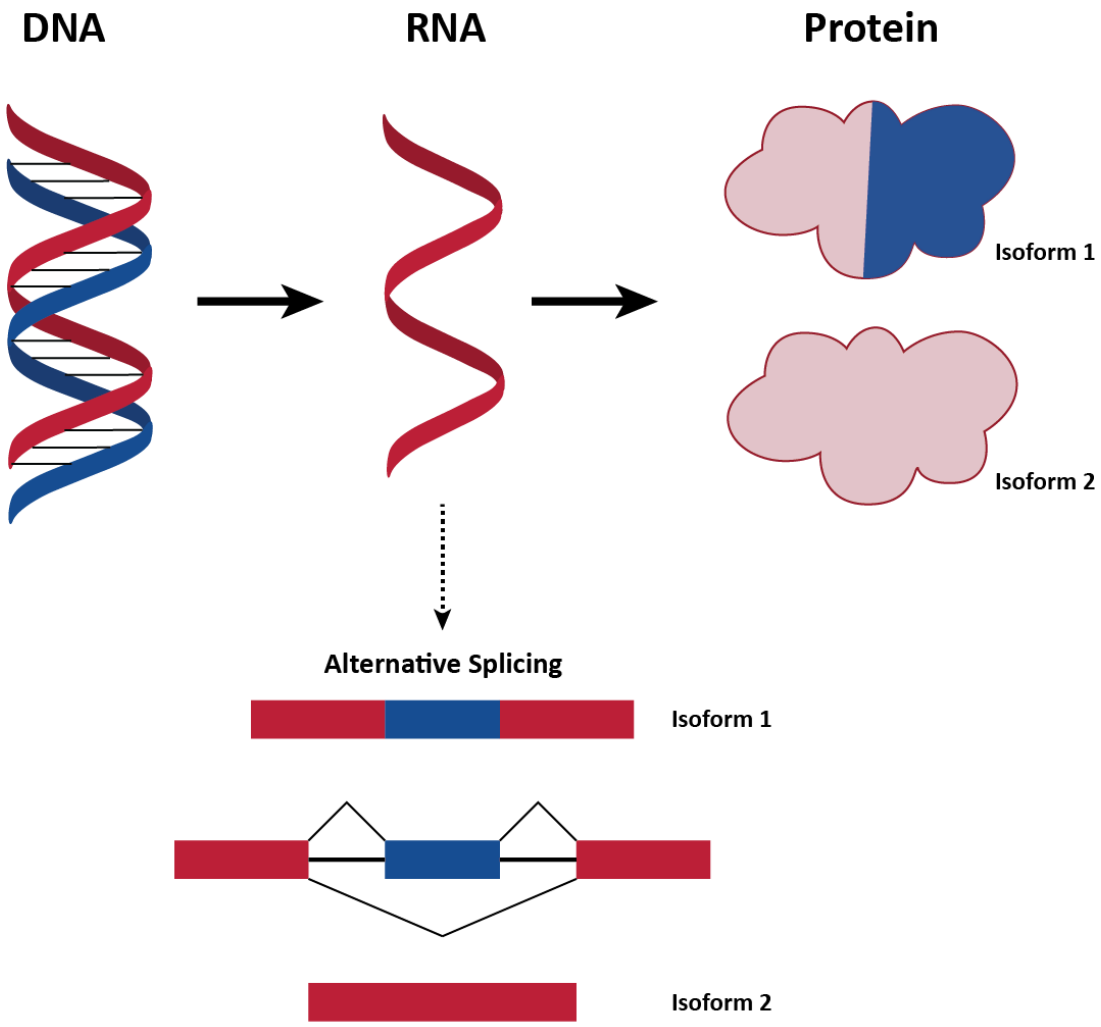


Figure 1.1 Central dogma of molecular biology.
Cartoon illustrating the central dogma of molecular biology, that DNA makes RNA, which makes protein. In the bottom panel, some detail has been added to illustrate how RNA splicing can lead to different protein isoforms.

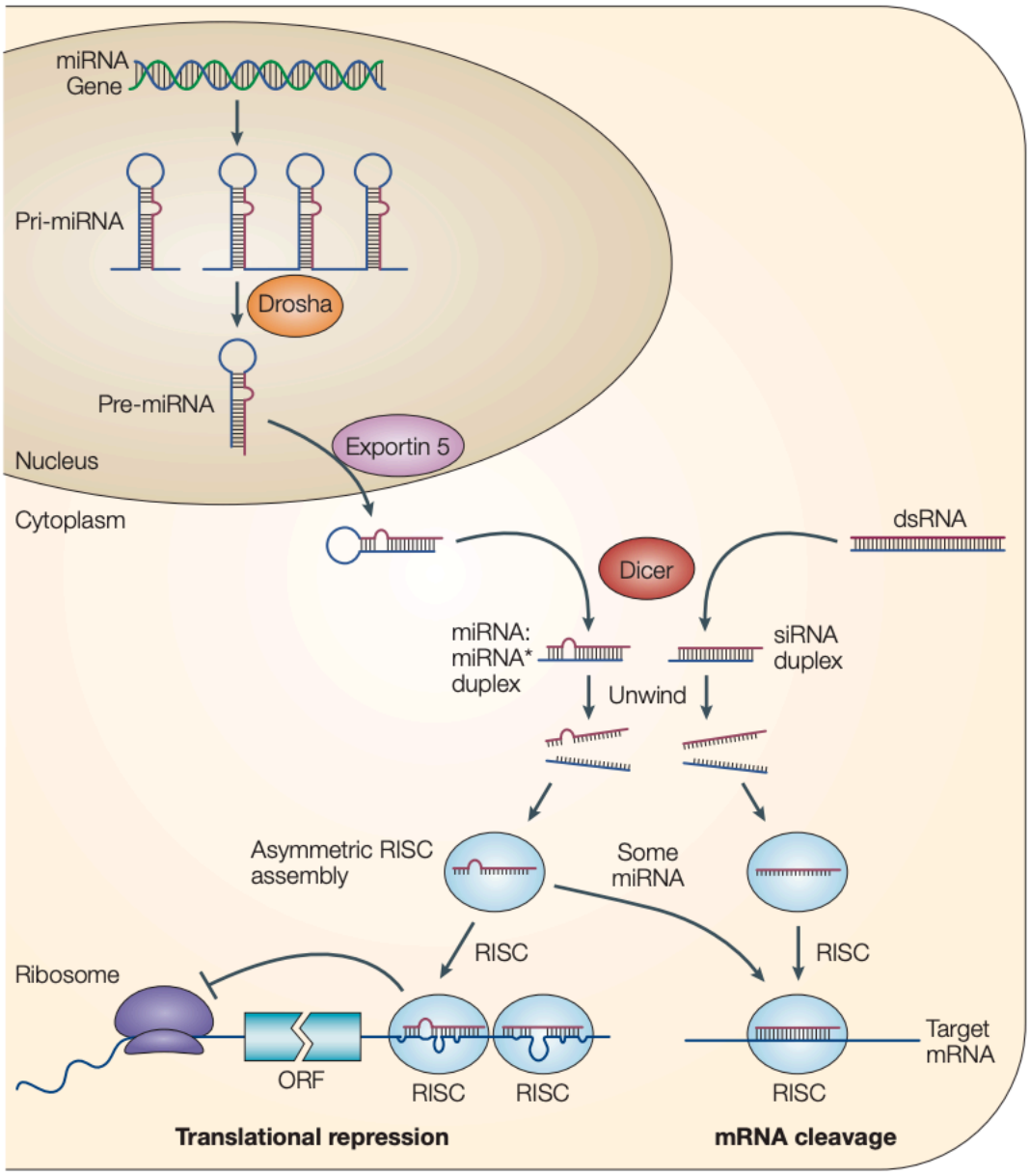


Figure 1.2 miRNA and siRNA biogenesis and RNAi mechanism.
Adapted with permission from He et al[12]. Model for production and action of miRNAs and siRNAs. miRNAs are processed from parent gene by Drosha protein, then processed by Dicer and incorporated into RISC complex for translational repression or mRNA cleavage. siRNAs have different initiation mechanisms, but are also processed by Dicer and perform similar downstream function to cleave target mRNA.

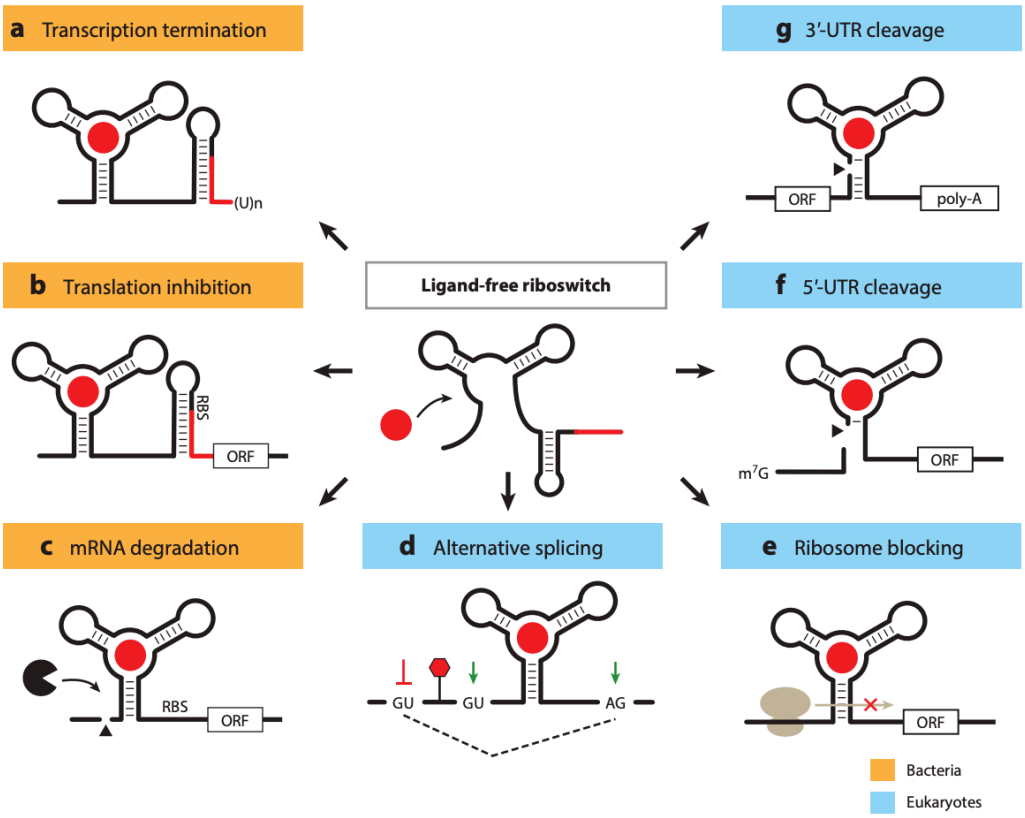


Figure 1.3 Mechanisms for riboswitch gene regulation

Adapted with permission from Hallberg et al[18]. This schematic illustrates the number of different ways in which riboswitches can influence gene expression. Orange mechanisms are in bacteria, blue in eukaryotes. The red circle represents target ligand. **(a)** Transcriptional attenuation: In the ligand-bound form, a terminator hairpin is formed, causing transcription termination. **(b)** Translation inhibition: Ligand binding causes an alternative structure to form, occluding the ribosome-binding site (RBS) and preventing initiation of translation. **(c)** Messenger RNA (mRNA) degradation caused by ribozyme activity: Ligand-induced self-cleavage by ribozyme activity destabilizes the mRNA, leading to rapid degradation. **(d)** Alternative splicing: Ligand binding leads to an alternate mature mRNA forming through changes in splice site selection. Depending on the organism, this leads to addition of a small upstream open reading frame (ORF), inclusion of a premature stop codon, or exclusion of the polyadenosine (poly-A) tail and mRNA degradation. **(e)** Ribosome blocking: In yeast, aptamer–ligand interactions can inhibit the ribosome from scanning from the 5’ 7-methylguanosine cap to the translation start site, preventing translation. **(f)** 5’-UTR (untranslated region) cleavage: Removal of the 5? 7-methylguanosine cap by aptazyme activity prevents ribosome progression and destabilizes the mRNA. **(g)** 3’-UTR cleavage: Removal of the poly-A tail by aptazyme (self-cleaving sequence) activity leads to rapid mRNA degradation.

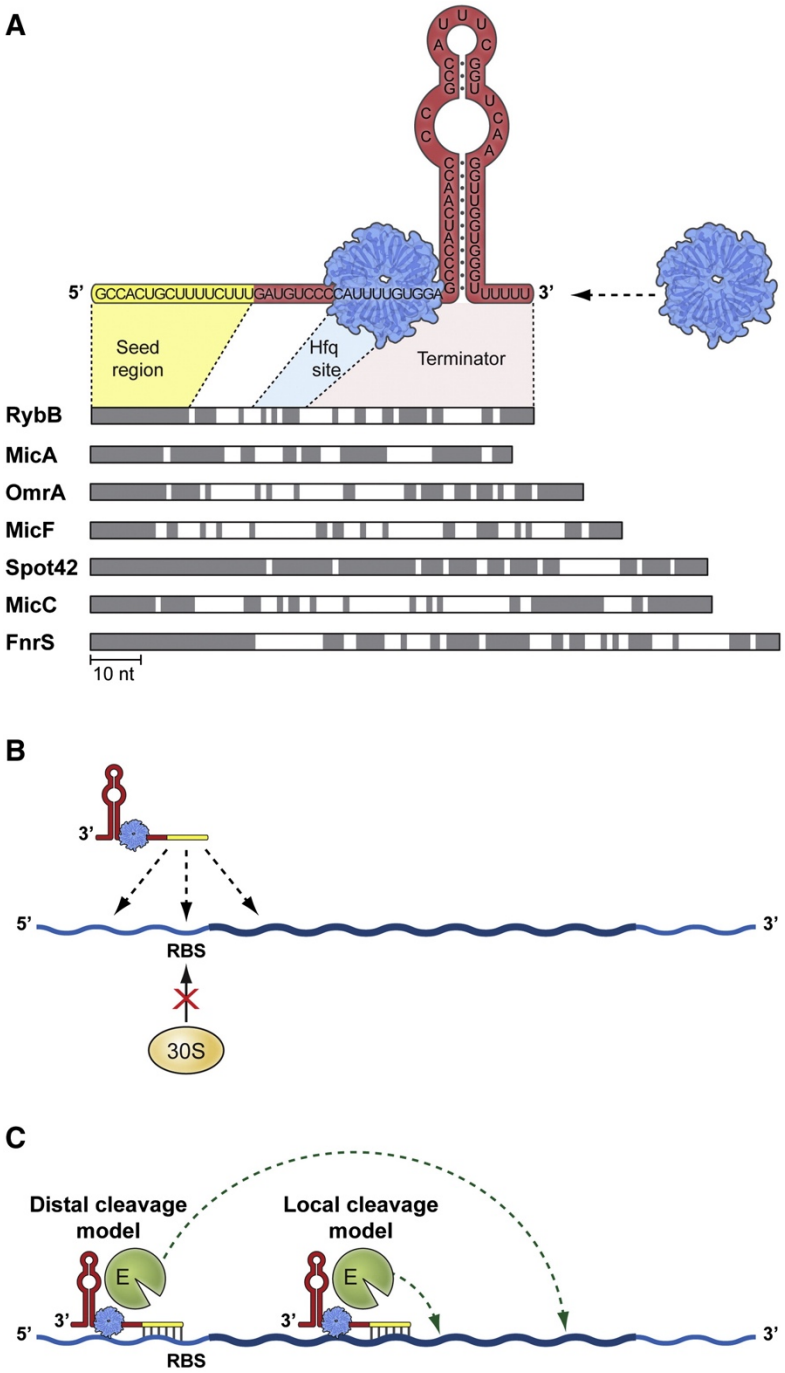


Figure 1.4 *Trans*-acting bacterial sRNAs.
Adapted with permission from Storz et al[31]. **(A)** Structure of Hfq-binding sRNAs, which highly conserved regions in grey. **(B)** Diagram showing different positions where sRNAs can block ribosome binding. **(C)** Different ways that base-paired sRNAs can direct mRNA cleavage using RNaseE.

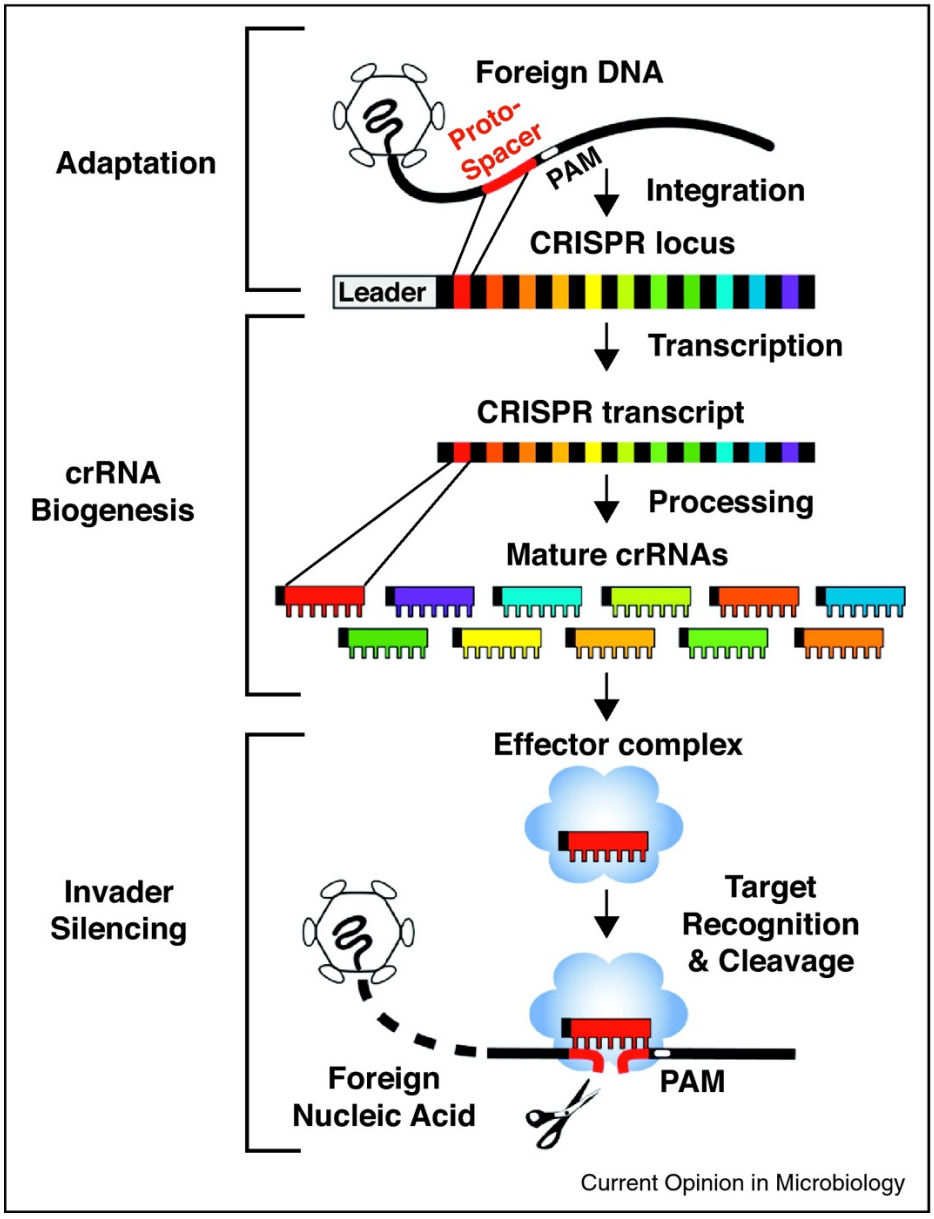


Figure 1.5 CRISPR/Cas response pathway.
Adapted with permission from Terns et al[25]. Overview of the CRISPR/Cas invader defense pathway. Foreign DNA is acquired from invading pathogen and integrated into the host CRISPR locus. These sequences are then transcribed and processed to produce mature cRNAs. crRNA-Cas protein effector complexes target complementary invading DNA or RNA sequences for cleavage, preventing foreign gene expression.

2 Alternative mRNA splicing for gene regulation in eukaryotes

2.1 Introduction

Although RNA features prominently in a diverse number of processes within the cell, it is most well-known in the "central dogma" of molecular biology as the template for protein translation.[32] In eukaryotes, pre-mRNA associates with the machinery of the spliceosome complex to splice out introns and alternative exons, producing functional, protein-coding mRNA.[33] This process is responsible for much of the protein diversity in the cell, allowing a single mRNA sequence to produce dozens of different protein isoforms depending on the pattern of alternative splicing. An interesting example is found in the fruit fly *Drosophila melanogaster*, where a single gene can encode up to 38,016 different isoforms – more than the total number of genes in the entire fruit fly genome!

Although there are a few rare sequences that can self-splice to form functional RNA, the majority of eukaryotic splicing is performed by an RNA-protein complex known as the spliceosome. The spliceosomal ribonucleoprotein complex (RNP) is highly varied depending on the local sequence and cell type and can consist of anywhere between 50-80 proteins and RNAs.[34] These factors localize to specific sequences regions in the pre-mRNA known as the 5' splice site, the 3' splice site, and the branch site (5'SS, 3'SS, and BS, respectively). Because these sequences are highly conserved, diversity in splice isoforms is created by the *trans*-acting factors within the spliceosome. There are several different types of interactions that can influence alternative splicing, including RNA-protein, RNA-RNA, and DNA-based effects (Figure 2.1). [7,35]

Certain factors, such as hnRNPs (hetereogeneous ribonucleoprotein particles) and SR (serine/arginine rich) proteins act as general splicing repressors or activators, respectively, when bound to pre-mRNA. These proteins can enhance the recruitment of other factors or physically block other splicing elements to influence alternative splicing patterns. RNA-RNA interactions can similarly alter the accessibility of splice sites, changing spliceosomal recognition and preference for exclusion/inclusion. Although the majority of splicing changes are effected through direct binding to RNA, it is also possible to alter these patterns by manipulating the DNA template. Protein factors can create PTMs (post-translational modifications) on chromatin, which affect the rate at which RNA polymerase reads through the DNA. Slower or faster transcription rates can lead to changes in RNA folding patterns and subsequent co-transcriptional splicing.[30,36,37] These RNA- and DNA-binding factors are broadly classified as either intron- or exon-defining, with intron definition occurring at the pre-mRNA level and exon definition at the DNA level.[38]

The complexity of eukaryotic splicing pathways affords a great deal of protein diversity, but also means that minor errors can (literally) translate into major downstream effects. Based on data from the Human Gene Mutation Database (as of April 8th, 2019) approximately 21,773 mutations have effects on pre-mRNA splicing, accounting for 8.7% of all mutation entries.[39] These sequences errors can obscure splice factor recognition sites and prevent appropriate spliceosomal assembly in key areas of the genome. Although direct gene therapy may one day provide a solution, another treatment option would be a system that provides targeted delivery of key *cis*- or *trans*-acting factors directly to a splice junction to promote specific alternative splicing events.

In this chapter, we report our efforts toward developing a targeted system to promote specific alternative splicing events in mammalian cells. We also describe our initial work towards the use of a tightly-inducible splicing-based expression in eukaryotic cells. Though these tools were developed orthogonally, it was initially thought that they might be used in parallel to provide inducible expression of a protein for on-demand changes in splicing patterns. Credit for the conception and initial work on the targeted splicing project goes towards Ryan Muller of the Ingolia lab and Yeon Lee of the Rio lab.

2.2 Results

2.2.1 Efforts towards validation of a splicing-regulated inducible protein expression system

Our initial work focused on using an RNA-based alternative splicing sequence developed by our group to provide tightly inducible expression of a gene editing protein (Figure 2.2). This "suicide cassette exon" is known as hybrid P5SM (*Hy*P5SM) and contains a premature termination codon (PTC) which, when retained in the gene of insertion, produces a non-functional splice product (SP-II) that is degraded through nonsense-mediated decay (NMD) pathways in the cell. Normally, the *Hy*P5SM sequence is bound by an unknown splice factor (SF), but this protein is released when the riboswitch binds to the L5 protein (Figure 2.2). *Hy*P5SM has been designed to function in dicots (dicotyledons) through mutation of the L5 binding sequence, allowing it to exclusively bind to the *Oryza sativa* L5 protein (*Os*L5) while avoiding the endogenous *Arabidopsis thaliana* L5. Once *Os*L5 is bound, the native SF protein is displaced and the *Hy*P5SM exon is spliced out, leaving a functional splice product (SP-I). *Hy*P5SM can act as a two-component inducible system in dicots through induction of OsL5. This sequence has been used previously to provide on-demand protein expression in response to a trigger compound, with minimal "leaky" expression in the absence of the compound (Figure 2.3).[40–42]

The *Hy*P5SM system provides an attractive solution for leaky protein expression, and our goal was to adapt it for use with gene editing enzymes. Our main target was the Cas9 protein, which has become the standard for fast, efficient gene engineering in the last decade. Cas9 is an RNA-guided protein that can produce double-stranded breaks (DSBs) in specific DNA sequences (Figure 2.4). Leaky Cas9 expression can be an issue for time-sensitive experiments where protein expression is lethal before a certain developmental stage.[43] High levels of the protein also increase the risk of off-target binding and DSBs, effects that are detrimental in whole-organism experiments.[44] Our aim was to use *Hy*P5SM two-component system to reduce leaky expression and the resulting side effects (Figure 2.5).

To begin, the *Hy*P5SM cassette was cloned into a version of Cas9 optimized for plant expression. Tests were limited to plant systems because the SF protein interacting with *Hy*P5SM is unidentified but likely absent from other eukaryotic systems. It was predicted that insertion of the splicing cassette alone would reduce protein expression compared to Cas9 under a

constitutive promoter (as shown in Figure 2.4 and Figure 2.5). Reduced protein levels should correspond to lower off-target cutting, one of the major side effects of leaky expression.

In order to quantify Cas9 activity *in planta*, we decided to use a system pioneered by collaborator Yan Liang in the Loque lab at JGI.[45] This system relies on the mechanism of DNA repair after DSBs in *Nicotiana benthamiana*, our model plant system. After a DSB is formed, the most common repair pathway is through NHEJ (non-homologous end joining), which can produce a frameshift in 2 out of every 3 repair events. The Loque lab cleverly designed a "GFP switch" system with GFP and a repressor linked to a gene containing a Cas9 target.[45] The repressor prevents GFP expression until Cas9 cuts the linked target sequence. Once a DSB is induced, NHEJ produces and inactivates the repressor, allowing the GFP to mature and produce a fluorescent signal (Figure 2.6).

The GFP switch system was ideal for comparing Cas9 activity between different targets – in this case, between DNA targets with perfect hybridization to the sgRNA and off-target sites with minor mismatches. It was predicted that Cas9-*Hy*P5SM would produce less protein and therefore less intense fluorescence when targeted to these mismatched regions when compared to Cas9 alone. However, after comparing the two systems it was determined that both on-target and off-target activity was much higher for Cas9-*Hy*P5SM (Figure 2.7). After sequence-confirming the Cas9-*Hy*P5SM DNA template to eliminate the possibility of a frameshift (which would destroy the PTC-based control of expression), the next step was to check RNA transcript and protein levels.

Previous research has shown that the presence of intronic elements can sometimes lead to increased gene expression through an unknown mechanism.[46–49] We ran RT-PCR on normalized total RNA for both Cas9 constructs and found that the Cas9-*Hy*P5SM had a small amount of functional SP-I, full-length Cas9 product and a large quantity of non-functional SP-II. This low level of leaky expression was expected and does not account for the increased activity of the Cas9-*Hy*P5SM construct. Next, we quantified the overall protein levels from leaf extracts for the two sets and found that (surprisingly enough) the Cas9-*Hy*P5SM sequence actually produced less functional Cas9 protein than the regular Cas9 sequence!

These results were somewhat baffling as they would imply that Cas9 expression under the control of *Hy*P5SM produced minimal leaky expression without *Os*L5, yet somehow displayed more potent DSB activity. One possible explanation might be that the splicing event improves the longevity of the RNA transcripts or protein product, although it is unclear how this effect would take place. As a result of these initial findings, we were unfortunately forced to discard the idea of adding Cas9-*Hy*P5SM into a two-component induction system.

2.2.2 Efforts towards targeted delivery of dCas9 fusion proteins to influence alternative splicing

Although our efforts towards development of a splicing-based gene regulation tool in plants were unsuccessful, we decided to transfer our relevant skills into a second project focused on native alternative splicing patterns in mammalian cells. This project was based upon previous work by

Ryan Muller of the Ingolia lab and Yeon Lee of the Rio lab, who hypothesized that dCas9 (catalytically inactive Cas9) could be fused to different histone modifying proteins to change the architecture of the DNA template, thereby altering the preferred splice products. At the beginning of the project, there was some precedent for influencing alternative splicing using DNA-targeted systems. The Kornblihtt lab had shown that addition of siRNAs (short interfering RNAs) promotes histone methylation and acetylation events that can alter the rate of RNA pol II (RNA polymerase II) transcription.[38,50] These "roadblocks" slow the enzyme and shift normal splicing patterns to promote inclusion of alternative exons (Figure 2.8).[38]

Three dCas9 fusions to histone modification proteins (dCas9-HMPs) were constructed by Ryan Muller and myself. The chosen proteins were 1) EHMT2, a H3K9 (histone 3 lysine 9) methyltransferase, 1) the catalytic core of P300, a H3K122 and H3K27 acetyltransferase, and 3) LSD1, a H3K4 and H3K9 demethylase. These HMPs were selected because there is some precedent for their use in dCas9 fusions to produce targeted demethylation or acetylation events, albeit for gene activation or repression (Figure 2.9).[51–53]The basic idea of this project was to target these dCas9-HMPs to DNase-sensitive region of the DNA adjacent to alternative exons and produce a change in the percentage spliced in (delta PSI) compared to cells with dCas9 fused to an inactive protein (Figure 2.10).

There are a large number of known alternative exons in humans, and the first step in this project was to identify exon targets that would be sensitive to HMP activity, Yeon and Ryan came up with the idea to treat cells with two different methyltransferase inhibitors and generate an RNA-seq library from the total cellular RNA (Figure 2.11). The data could be analyzed for region with changes in splicing patterns compared to control RNA. Sequences sensitive to HMT (histone methyltransferase) activity would also be potential sites for acetyltransferase modification, allowing us to pick sequences to target with the dCas9-HMPs.

The library preparation required some optimization of the drug treatment to provide maximum inhibitory activity without causing cell death. In the end, it was found that a 0.5 μM concentration of the drugs produced minimal changes in cell viability and this treatment was used for the final RNA library. Two separate libraries were generated using chaetocin, a general HMT inhibitor, and A366, an inhibitor that specifically reduces EHMT2 activity.[54,55] The total RNA was prepared (see Materials and Methods) and RNA integrity was verified on a Bioanalyzer before it was sent off for sequencing at the university core facility. The data came back within 1-2 months and was analyzed by Yeon Lee of the Rio lab using both MISO (Mixture of Isoforms)[56] and JUM (Junction Usage Model)[57] software. The A366 treatment seemed to have a number of promising targets compared to chaetocin, so these results were chosen moving forward. Initial analysis with MISO generated a list of 6 potential hits that displayed a ΔPSI >20 / ΔPSI <-20 and had a high Bayes factor.[56]

The six candidates are shown in Figure 2.12 below, along with their corresponding delta PSI values. The convention in this case is that delta PSI describes the change in percentage of exon spliced in versus the control treatment (DMSO only). A ΔPSI > 0 means that the treatment has

lead to increased production of SP-II (splice product II), or increased exon inclusion. Conversely, a ΔPSI < 0 means that the treatment favored production of SP-I (splice product I), or increased exon exclusion.

In order to verify these targets, the first step in the process was to design primer pairs that could span the alternative exon and provide a readout of ΔPSI when analyzed with RT-PCR. I optimized primer sets for all six genes and the final pairs were chosen for high specificity and minimal side product banding on-gel. Next, Yeon and I performed a second round of drug treatment with A366 and total RNA was isolated from treated and control cells. The RNA was reverse transcribed into cDNA, which was normalized and used as a template for RT-PCR with the chosen primer sets. Delta PSI values were calculated for each sample using the intensity of the two product bands (exon included and excluded), allowing us to verify which RNA-seq splicing changes could be reproduced (Figure 2.12)

Unfortunately, only a single candidate (UAP1, alternative exon 9) showed reproducible changes in alternative exon splicing in response to the A366 treatment. We were surprised to observe that this change did not occur at the previously optimized 0.5 µM dosage, but instead at 1.5 µM -- three times the amount used for library preparation. After this discovery, Yeon conducted a second of data analysis, this time with the JUM software. She found that the previously identified hits were most likely misannotated in MISO, which has a tendency to misinterpret regions with complex splicing patterns (e.g. more than one alternative exon that can splice together or separately). JUM did not produce any new hits from the library analysis, so we decided to continue with UAP1 in spite of the splicing inconsistency observed after repeating drug treatment.

Now that a relevant target was in sight, we formulated hypotheses about the effects that the dCas9-HMPs would have on ΔPSI values for UAP1. Based upon information in the literature, it was likely that EHMT2 (a methyltransferase) would produce ΔPSI >0 and LSD1 (a demethylase) would induce the opposite effect, with ΔPSI < 0.[50] These predictions were also supported by the fact that A366, which inhibits EHMT2 activity, produced ΔPSI <0.There was less precedent for acetyltransferase effects on splicing, so it was unknown how this fusion would perform. decided A pooled guide RNA strategy was used to target the fusions to multiple sites within DNase-sensitive regions near the UAP1 target exon. Four different sgRNAs (single guide RNAs) were designed using the CRISPR software available online through the Zhang lab[58]. Targets were selected within DNase hypersensitive regions annotated on the UCSC genome browser.[59] Guides were chosen for high specificity to the target region, with minimal off-target sequences in the human genome.[58,60] pooled and transfected along with the dCas9 fusions in 5 different ratios and the resulting splicing changes were quantified on-gel via RT-PCR. The dCas9-HMPs were compared to dCas9-HaloTag, which was meant to serve as a non-modifying fusion control.

The 1:3 ratio of dCas9 fusion to sgRNA was eventually chosen due to the large splicing effects and minimal error compared other ratios. These conditions were verified again via Bioanalyzer run to provide more sensitive quantification of the data. Although the P300 acetyltransferase

fusion showed variable results in both cases, EHMT2 and LSD1 both showed ΔPSI>0, an unexpected result (Figure 2.13). This would imply that both a methyltransferase and demethylase had the same effects on alternative splicing, albeit to different degrees (EHMT2 had a larger ΔPSI than LSD1). One possible explanation is that the dCas9-HaloTag fusion is not a proper control for this experiment due to its small size (31.7 kDa) compared to the other proteins (average size of 90.6 kDa). We began cloning and testing catalytically inactive LSD1 and EHMT2 (dLSD1 and dEHMT2, respectively) as controls for future experiments. These two dCas9-HMPs were the main focus going forward due to the extreme variability seen in the dCas9-P300 samples.

Final testing for the dCas9-HMP constructs was conducted while dLSD1 and dEHMT2 were being cloned, and the results show the subsequent inclusion of these new constructs (Figure 2.14). Although initial testing showed similar positive ΔPSI values for LSD1 and EHMT2, the relative size of this change had reversed, with LSD1 now showing a larger ΔPSI than EHMT2. Subsequent testing revealed extreme variation between biological and technical replicates, with any ΔPSI obscured by error during the quantification process. Although these results were disappointing, they may simply indicate that the effects of HMP modification are minimal for this particular target exon. Based upon the combined results presented here, it was concluded that further screening might not be possible due to the fact that the drug treatment method was unable to reliably produce candidate gene targets.

2.3 Discussion

2.3.1 Splicing-regulated protein expression

Although our efforts toward applying the *Hy*P5SM suicide exon cassette for inducible Cas9 expression failed, this system could still be useful for other proteins. For example, previous work has shown that this suicide cassette exon can provide tight regulation of inducible protein expression *in planta*[42]. Interestingly, despite reduced transcript and protein levels of Cas9 in the uninduced *Hy*P5SM construct, the enzyme showed increased on- and off-target activity. Perhaps further investigation could uncover the mechanism behind this improved efficiency and apply it to other enzymes in need of an activity boost.

Since the end of this project in 2016, a number of alternate systems for tight regulation of Cas9 expression have emerged, including several light- and chemically-inducible constructs.[61–67] These engineered versions of the enzyme have solved many of the issues that we sought to address with our two-component system, with the added advantage that they can function in a variety of eukaryotic systems. In short, our aims were ambitious on this project but the system could not meet the current needs in the field.

Although our work with this splicing system did not produce the desired results, the skills and expertise that were necessary for these experiments paved the way for another RNA-based plant project in the lab. Our experience with the *Hy*P5SM system and chemically inducible protein expression were transferred into plant growth studies under microgravity conditions, outlined below in Chapter 3 of this work.

2.3.2 Targeted control of alternative splicing

Despite promising initial results, our dCas9 fusions for targeted histone modification failed to produce consistent, measurable changes in alternative splicing patterns. The drug treatment protocol only produced a limited number of targets, and we were unable to continue testing the HMP fusions further due to a lack of candidate genes. Continued work on this project would require refining the target selection strategy and verifying HMP activity *in vivo*. With a longer target list, it is possible that these dCas9 fusions could produce the desired changes in alternative splicing. Additionally, further optimization could have been performed on the dCas9 fusion designs. Different fusion linker lengths, arrangements, and protein tags may have influenced the activity of the HMPs and changed the experimental outcome.

Although our histone modification strategy did not function as expected, Cas enzyme fusions have been successfully employed to direct alternative splicing patterns since this project was discontinued in 2017. These projects were designed to focus on RNA-targeted Cas proteins (a more recent discovery in the field) rather than the traditional DNA-targeted dCas9. Two different research groups have successfully employed dCasRx, the catalytically inactive form of an RNA-targeted ribonuclease, to induce changes in exon inclusion/exclusion. The Hsu lab has shown that dCasRx alone can promote exon exclusion through steric effects, while fusions to the hnRNPA1 SF protein made this effect even more pronounced.[68] Meanwhile, the Chang group has tested alternate dCasRx fusion to other splice factors to activate exon inclusion in certain target genes.[69] These results suggest that RNA-targeted systems are likely to have stronger effects on alternative splicing patterns. Although targeted histone modifications may still be a candidate for future testing, our results suggest that a better screening method for candidate exons is the first step towards validating these splicing changes. Perhaps future researchers could optimize this system to determine if Cas protein fusions can indeed change splicing patterns through histone markers.

Author contributions

Tania Gonzalez and Yan Liang provided valuable feedback, guidance, and constructs for the Cas9-*Hy*P5SM work. Ryan Muller of the Ingolia lab cloned the dCas9-EHMT2 construct and gathered preliminary data to make the dCas9 project possible. Yeon Lee helped prepare the drug treatment library side-by-side, analyzed the data, and worked to validate target candidates, including UAP1 (although the data shown here was collected independently with a separately developed primer set). She worked in parallel on all of the preliminary experiments prior to final dCas9 fusion testing. Yeon went out of her way to mentor Ryan and myself on this project and made an effort to teach us the basics of mammalian tissue culture, transfection, and RNA processing.

2.4 Materials and Methods

Leaf imaging experiments

Agroinfiltration experiments were conducted in *Nicotiana benthamiana*, as described in Gonzalez et al.[42] with *Agrobacterium tumefaciens* strain GV3101. Cas9 constructs and guide RNAs were expressed on a pCR-EF plasmid. The target plasmid sequence is as described in Liang et al., on a pTKan vector.[45] Constructs tested with one on-target sgRNA and 5 off-target guides and fluorescence intensity was measured using cut leaves scanned on a Typhoon Trio + on the GFP wavelength channels. Relative fluorescence intensity was calculated in ImageQuant software as described in Gonzalez et al. and the value were compared for an average of 5 leaves per construct. Protein sequences used in this project are described in Table 2.1. Constructs for GFP switch system were generously shared by the Loque lab.

Molecular cloning

Cas9 and pcoCas9 sequences were obtained from the Loque lab in plant expression plasmid pCR3-EF. Cas9-HyP5SM was cloned through Gibson assembly of the HyP5SM cassette into the hCas9 sequence approximately 2/3 into the sequence (similar to use in past genes). Cas9-HyP5SM was inserted into a pBinAR plasmid for initial splicing response tests. Constructs were expressed under the constitutive Cauliflower mosaic virus promoter for all plant tests. The single guide RNA sequences were ordered as complementary oligonucleotides with overhang regions for insertion into the pCR3-EF expression plasmid. Oligos were annealed at room temperature before ligation with cut plasmid using T4 DNA ligase.

dCas9 fusions were expressed in a pCDNA3.1 vector provided by Ryan Muller and the Rio lab. Ryan cloned the dCas9-EHMT2 fusion and we cloned the remaining fusions using Gibson assembly into the vector backbone. The sgRNA constructs were expressed on a pgRNA vector, which we modified to contain BsmBI and BsaI cut sites after the U6 promoter sequence. This allowed for ordering of sgRNAs as short oligonucleotides with complementary overhangs to anneal scarlessly after the promoter (rather than ordering the entire sequence as an ultramer, which was the previous method). Fusions were sequence verified using Sanger sequencing through Genewiz prior to transfection experiments.

Cell transfection experiments

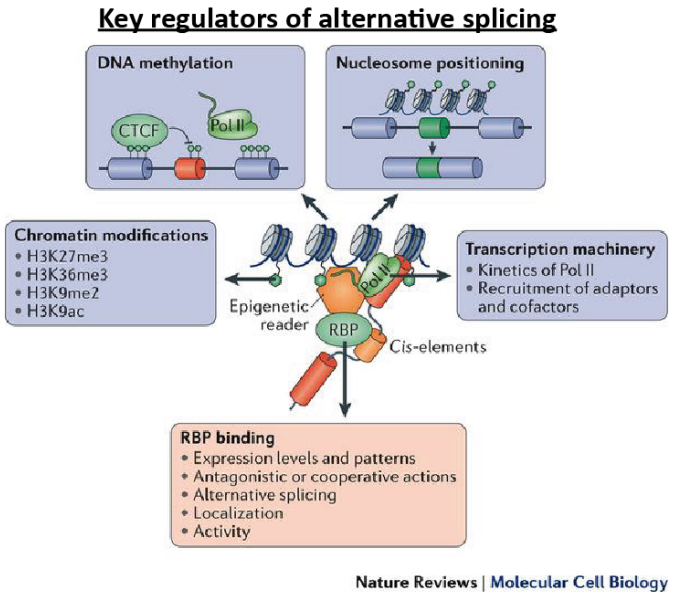
All experiments were conducted in HEK293 cells from the ATCC (American Type Culture Collection). Cells were grown in 6-, 12-, 24-, and 96-well format for initial experiments but 96-well format was chosen for final testing to save on material. Transfections were performed using Lipofectamine 2000 reagent with standard protocol for 120 ng total plasmid DNA. Plasmids were pooled for transfection and a pMAXGFP standard was used as a control for transfection efficiency. Cells were split to allocate 20,000 cells / well before transfection. 2 x 96-well plate volumes were transfected and then duplicate wells were combined into 24-well plates after 24 hours to allow cells to reach higher confluency. Cellular RN was harvested 72 hours after transfection using Trizol extraction. Stocks were frozen in Trizol at -80°C until processing.

RT-PCR analysis

Total RNA was processed according to Trizol Reagent (Thermo Fisher Scientific) protocol and normalized prior to reverse transcription into cDNA using iScript Select cDNA Synthesis Kit (Bio-

Rad Laboratories) with random hexamers.Approximately 500 ng of total cDNA was taken through to RT-PCR and products were run on an 8% native PAGE gel, before staining with SYBR Gold. Controls with no RNA and no reverse transcriptase were run through cDNA synthesis and RT-PCR to ensure that all bands were due to RNA rather than DNA contamination. This step was essential as DNase processing could not be performed – the digestion often led to loss in RNA yield, which was already low due to small cell quantities. ImageLab (Bio-Rad Laboratories) software was used to process the images and provide pixel intensity quantitation for the two splice product gel bands. Global background subtraction was used for all samples. For more sensitive analysis of top hits (rather than bulk screening), the RT-PCR products were also run on a Bioanalyzer with assistance from Yeon Lee.

2.5 Figures



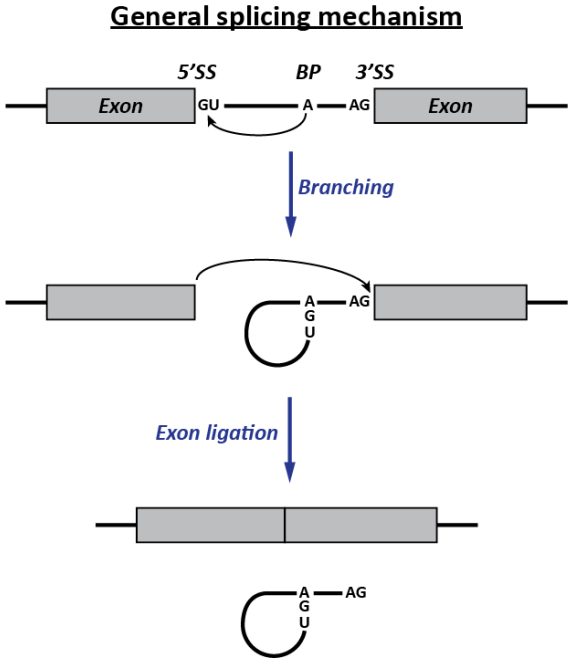


Figure 2.1 Key regulators and general mechanism of splicing.
Top panel adapted with permission from Baralle et al.[35] showing key regulators of alternative splicing patterns and recognition. Bottom panel shows general mechanism of alternative splicing, with the 5' splice site (5'SS), branch point (BP), and 3' splice site (3'SS) shown in the context of an intron between two exon elements.

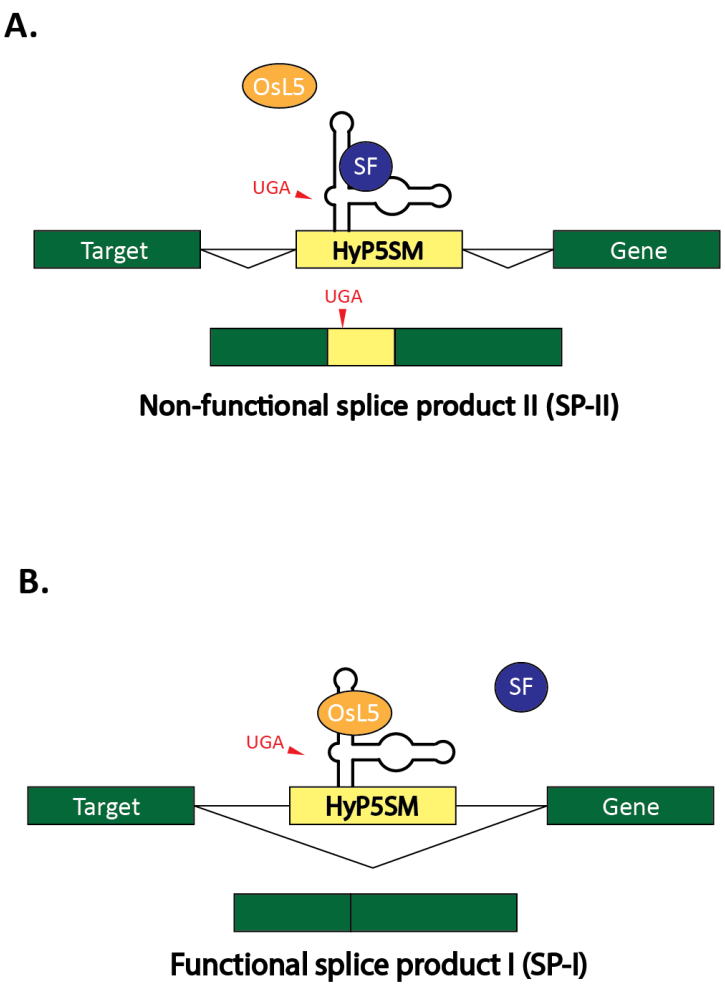


Figure 2.2 *Hy*P5SM Schematic.
(A) Plant leaf showing response (leaf cell death, yellowing) in response to expression of protein. **(B)** Traffic light schematic showing constitutive, leaky, and two-input response in protein expression. Leaky expression is not seen on-leaf with the two-input, splicing-regulated system.

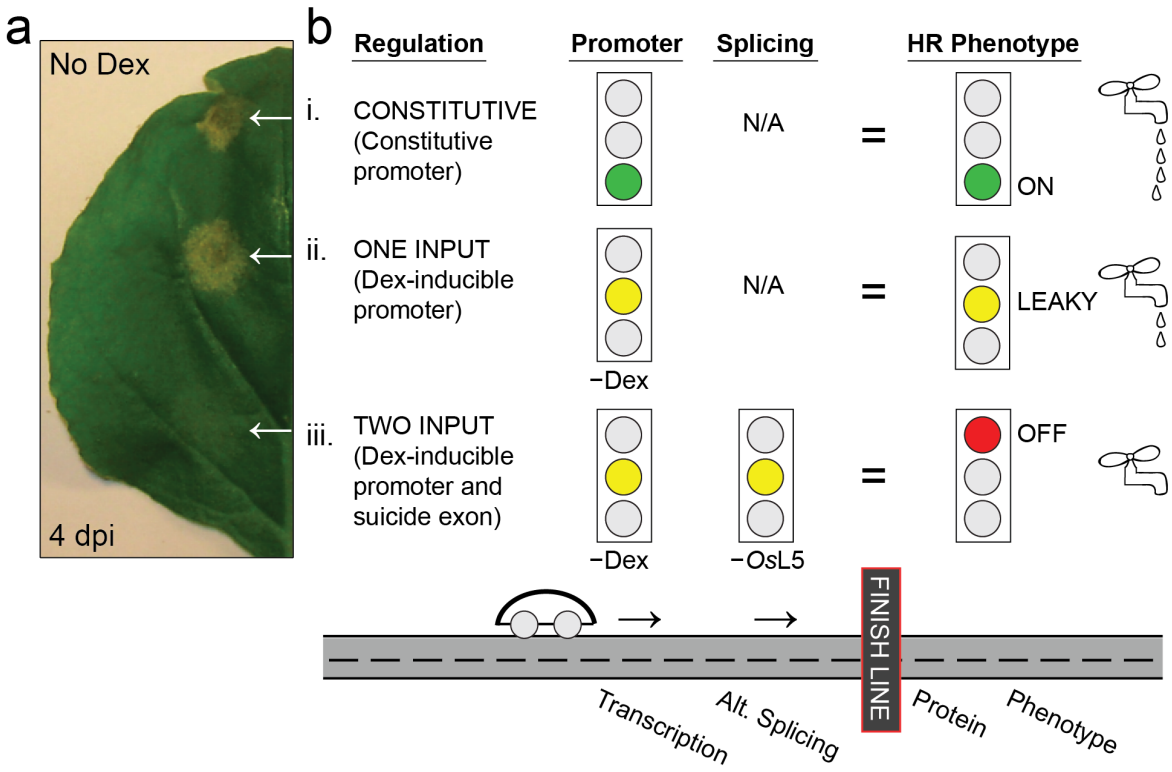


Figure 2.3 Illustration of leaky protein expression.
Adapted from Gonzalez *et al.*[42] **(A)** Plant leaf showing response (leaf cell death, yellowing) in response to expression of protein. **(B)** Traffic light schematic showing constitutive, leaky, and two-input response in protein expression. Leaky expression is not seen on-leaf with the two-input, splicing-regulated system.

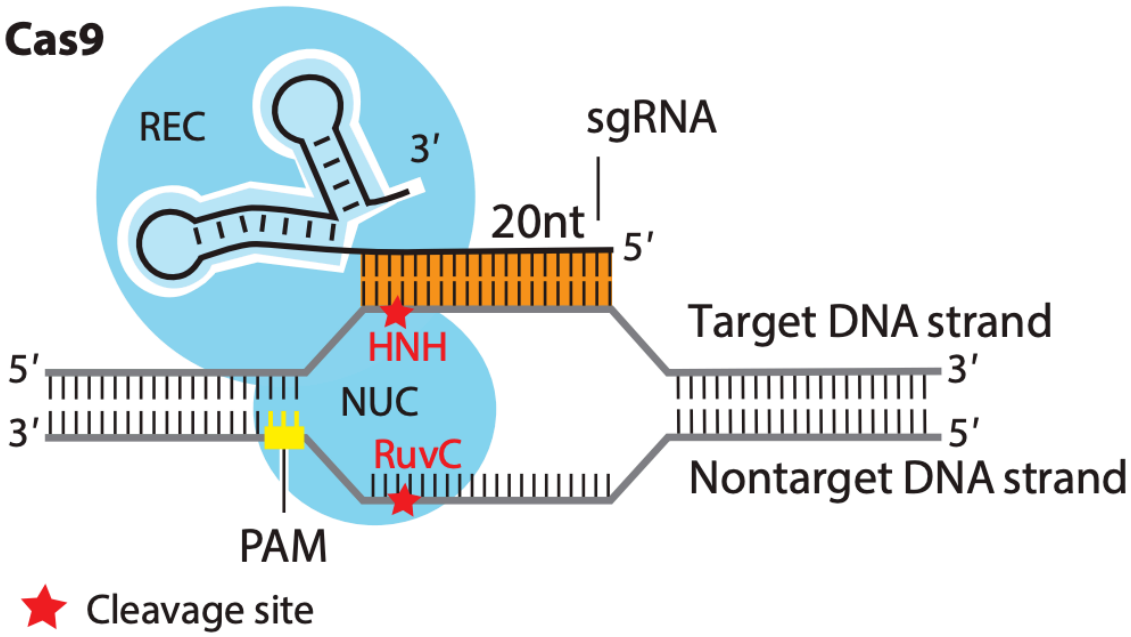


Figure 2.4 Cas9 RNA-guided cleavage of target DNA.
Adapted with journal permission from Wang *et al.*[70] Schematic shows the Cas9 protein (blue) bound to its single guide RNA (sgRNA) interacting a target dsDNA strand. The HNH and RuvC nuclease domains cleave at two distinct sites (red stars), resulting in a blunt-end DNA break. This DSB can be repaird

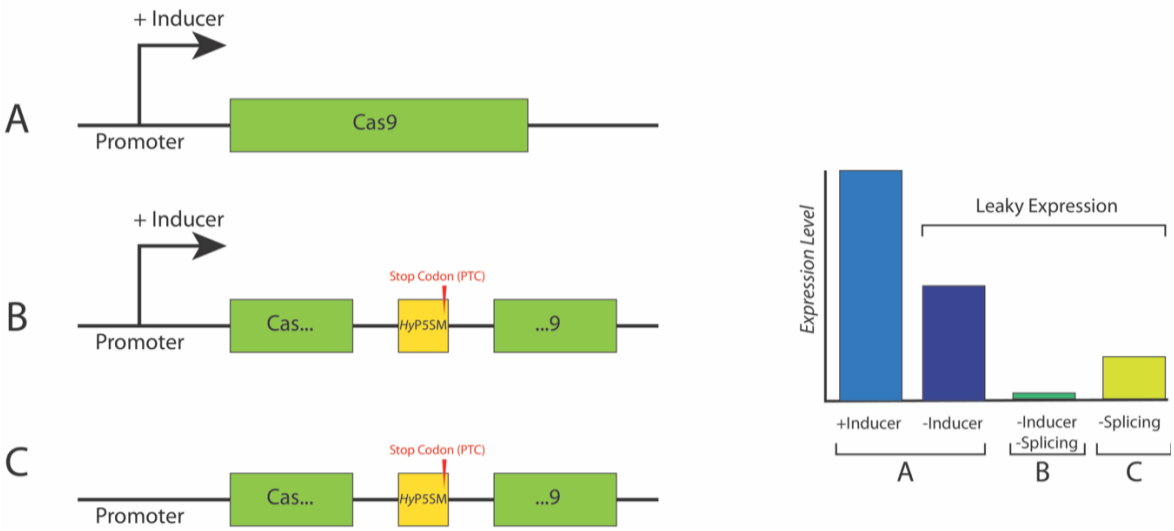


Figure 2.5 Two-component regulation of Cas9 expression.
(A) One-component expression of Cas9 under an inducible promoter. **(B)** Two-component expression of Cas9 using an inducible promoter and splicing regulation. **(C)** One-component expression of Cas9 using splicing regulation alone. Panel to the right shows A, B, and C and theoretical expression levels of Cas9 for each system +/- inducer and without splicing

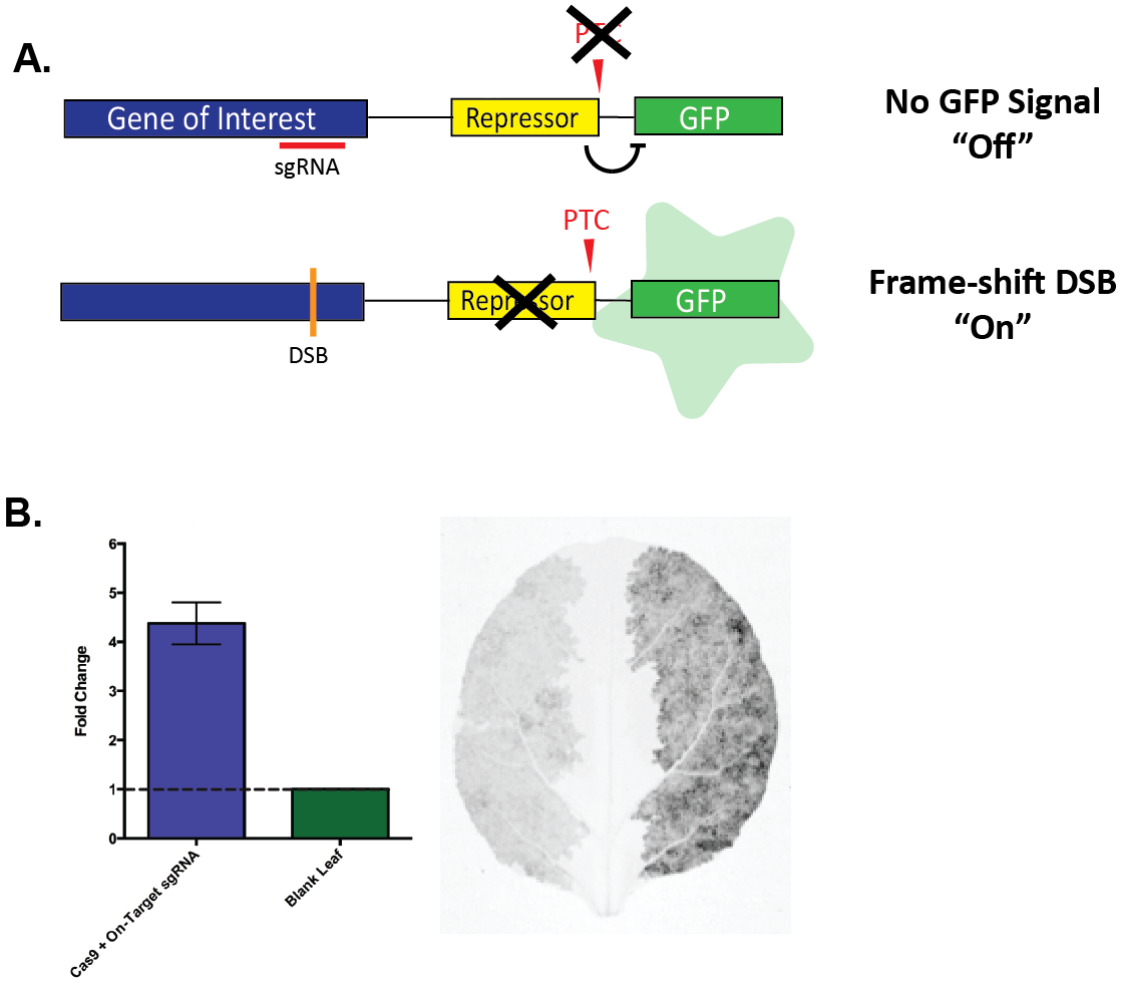
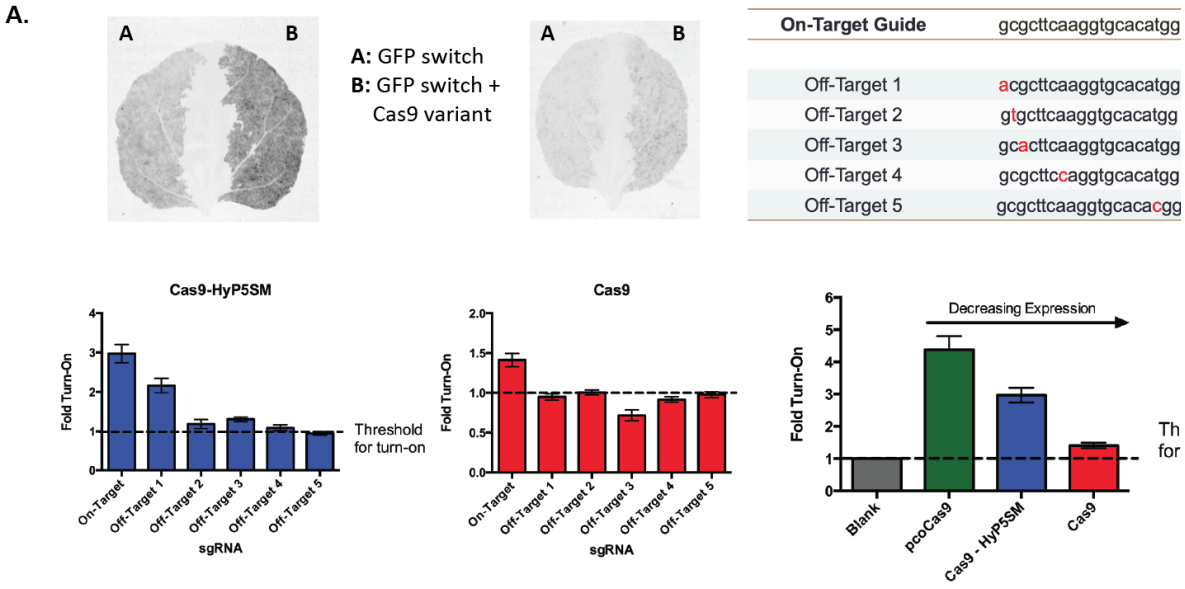


Figure 2.6 Loque lab Cas9 activity visualization system.
System developed by Dr. Yan Liang from the Loque lab.[45] **(A)** GFP switch construct containing GFP linked to a repressor domain. Without DSB-induced cleavage at the sgRNA binding site, the GFP is repressed. Once the DSB is induced, 2/3 events will produce a frameshift and a PTC is added to the repressor. The repressor is then turned off and GFP is expressed, producing fluorescent signal. **(B)** Quantification of fluorescence (left) and representative leaf image (right) after agroinfiltration of plasmids containing the GFP switch and/or Cas9 with sgRNA. The left leaf half has no Cas9 and is "blank" while the right leaf half has both constructs. There is a clear turn-on in fluorescence after once Cas9 is added and induced DSBs in the GFP switch target.



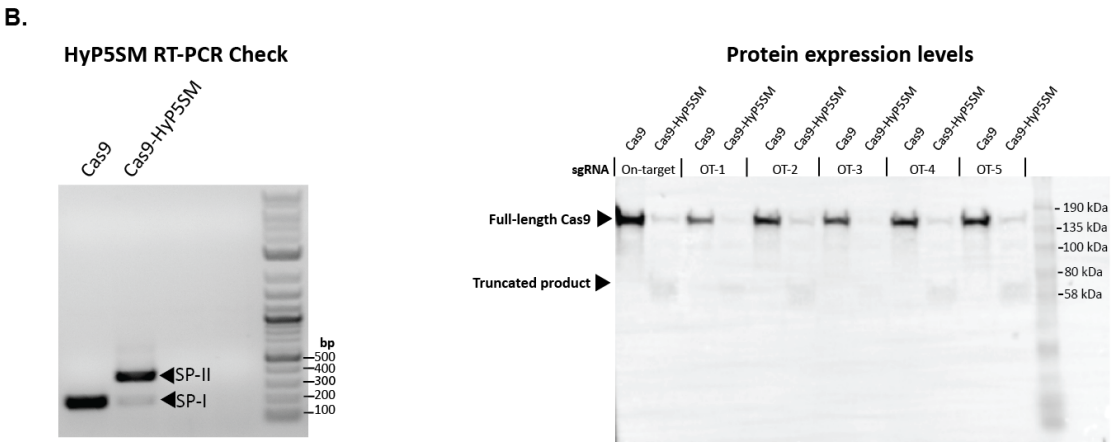


Figure 2.7 Cas9 vs. Cas9-*Hy*P5SM performance

(A) Representative leaf images and corresponding quantification of Cas9 activity for on-target and off-target sequences inserted into the GFP switch. Off-target mismatches are highlighted in red within the table. Overall on-target expression for Cas9 and Cas9-*Hy*P5SM compared to negative control (Blank) and high expression positive control (pcoCas9). **(B)** RT-PCR check (left) and Western Blot (right) of Cas9 and Cas9-*Hy*P5SM to determine if differences in Cas9 activity are due to transcript or protein levels in the cell.

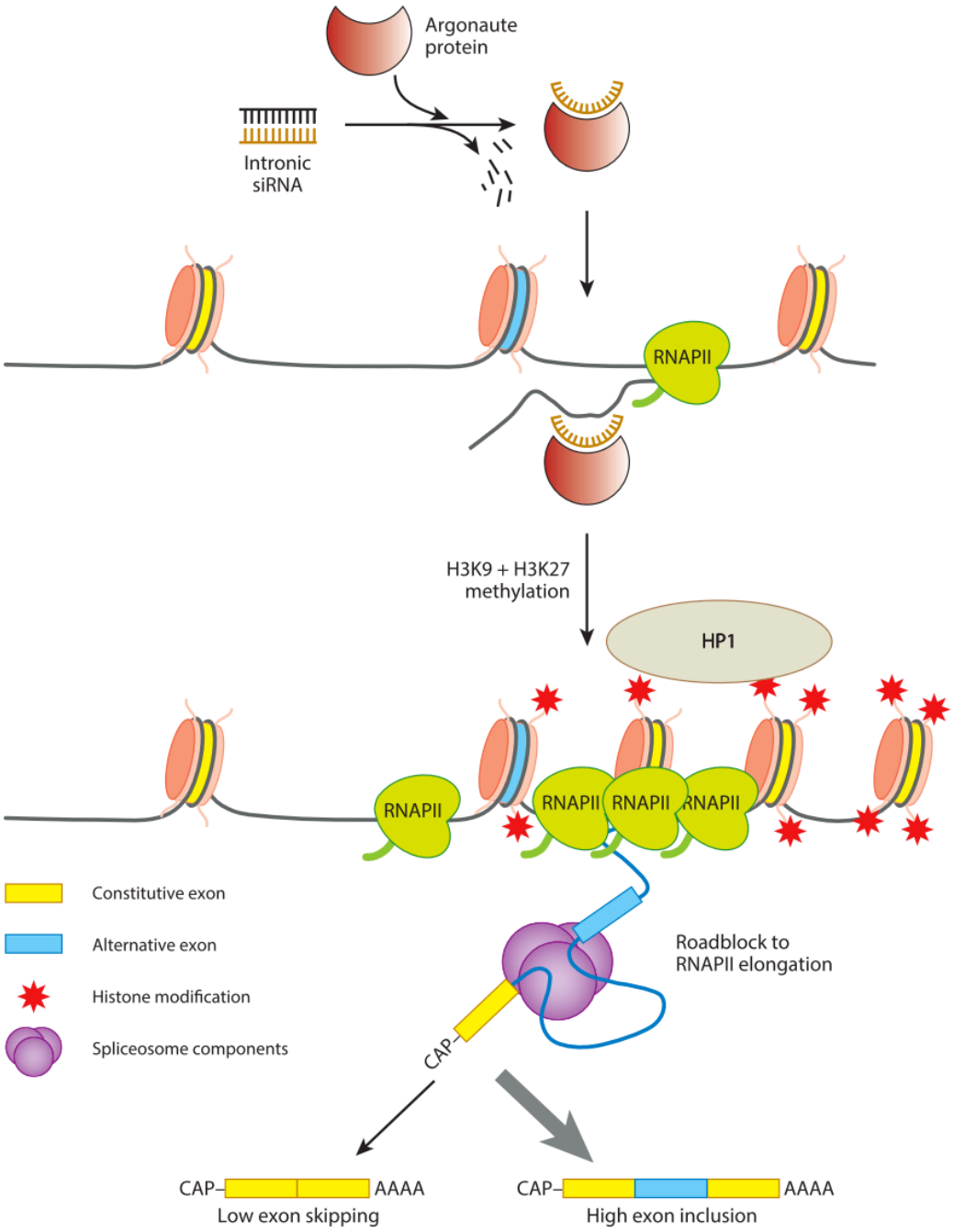


Figure 2.8 siRNA-mediated regulation of alternative splicing

Adapted with permission from Naftelberg et *al*.[38] Expressing siRNAs targeted to intronic RNA regions increases the abundance of histone methylation and acetylation markers on the DNA. These modifications act as "roadblocks" for RNA polymerase elongation, promoting an increase in alternative exon inclusion.

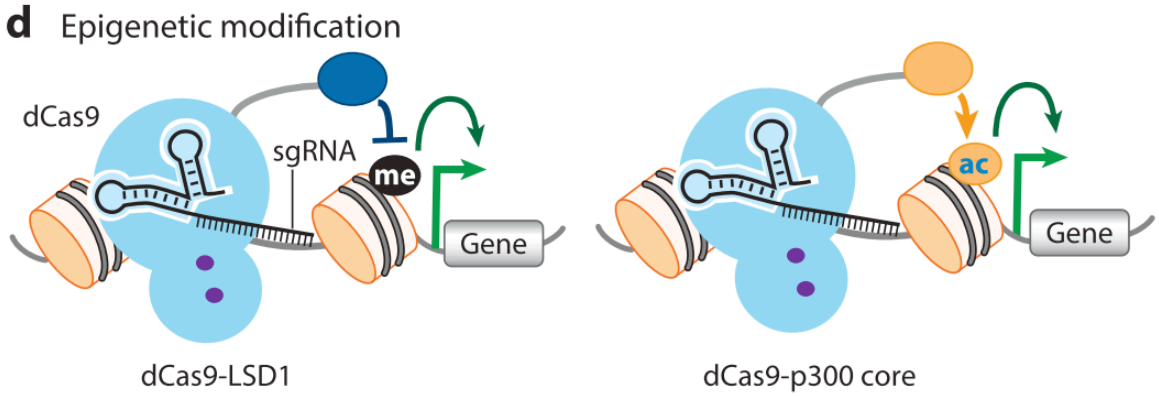


Figure 2.9 dCas9-HMP fusions for gene regulation
Adapted with permission from Wang et *al*.[70] dCas9 fusions to LSD1 and the p300 HAT core have been shown to effectively produce histone modifications that activate or repress gene expression[51,52,70].

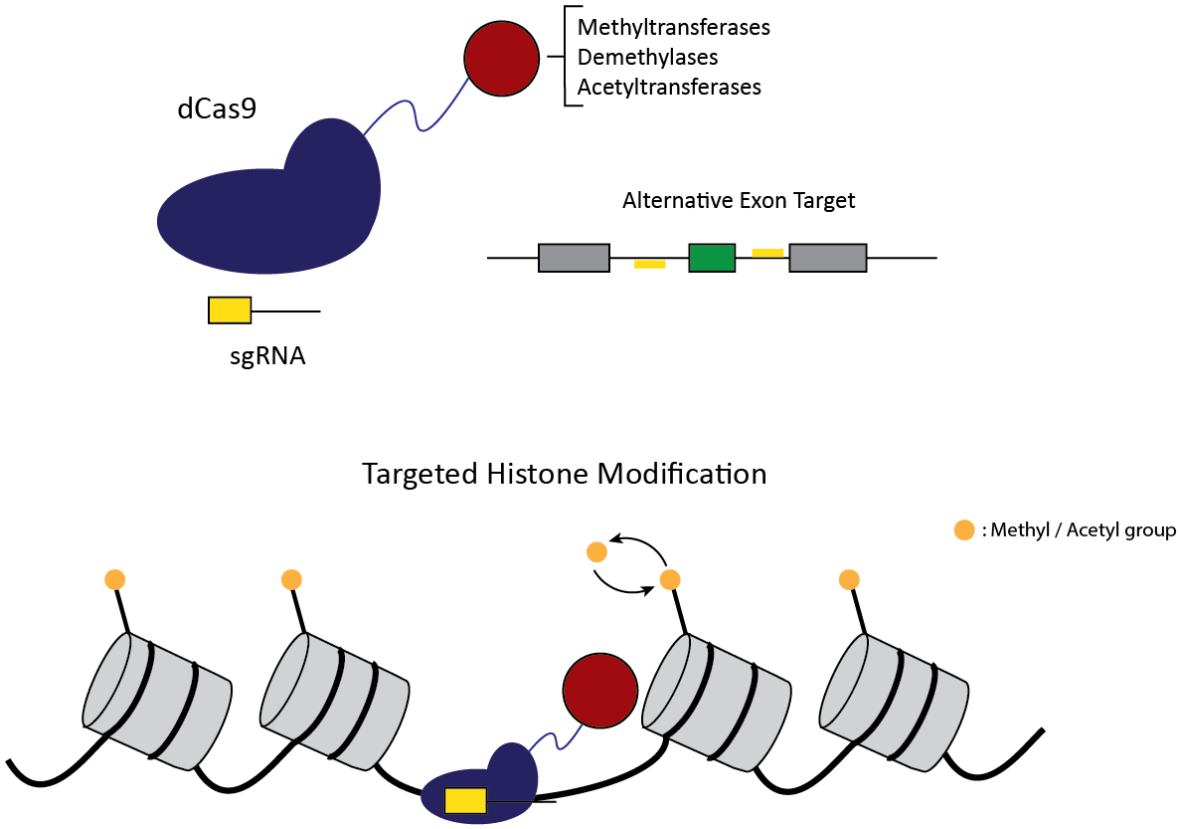


Figure 2.10 Targeted dCas9 fusions for histone modification
dCas9 (blue) fused to various histone-modifying proteins (red) paired with an sgRNA (yellow) designed to target intron flanking regions near an alternative exon target. These dCas9 fusions lead to addition / removal of modifications to alter the rate of RNA Pol II transcription and bring about changes in exon inclusion / exclusion ratios.

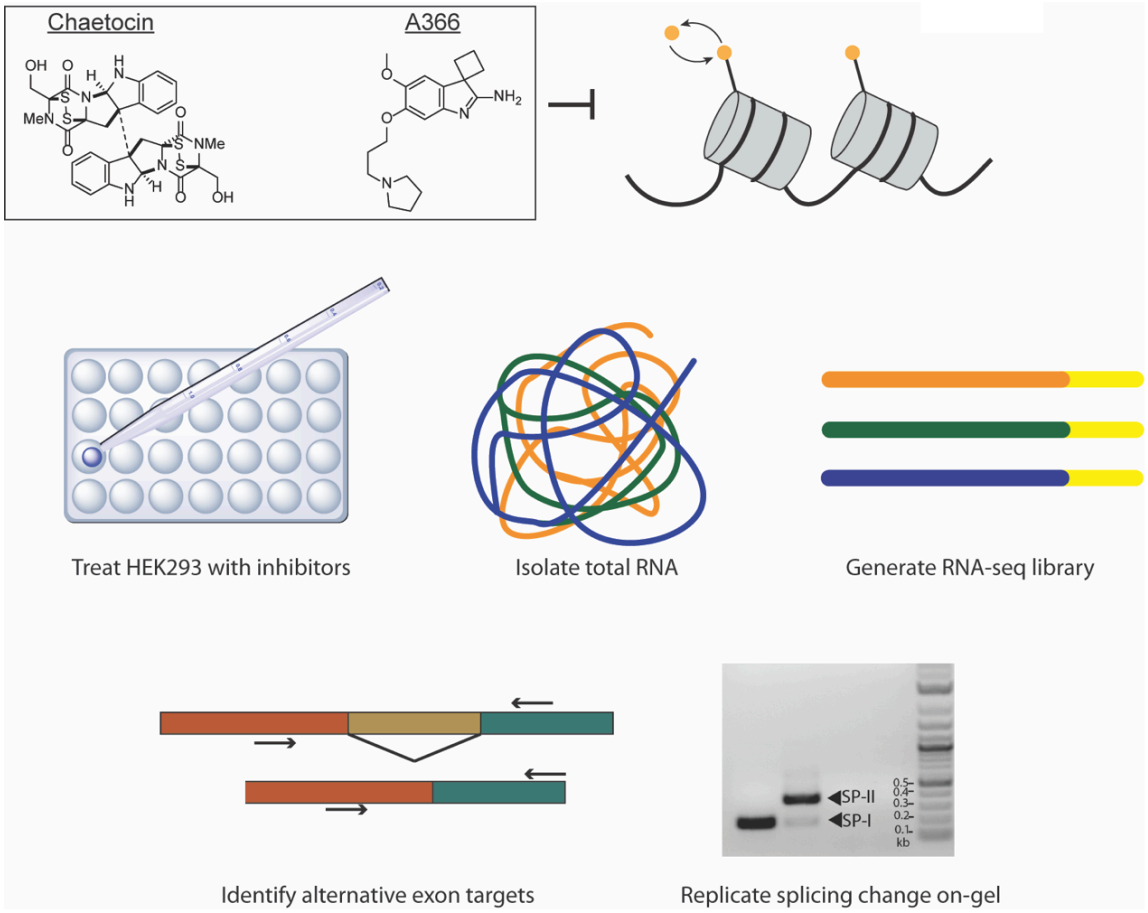


Figure 2.11 Identifying dCas9-HMP targets through RNA-seq
Schematic illustrating the protocol for RNA-seq library preparation. Cells were treated with Chaetocin or A366 to inhibit histone methyltransferase activity. After the treatment period, total cellular RNA was isolated and used to generate an RNA-seq library. The library was sent for sequencing and the treated cells were compared to a negative control to identify alternative exons with altered splicing patterns due to treatment. These exons were then targeted with dCas9-HMP fusions and analyzed through RT-PCR replicate the exon inclusion/exclusion change.

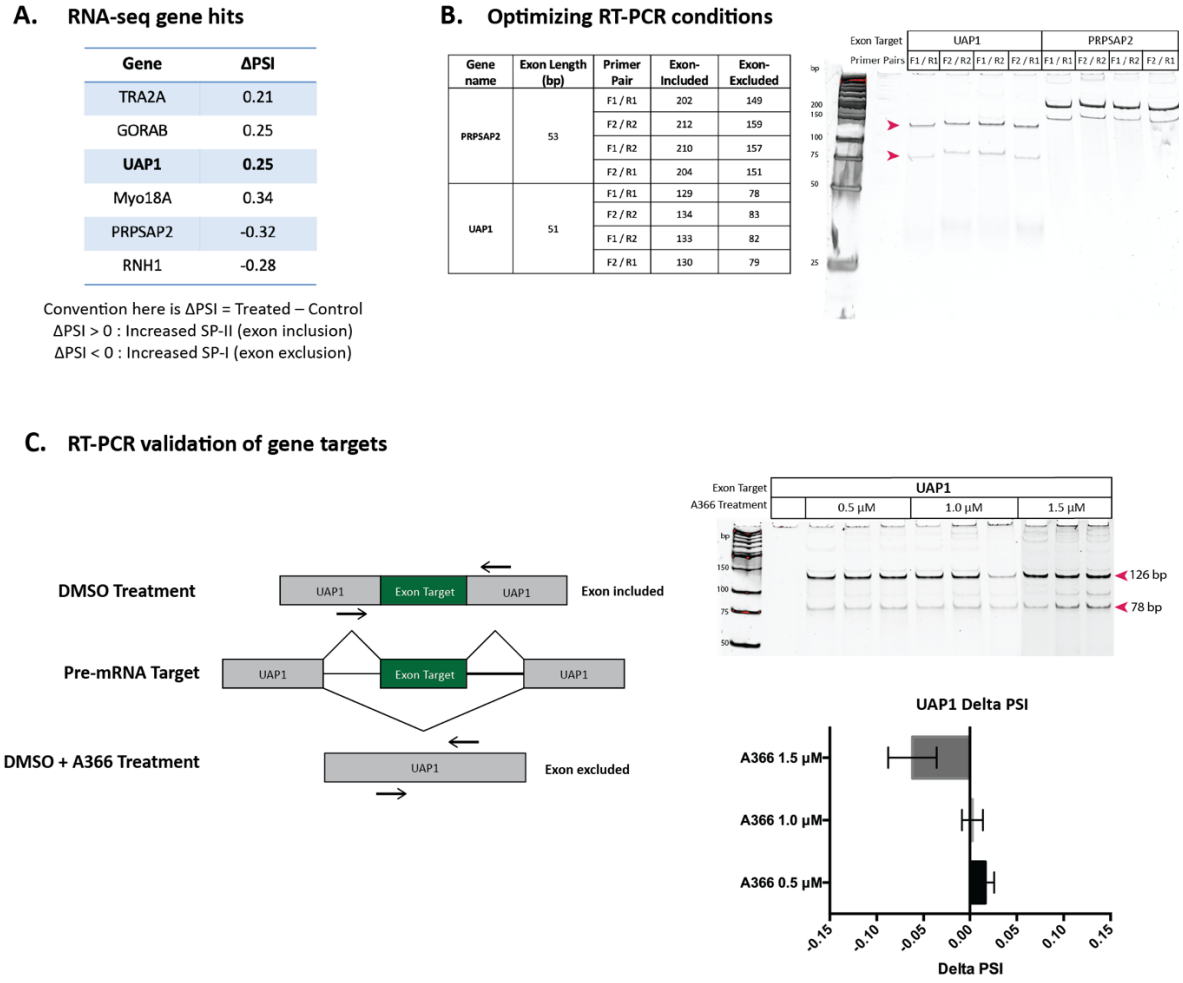
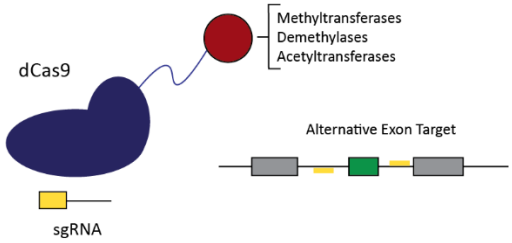


Figure 2.12 Characterizing dCas9-HMP targets from RNA-seq library

(A) Table of potential target genes containing alternative exons that were influenced by the drug treatment procedure. These hits were extracted from the RNA-seq data by collaborator Yeon Lee from the Rio lab using MISO software. Although 6 potential hits were found, only one (UAP1, in bold) was eventually confirmed through drug treatment and RT-PCR analysis. **(B)** Table showing primer candidates for RT-PCR analysis of UAP1 and PRPSAP2 exons. Example gel showing primer pair optimization for RT-PCR analysis of splice product ratios across alternative exon for two gene targets. **(C)** Schematic showing exon excluded and included PCR products with and without A366 for UAP1. The corresponding gel images and quantification are shown. UAP1 matched the RNA-seq findings but only at higher-than-predicted levels of A366 (1.5 μM compared to 0.5 μM). All other targets could not be validated through RT-PCR.

A. dCas9 fusions

Expected splicing changes for dCas9 fusions	
ΔPSI	Construct
ΔPSI > 0	**EHMT2/G9a**: H3K9 methyltransferase (110.2 kDa)
ΔPSI = ?	**P300**: H3K122 and H3K27 acetyltransferase (67.9 kDa)
ΔPSI < 0	**LSD1**: H3K4 and H3K9 demethylase (93.6 kDa)
ΔPSI = 0	**HaloTag**: (31.7 kDa)



B. Screening transfection conditions

dCas9 plasmid (A)	sgRNA plasmids (B)	Ratio A:B
90 ng	30 ng	3:1
80 ng	40 ng	2:1
60 ng	60 ng	1:1
40 ng	80 ng	1:2
30 ng	90 ng	1:3

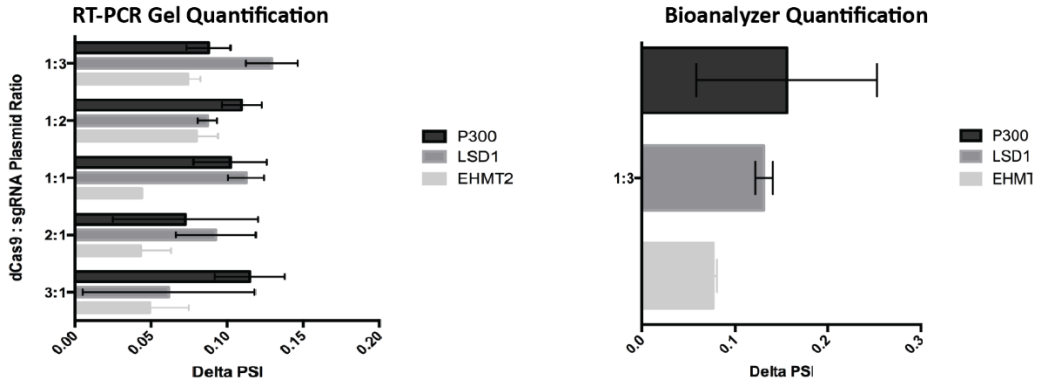


Figure 2.13 Optimizing transfection conditions for dCas9-HMP testing
(A) Expected delta PSI trends for each of the dCas9 fusions along with an illustration of the fusion system. **(B)** Different transfection conditions for dCas9-HMPs and sgRNAs along with accompanying delta PSI values for RT-PCR on-gel and Bioanalyzer analysis. The 1:3 dCas9:sgRNA ratio was chosen for future experiments.

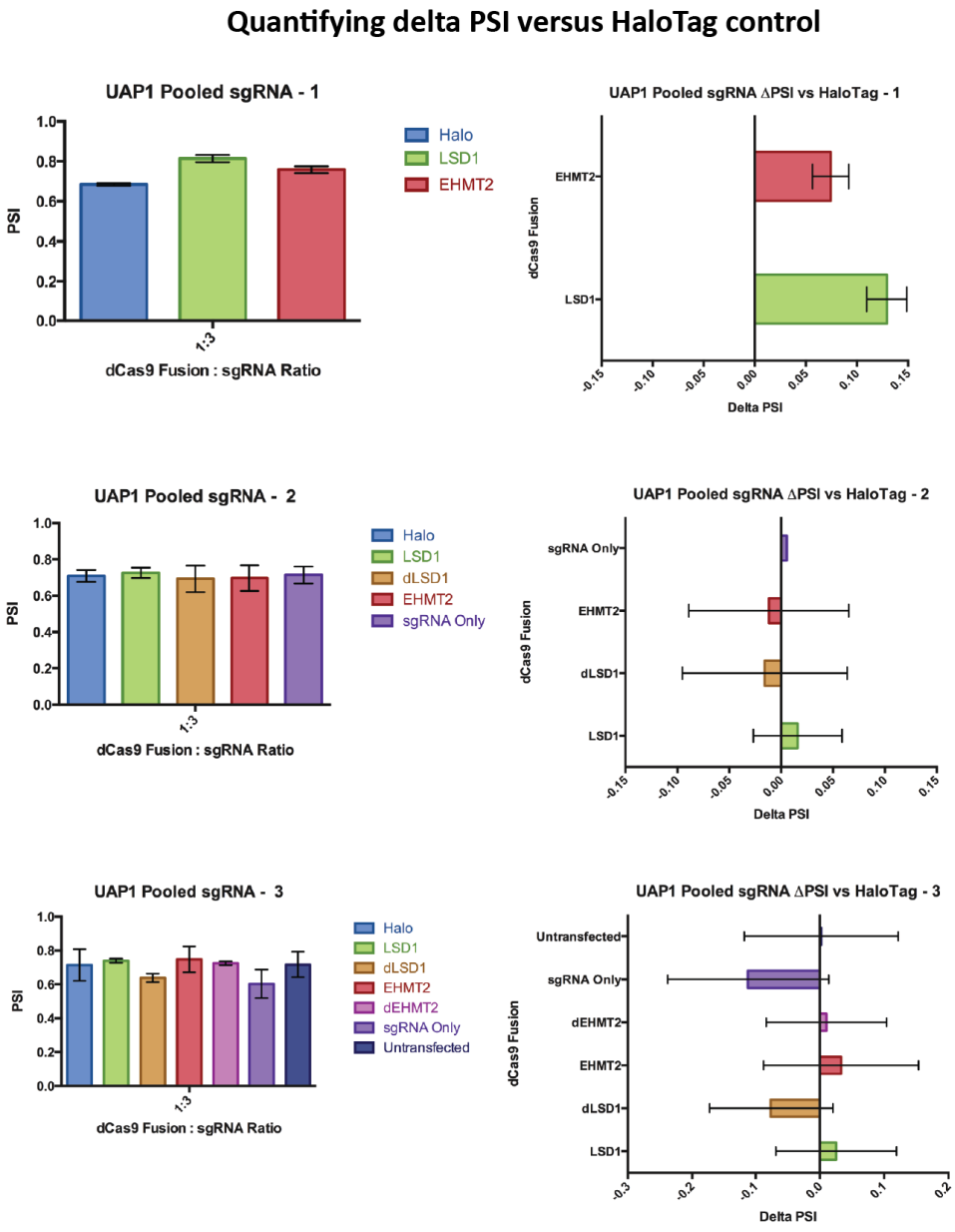


Figure 2.14 Testing dCas9-HMP fusions targeted to the UAP1 alternative exon
Quantification of PSI and delta PSI value for all fusions in three replicates. Different fusions were added in each as they were cloned. *Left:* Raw PSI value for each dCas9 fusion, *Right:* Delta PSI values for fusions compared to the dCas9-Halo construct. The data were unfortunately obscured by variation between biological replicates, leading to the eventual abandonment of this project.

2.6 Tables

Table 2.1: Construct sequences for Cas9-*Hy*P5SM project
Annotations included are: FLAG tag(orange), SV40 NLS (red), *Hy*P5SM cassette (green), introns (blue).

Construct	Sequence
Cas9	ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATG GCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCT GGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAA GGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCG AAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCC TTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTAC CACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG CTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCG ACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCA TCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATC TGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGA CCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACG ACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTC CGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATG ATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAG AAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAG GAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTG AACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGA GAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAG AAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGA CCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAG AGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGCTGCCCAAGCACAGCCTGCTGTACG AGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCC TGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAG CTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAA AACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGG CTGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGG GGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTG AAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACA TCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCG CCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAG CCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGA GAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAA ACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGG AACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTC CATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGG TCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACA ATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTG GAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAAT GACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTC CAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAA CCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGC GGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCA TGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACG GCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCC CAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAG

	GAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCA CCGTGGCCTATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAA GAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAG GGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGC CGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTG AACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTG TTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATC CTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAG GCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCAC CATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGG CCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGAC
Cas9- *Hy*P5SM	ATGGACTATAAGGACCACGACGGAGACTACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATG GCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCT GGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAA GGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCG AAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCC TTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTAC CACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGG CTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCG ACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCA TCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATC TGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGAAACCTGATTGCCCTGAGCCTGGGCCTGA CCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACG ACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTC CGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCTATG ATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAG AAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAG GAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTG AACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGA GAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAG AAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGA CCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAG AGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTAGATTTATGCATCCTCTTGT CATGAGAAGTCGAATTGTTCCCATTCTGTGTGTTGCAGCTACAGATGGAGATACATAGAGATACTCGTGGATT TTGCTTAGTGTTGAGTTTTGTTCTGGTTGTGAACTAAAAGTTTATACATTTGCAGGAAATAAATAGCCTTTTGTT TAAATCAAAAGGTCTTACCTATGTTAGTGTGAAGCATTGGATCCCAAAGAGAGAACTCCAAAATGCGATGAG GCATATTTAATCTTGTCTGGACTAGTAACAGGTTGGGATGACCACCTGTGAAGCTCCAACAGGATTGCCTCCTC ACGCAATGTTTGAGGTCTGATGTTCAATAGCTTGTTTTGTTTCACTTTGCTTTGGACTTTCTTTTCGCCAATGAG CTATGTTTCTGATGGTTTTCACTCTTTTGGTGTGTAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCA CCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCG AGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAG GACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTCAACGCCTCCC TGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACA TTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCT ATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTG AGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGAC GGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAA GCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAG AAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAA CATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGA AGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACCCAG CTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACTGGAC ATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAG AAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACC AAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCG GCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCT

	GATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTG ATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATG ATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTT TCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACC GGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAA TATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCG ATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCT ATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTG GGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAA GAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGA ATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGT ACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAAC AGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACG CTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATA TCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGG AAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG ACACGGATCGACCTGTCTCAGCTGGGAGGCGAC
pcoCas9	ATGGATTACAAGGATGATGATGATAAGGATTACAAGGATGATGATGATAAGATGGCTCCAAAGAAGAAGAG AAAGGTTGGAATCCACGGAGTTCCAGCTGCTGATAAGAAGTACTCTATCGGACTTGACATCGGAACCAACTCT GTTGGATGGGCTGTTATCACCGATGAGTACAAGGTTCCATCTAAGAAGTTCAAGGTTCTTGGAAACACCGATA GACACTCTATCAAGAAGAACCTTATCGGTGCTCTTCTTTTCGATTCTGGAGAGACCGCTGAGGCTACCAGATTG AAGAGAACCGCTAGAAGAAGATACACCAGAAGAAAGAACAGAATCTGCTACCTTCAGGAAATCTTCTCTAAC GAGATGGCTAAGGTTGATGATTCTTTCTTCCACAGACTTGAGGAGTCTTTCCTTGTTGAGGAGGATAAGAAGC ACGAGAGACACCCAATCTTCGGAAACATCGTTGATGAGGTTGCTTACCACGAGAAGTACCCAACCATCTACCA CCTTAGAAAGAAGTTGGTTGATTCTACCGATAAGGCTGATCTTAGACTTATCTACCTTGCTCTTGCTCACATGA TCAAGTTCAGAGGACACTTCCTTATCGAGGGAGACCTTAACCCAGATAACTCTGATGTTGATAAGTTGTTCATC CAGCTTGTTCAGACCTACAACCAGCTTTTCGAGGAGAACCCAATCAACGCTTCTGGAGTTGATGCTAAGGCTA TCCTTTCTGCTAGACTTTCTAAGTCTCGTAGACTTGAGAACCTTATCGCTCAGCTTCCAGGAGAGAAGAAGAAC GGACTTTTCGGAAACCTTATCGCTCTTTCTCTTGGACTTACCCCAAACTTCAAGTCTAACTTCGATCTTGCTGAG GATGCTAAGTTGCAGCTTTCTAAGGATACCTACGATGATGATCTTGATAACCTTCTTGCTCAGATCGGAGATCA GTACGCTGATCTTTTCCTTGCTGCTAAGAACCTTTCTGATGCTATCCTTCTTTCTGACATCCTTAGAGTTAACACC GAGATCACCAAGGCTCCACTTTCTGCTTCTATGATCAAGAGATACGATGAGCACCACCAGGATCTTACCCTTTT GAAGGCTCTTGTTAGACAGCAGCTTCCAGAGAAGTACAAGGAAATCTTCTTCGATCAGTCTAAGAACGGATAC GCTGGATACATCGATGGAGGAGCTTCTCAGGAGGAGTTCTACAAGTTCATCAAGCCAATCCTTGAGAAGATG GATGGAACCGAGGAGCTTCTTGTTAAGTTGAACAGAGAGGATCTTCTTAGAAAGCAGAGAACCTTCGATAAC GGATCTATCCCACACCAGATCCACCTTGGAGAGCTTCACGCTATCCTTCGTAGACAGGAGGATTTCTACCCATT CTTGAAGGATAACAGAGAGAAGATCGAGAAGATCCTTACCTTCAGAATCCCATACTACGTTGGACCACTTGCT AGAGGAAACTCTCGTTTCGCTTGGATGACCAGAAAGTCTGAGGAGACCATCACCCCTTGGAACTTCGAGGAG GTAAGTTTCTGCTTCTACCTTTGATATATATATAATAATTATCATTAATTAGTAGTAATATAATATTTCAAATATT TTTTTCAAAATAAAAGAATGTAGTATATAGCAATTGCTTTTCTGTAGTTTATAAGTGTGTATATTTTAATTTATA ACTTTTCTAATATATGACCAAAATTTGTTGATGTGCAGGTTGTTGATAAGGGAGCTTCTGCTCAGTCTTTCATC GAGAGAATGACCAACTTCGATAAGAACCTTCCAAACGAGAAGGTTCTTCCAAAGCACTCTCTTCTTTACGAGT ACTTCACCGTTTACAACGAGCTTACCAAGGTTAAGTACGTTACCGAGGGAATGAGAAAGCCAGCTTTCCTTTCT GGAGAGCAGAAGAAGGCTATCGTTGATCTTCTTTTCAAGACCAACAGAAAGGTTACCGTTAAGCAGTTGAAG GAGGATTACTTCAAGAAGATCGAGTGCTTCGATTCTGTTGAAATCTCTGGAGTTGAGGATAGATTCAACGCTT CTCTTGGAACCTACCACGATCTTTTGAAGATCATCAAGGATAAGGATTTCCTTGATAACGAGGAGAACGAGGA CATCCTTGAGGACATCGTTCTTACCCTTACCCTTTTCGAGGATAGAGAGATGATCGAGGAGAGACTCAAGACC TACGCTCACCTTTTCGATGATAAGGTTATGAAGCAGTTGAAGAGAAGAAGATACACCGGATGGGGTAGACTT TCTCGTAAGTTGATCAACGGAATCAGAGATAAGCAGTCTGGAAAGACCATCCTTGATTTCTTGAAGTCTGATG GATTCGCTAACAGAAACTTCATGCAGCTTATCCACGATGATTCTCTTACCTTCAAGGAGGACATCCAGAAGGCT CAGGTTTCTGGACAGGGAGATTCTCTTCACGAGCACATCGCTAACCTTGCTGGATCTCCAGCTATCAAGAAGG GAATCCTTCAGACCGTTAAGGTTGTTGATGAGCTTGTTAAGGTTATGGGTAGACACAAGCCAGAGAACATCGT TATCGAGATGGCTAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACTCTCGTGAGAGAATGAAGAGAA TCGAGGAGGGAATCAAGGAGCTTGGATCTCAAATCTTGAAGGAGCACCCAGTTGAGAACACCCAGCTTCAGA ACGAGAAGTTGTACCTTTACTACCTTCAGAACGGAAGAGATATGTACGTTGATCAGGAGCTTGACATCAACAG ACTTTCTGATTACGATGTTGATCACATCGTTCCACAGTCTTTCTTGAAGGATGATTCTATCGATAACAAGGTTCT

	TACCCGTTCTGATAAGAACAGAGGAAAGTCTGATAACGTTCCATCTGAGGAGGTTGTTAAGAAGATGAAGAA CTACTGGAGACAGCTTCTTAACGCTAAGTTGATCACCCAGAGAAAGTTCGATAACCTTACCAAGGCTGAGAGA GGAGGACTTTCTGAGCTTGATAAGGCTGGATTCATCAAGAGACAGCTTGTTGAGACCAGACAGATCACCAAG CACGTTGCTCAGATCCTTGATTCTCGTATGAACACCAAGTACGATGAGAACGATAAGTTGATCAGAGAGGTTA AGGTTATCACCTTGAAGTCTAAGTTGGTTTCTGATTTCAGAAAGGATTTCCAGTTCTACAAGGTTAGAGAGATC AACAACTACCACCACGCTCACGATGCTTACCTTAACGCTGTTGTTGGAACCGCTCTTATCAAGAAGTACCCAAA GTTGGAGTCTGAGTTCGTTTACGGAGATTACAAGGTTTACGATGTTAGAAAGATGATCGCTAAGTCTGAGCAG GAGATCGGAAAGGCTACCGCTAAGTACTTCTTCTACTCTAACATCATGAACTTCTTCAAGACCGAGATCACCCT TGCTAACGGAGAGATCAGAAAGAGACCACTTATCGAGACCAACGGAGAGACCGGAGAGATCGTTTGGGATA AGGGAAGAGATTTCGCTACCGTTAGAAAGGTTCTTTCTATGCCACAGGTTAACATCGTTAAGAAAACCGAGGT TCAGACCGGAGGATTCTCTAAGGAGTCTATCCTTCCAAAGAGAAACTCTGATAAGTTGATCGCTAGAAAGAAG GATTGGGACCCAAAGAAGTACGGAGGATTCGATTCTCCAACCGTTGCTTACTCTGTTCTTGTTGTTGCTAAGGT TGAGAAGGGAAAGTCTAAGAAGTTGAAGTCTGTTAAGGAGCTTCTTGGAATCACCATCATGGAGCGTTCTTCT TTCGAGAAGAACCCAATCGATTTCCTTGAGGCTAAGGGATACAAGGAGGTTAAGAAGGATCTTATCATCAAGT TGCCAAAGTACTCTCTTTTCGAGCTTGAGAACGGAAGAAAGAGAATGCTTGCTTCTGCTGGAGAGCTTCAGAA GGGAAACGAGCTTGCTCTTCCATCTAAGTACGTTAACTTCCTTTACCTTGCTTCTCACTACGAGAAGTTGAAGG GATCTCCAGAGGATAACGAGCAGAAGCAGCTTTTCGTTGAGCAGCACAAGCACTACCTTGATGAGATCATCG AGCAAATCTCTGAGTTCTCTAAGAGAGTTATCCTTGCTGATGCTAACCTTGATAAGGTTCTTTCTGCTTACAACA AGCACAGAGATAAGCCAATCAGAGAGCAGGCTGAGAACATCATCCACCTTTTCACCCTTACCAACCTTGGTGC TCCAGCTGCTTTCAAGTACTTCGATACCACCATCGATAGAAAAAGATACACCTCTACCAAGGAGGTTCTTGATG CTACCCTTATCCACCAGTCTATCACCGGACTTTACGAGACCAGAATCGATCTTTCTCAGCTTGGAGGAGATAAG AGACCAGCTGCTACCAAGAAGGCTGGACAGGCTAAGAAGAAGAAG

Table 2.2: Construct sequences for dCas9-HMP project

Annotations included are: FLAG tag(orange),NLS (red), HMP (green), mutation site(highlighted dLSD1 K661A and dEHMT2 C959A).

Construct	Sequence
dCas9-HaloTag	ATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGT ACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAG CAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTC GACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAA CCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTG GAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAG GTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCG ACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTG AACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAA ACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG AAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGG CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTAC GACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACC TGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTG AGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCA GGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTG AACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGA GAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGA AGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACC AGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGC TTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGT ACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTT CCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCA GCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAA ACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGC TGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGG GCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGA AGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATC CAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCC ATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAG AATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACT

	GGACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCG ACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTG AAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGA CCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCC GGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGC TGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGA TCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGA TCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGG GAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATC GTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAG CTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTG TGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATC ACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGA AAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGC CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCA GCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGC ACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGA CAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACAC CAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCAC<u>CGGCCTGTACGAGACACGGATCGAC</u> <u>CTGTCTCAGCTGGGAGGCGAC</u>AGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCGGATCCGGTTCCGGACG GGCTGGCGGAGGTAGCGGCGGTGGATCAGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGGCA GAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGT TGGTCCGCGCGATGGCACCCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCA TCCCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGAC CTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGT CCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGT ATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCA GGCCTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGA TGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTTGACCGC GAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAG AATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCA CCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATC TGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCTACTCTG<u>GAGATTTCCGG</u> TGATCCAAAAAAGAAGAGAAAGGTACGTACCGGTTGA
dCas9-EHMT2	ATGTACCCATACGATGTTCCAGATTACGCTTCG<u>CCGAAGAAAAAGCGCAAGGTC</u>GAAGCGTCCGACAAGAAGT ACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAG CAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTC GACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAA CCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTG GAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAG GTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCG ACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTG AACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAA ACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG AAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGG CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTAC GACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACC TGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTG AGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCA GGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTG AACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGA GAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGA AGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACC AGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGC

	TTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGT ACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTT CCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCA GCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAA ACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGC TGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGG GCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGA AGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATC CAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCC ATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAG AATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACT GGACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCG ACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTG AAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGA CCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCC GGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGC TGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGA TCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGA TCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGG GAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATC GTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAG CTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTG TGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATC ACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGA AAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGC CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCA GCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGC ACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGA CAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACAC CAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGAC CTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCNGCGGATCCATGAGTGATGA TGTCCACTCACTGGGAAAGGTGACCTCAGATCTGGCCAAAAGGAGGAAGCTGAACTCAGGAGGTGGCCTGTC AGAGGAGTTAGGTTCTGCCCGGCGTTCAGGAGAAGTGACCCTGACGAAAGGGGACCCCGGGTCCCTGGAGGA GTGGGAGACGGTGGTGGGTGATGACTTCAGTCTCTACTATGATTCCTACTCTGTGGATGAGCGCGTGGACTCC GACAGCAAGTCTGAAGTTGAAGCTCTAACTGAACAACTAAGTGAAGAGGAGGAGGAGGAAGAGGAGGAAGA AGAAGAAGAGGAAGAGGAGGAGGAAGAGGAAGAAGAAGAGGAAGATGAGGAGTCAGGGAATCAGTCAGA TAGGAGTGGTTCCAGTGGCCGGCGCAAGGCCAAGAAGAAATGGCGAAAAGACAGCCCATGGGTGAAGCCGT CTCGGAAACGGCGCAAGCGGGAGCCTCCGCGGGCCAAGGAGCCACGAGGAGTGAATGGTGTGGGCTCCTCA GGCCCCAGTGAGTACATGGAGGTCCCTCTGGGGTCCCTGGAGCTGCCCAGCGAGGGGACCCTCTCCCCCAACC ACGCTGGGGTGTCCAATGACACATCTTCGCTGGAGACAGAGCGAGGGTTTGAGGAGTTGCCCCTGTGCAGCTG CCGCATGGAGGCACCCAAGATTGACCGCATCAGCGAGAGGGCGGGGCACAAGTGCATGGCCACTGAGAGTGT GGACGGAGAGCTGTCAGGCTGCAATGCCGCCATCCTCAAGCGGGAGACCATGAGGCCATCCAGCCGTGTGGC CCTGATGGTGCTCTGTGAGACCCACCGCGCCCGCATGGTCAAACACCACTGCTGCCCGGGCTGCGGCTACTTCT GCACGGCGGGCACCTTCCTGGAGTGCCACCCTGACTTCCGTGTGGCCCACCGCTTCCACAAGGCCTGTGTGTCT CAGCTGAATGGGATGGTCTTCTGTCCCCACTGTGGGGAGGATGCTTCTGAAGCTCAAGAGGTGACCATCCCCC GGGGTGACGGGGTGACCCCACCGGCCGGCACTGCAGCTCCTGCACCCCCACCCCTGTCCCAGGATGTCCCCGG GAGAGCAGACACTTCTCAGCCCAGTGCCCGGATGCGAGGGCATGGGGAACCCCGGCGCCCGCCCTGCGATCC CCTGGCTGACACCATTGACAGCTCAGGGCCCTCCCTGACCCTGCCCAATGGGGGCTGCCTTTCAGCCGTGGGGC TGCCACTGGGGCCAGGCCGGGAGGCCCTGGAAAAGGCCCTGGTCATCCAGGAGTCAGAGAGGCGGAAGAAG CTCCGTTTCCACCCTCGGCAGTTGTACCTGTCCGTGAAGCAGGGCGAGCTGCAGAAGGTGATCCTGATGCTGTT GGACAACCTGGACCCCAACTTCCAGAGCGACCAGCAGAGCAAGCGCACGCCCCTGCATGCAGCCGCCCAGAA GGGCTCCGTGGAGATCTGCCATGTGCTGCTGCAGGCTGGAGCCAACATAAATGCAGTGGACAAACAGCAGCG

	GACGCCACTGATGGAGGCCGTGGTGAACAACCACCTGGAGGTAGCCCGTTACATGGTGCAGCGTGGTGGCTG TGTCTATAGCAAGGAGGAGGACGGTTCCACCTGCCTCCACCACGCAGCCAAAATCGGGAACTTGGAGATGGTC AGCCTGCTGCTGAGCACAGGACAGGTGGACGTCAACGCCCAGGACAGTGGGGGGTGGACGCCCATCATCTGG GCTGCAGAGCACAAGCACATCGAGGTGATCCGCATGCTACTGACGCGGGGCGCCGACGTCACCCTCACTGACA ACGAGGAGAACATCTGCCTGCACTGGGCCTCCTTCACGGGCAGCGCCGCCATCGCCGAAGTCCTTCTGAATGC GCGCTGTGACCTCCATGCTGTCAACTACCATGGGGACACCCCCCTGCACATCGCAGCTCGGGAGAGCTACCATG ACTGCGTGCTGTTATTCCTGTCACGTGGGGCCAACCCTGAGCTGCGGAACAAAGAGGGGGACACAGCATGGG ACCTGACTCCCGAGCGCTCCGACGTGTGGTTTGCGCTTCAACTCAACCGCAAGCTCCGACTTGGGGTGGGAAAT CGGGCCATCCGCACAGAGAAGATCATCTGCCGGGACGTGGCTCGGGGCTATGAGAACGTGCCCATTCCCTGTG TCAACGGTGTGGATGGGGAGCCCTGCCCTGAGGATTACAAGTACATCTCAGAGAACTGCGAGACGTCCACCAT GAACATCGATCGCAACATCACCCACCTGCAGCACTGCACGTGTGTGGACGACTGCTCTAGCTCCAACTGCCTGT GCGGCCAGCTCAGCATCCGGTGCTGGTATGACAAGGATGGGCGATTGCTCCAGGAATTTAACAAGATTGAGCC TCCGCTGATTTTCGAGTGTAACCAGGCGTGCTCATGCTGGAGAAACTGCAAGAACCGGGTCGTACAGAGTGGC ATCAAGGTGCGGCTACAGCTCTACCGAACAGCCAAGATGGGCTGGGGGGTCCGCGCCCTGCAGACCATCCCAC AGGGGACCTTCATCTGCGAGTATGTCGGGGAGCTGATCTCTGATGCTGAGGCTGATGTGAGAGAGGATGATTC TTACCTCTTCGACTTAGACAACAAGGATGGAGAGGTGTACTGCATAGATGCCCGTTACTATGGCAACATCAGCC GCTTCATCAACCACCTGTGTGACCCCAACATCATTCCCGTCCGGGTCTTCATGCTGCACCAAGACCTGCGATTTC CACGCATCGCCTTCTTCAGTTCCCGAGACATCCGGACTGGGGAGGAGCTAGGGTTTGACTATGGCGACCGCTTC TGGGACATCAAAAGCAAATATTTCACCTGCCAATGTGGCTCTGAGAAGTGCAAGCACTCAGCCGAAGCCATTG CCCTGGAGCAGAGCCGTCTGGCCCGCCTGGACCCACACCCTGAGCTGCTGCCCGAGCTCGGCTCCCTGCCCCCT GTCAACACATGA
dCas9-LSD1	ATGTACCCATACGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGT ACAGCATCGGCCTGGCCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAAGGTGCCCAG CAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTC GACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAA CCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTG GAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAG GTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCG ACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTG AACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAA ACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGG AAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGCCCTGAGCCTGGG CCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTAC GACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACC TGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTG AGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCA GGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTG AACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGA GAGCTGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGA AGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACC AGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGC TTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGT ACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTT CCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCA GCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAA ACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGC TGAAAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGG GCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGA AGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATC CAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCC ATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCC GAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAG AATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACAC CCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGGACCAGGAACT GGACATCAACCGGCTGTCCGACTACGATGTGGACGCTATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCG ACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTG

	AAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGA CCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCC GGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGC TGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTAC AAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGA TCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGATGA TCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCATGAACTTTTTC AAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGG GAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATC GTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAG CTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTG TGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATC ACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGA AAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGC CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCA GCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGC ACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGA CAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTG TTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACAC CAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGAC CTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCGGATCCTTATCTGGGAAG AAGGCGGCAGCCGCGGCGGCGGCGGCTGCAGCGGCAGCAACCGGGACGGAGGCTGGCCCTGGGACAGCAG GCGGCTCCGAGAACGGGTCTGAGGTGGCCGCGCAGCCCGCGGGCCTGTCGGGCCCAGCCGAGGTCGGGCCG GGGGCGGTGGGGGAGCGCACACCCCGCAAGAAAGAGCCTCCGCGGGCCTCGCCCCCCGGGGGCCTGGCGGA ACCGCCGGGGTCCGCAGGGCCTCAGGCCGGCCCTACTGTCGTGCCTGGGTCTGCGACCCCCATGGAAACTGGA ATAGCAGAGACTCCGGAGGGGCGTCGGACCAGCCGGCGCAAGCGGGCGAAGGTAGAGTACAGAGAGATGGA TGAAAGCTTGGCCAACCTCTCAGAAGATGAGTATTATTCAGAAGAAGAGAGAAATGCCAAAGCAGAGAAGGA AAAGAAGCTTCCCCCACCACCCCCTCAAGCCCCACCTGAGGAAGAAAATGAAAGTGAGCCTGAAGAACCATCG GGTGTGGAGGGCGCAGCTTTCCAGAGCCGACTTCCTCATGACCGGATGACTTCTCAAGAAGCAGCCTGTTTTCC AGATATTATCAGTGGACCACAACAGACCCAGAAGGTTTTTCTTTTCATTAGAAACCGCACACTGCAGTTGTGGT TGGATAATCCAAAGATTCAGCTGACATTTGAGGCTACTCTCCAACAATTAGAAGCACCTTATAACAGTGATACT GTGCTTGTCCACCGAGTTCACAGTTATTTAGAGCGTCATGGTCTTATCAACTTCGGCATCTATAAGAGGATAAA ACCCCTACCAACTAAAAAGACAGGAAAGGTAATTATTATAGGCTCTGGGGTCTCAGGCTTGGCAGCAGCTCGA CAGTTACAAAGTTTTGGAATGGATGTCACACTTTTGGAAGCCAGGGATCGTGTGGGTGGACGAGTTGCCACAT TTCGCAAAGGAAACTATGTAGCTGATCTTGGAGCCATGGTGGTAACAGGTCTTGGAGGGAATCCTATGGCTGT GGTCAGCAAACAAGTAAATATGGAACTGGCCAAGATCAAGCAAAAATGCCCACTTTATGAAGCCAACGGACAA GCTGTTCCTAAAGAGAAAGATGAAATGGTAGAGCAAGAGTTTAACCGGTTGCTAGAAGCTACATCTTACCTTA GTCATCAACTAGACTTCAATGTCCTCAATAATAAGCATGTGTCCCTTGGCCAGGCATTGGAAGTTGTCATTCAGT TACAAGAGAAGCATGTCAAAGATGAGCAGATTGAACATTGGAAGAAGATAGTGAAAACTCAGGAAGAATTGA AAGAACTTCTTAATAAGATGGTAAATTTGAAAGAGAAAATTAAAGAACTCCATCAGCAATACAAAGAAGCATCT GAAGTAAAGCCACCCAGAGATATTACTGCCGAGTTCTTAGTGAAAAGCAAACACAGGGATCTGACCGCCCTAT GCAAGGAATATGATGAATTAGCTGAAACACAAGGAAAGCTAGAAGAAAAACTTCAGGAGTTGGAAGCGAATC CCCCAAGTGATGTATATCTCTCATCAAGAGACAGACAAATACTTGATTGGCATTTTGCAAATCTTGAATTTGCTA ATGCCACACCTCTCTCAACTCTCTCCCTTAAGCACTGGGATCAGGATGATGACTTTGAGTTCACTGGCAGCCACC TGACAGTAAGGAATGGCTACTCGTGTGTGCCTGTGGCTTTAGCAGAAGGCCTAGACATTAAACTGAATACAGC AGTGCGACAGGTTCGCTACACGGCTTCAGGATGTGAAGTGATAGCTGTGAATACCCGCTCCACGAGTCAAACC TTTATTTATAAATGCGACGCAGTTCTCTGTACCCTTCCCCTGGGTGTGCTGAAGCAGCAGCCACCAGCCGTTCAG TTTGTGCCACCTCTCCCTGAGTGGAAAACATCTGCAGTCCAAAGGATGGGATTTGGCAACCTTAACAAGGTGGT GTTGTGTTTTGATCGGGTGTTCTGGGATCCAAGTGTCAATTTGTTCGGGCATGTTGGCAGTACGACTGCCAGCA GGGGTGAGCTCTTCCTCTTCTGGAACCTCTATAAAGCTCCAATACTGTTGGCACTAGTGGCAGGAGAAGCTGCT GGTATCATGGAAAACATAAGTGACGATGTGATTGTTGGCCGATGCCTGGCCATTCTCAAAGGGATTTTTGGTA GCAGTGCAGTACCTCAGCCCAAAGAAACTGTGGTGTCTCGTTGGCGTGCTGATCCCTGGGCTCGGGGCTCTTAT TCCTATGTTGCTGCAGGATCATCTGGAAATGACTATGATTTAATGGCTCAGCCAATCACTCCTGGCCCCTCGATT CCAGGTGCCCCACAGCCGATTCCACGACTCTTCTTTGCGGGAGAACATACGATCCGTAACTACCCAGCCACAGT GCATGGTGCTCTGCTGAGTGGGCTGCGAGAAGCGGGAAGAATTGCAGACCAGTTTTTGGGGGCCATGTATAC GCTGCCTCGCCAGGCCACACCAGGTGTTCCTGCACAGCAGTCCCCAAGCATG

dCas9-P300	ATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGG CCCCCAAGAAGAAGAGGAAGGTGGGCCGCGGAATGGATAAGAAATACTCAATAGGCTTAGCTATCGGCACAA ATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATAC AGACCGCCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACT CGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTC AAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAG AAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTAT CATCTGCGAAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATAT GATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTAT CCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGA TTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAAT GGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAA GATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCA ATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACT GAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTA AAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGC AGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATG GTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTC TATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAA AGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCA ATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGAAGAAGTTGTCGAT AAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTA CCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGG AATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAA AAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGA GTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTG GATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGAT TGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATA CTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGAT TTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAA GACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCC TGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAG CCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGC GTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATAC TCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGGACCAAGAATTAG ATATTAATCGTTTAAGTGATTATGATGTCGATGCCATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACA ATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAA GATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAG CTGAACGTGGAGGTTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATC ACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGA GGTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGA GATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCC AAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGC AAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACAC TTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAA AGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTA CAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAG ACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTG GAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCT TTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTA CCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAG GAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGT AGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGC AAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAAC ATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAATCTTGGAGCTCCC GCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACT CTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACCCAATTGCC GGATCCAAGGCTAGCCCGAAAAAGAAACGCAAAGTTGGGCGCGCCATTTTCAAACCAGAAGAACTACGACAG

	GCACTGATGCCAACTTTGGAGGCACTTTACCGTCAGGATCCAGAATCCCTTCCCTTTCGTCAACCTGTGGACCCT CAGCTTTTAGGAATCCCTGATTACTTTGATATTGTGAAGAGCCCCATGGATCTTTCTACCATTAAGAGGAAGTTA GACACTGGACAGTATCAGGAGCCCTGGCAGTATGTCGATGATATTTGGCTTATGTTCAATAATGCCTGGTTATA TAACCGGAAAACATCACGGGTATACAAATACTGCTCCAAGCTCTCTGAGGTCTTTGAACAAGAAATTGACCCAG TGATGCAAAGCCTTGGATACTGTTGTGGCAGAAAGTTGGAGTTCTCTCCACAGACACTGTGTTGCTACGGCAAA CAGTTGTGCACAATACCTCGTGATGCCACTTATTACAGTTACCAGAACAGGTATCATTTCTGTGAGAAGTGTTTC AATGAGATCCAAGGGGAGAGCGTTTCTTTGGGGGATGACCCTTCCCAGCCTCAAACTACAATAAATAAAGAAC AATTTTCCAAGAGAAAAAATGACACACTGGATCCTGAACTGTTTGTTGAATGTACAGAGTGCGGAAGAAAGAT GCATCAGATCTGTGTCCTTCACCATGAGATCATCTGGCCTGCTGGATTCGTCTGTGATGGCTGTTTAAAGAAAA GTGCACGAACTAGGAAAGAAAATAAGTTTTCTGCTAAAAGGTTGCCATCTACCAGACTTGGCACCTTTCTAGAG AATCGTGTGAATGACTTTCTGAGGCGACAGAATCACCCTGAGTCAGGAGAGGTCACTGTTAGAGTAGTTCATG CTTCTGACAAAACCGTGGAAGTAAAACCAGGCATGAAAGCAAGGTTTGTGGACAGTGGAGAGATGGCAGAAT CCTTTCCATACCGAACCAAAGCCCTCTTTGCCTTTGAAGAAATTGATGGTGTTGACCTGTGCTTCTTTGGCATGC ATGTTCAAGAGTATGGCTCTGACTGCCCTCCACCCAACCAGAGGAGAGTATACATATCTTACCTCGATAGTGTT CATTTCTTCCGTCCTAAATGCTTGAGGACTGCAGTCTATCATGAAATCCTAATTGGATATTTAGAATATGTCAAG AAATTAGGTTACACAACAGGGCATATTTGGGCATGTCCACCAAGTGAGGGAGATGATTATATCTTCCATTGCCA TCCTCCTGACCAGAAGATACCCAAGCCCAAGCGACTGCAGGAATGGTACAAAAAAATGCTTGACAAGGCTGTA TCAGAGCGTATTGTCCATGACTACAAGGATATTTTTAAACAAGCTACTGAAGATAGATTAACAAGTGCAAAGGA ATTGCCTTATTTCGAGGGTGATTTCTGGCCCAATGTTCTGGAAGAAAGCATTAAGGAACTGGAACAGGAGGAA GAAGAGAGAAAACGAGAGGAAAACACCAGCAATGAAAGCACAGATGTGACCAAGGGAGACAGCAAAAATGC TAAAAAGAAGAATAATAAGAAAACCAGCAAAAATAAGAGCAGCCTGAGTAGGGGCAACAAGAAGAAACCCG GGATGCCCAATGTATCTAACGACCTCTCACAGAAACTATATGCCACCATGGAGAAGCATAAAGAGGTCTTCTTT GTGATCCGCCTCATTGCTGGCCCTGCTGCCAACTCCCTGCCTCCCATTGTTGATCCTGATCCTCTCATCCCCTGCG ATCTGATGGATGGTCGGGATGCGTTTCTCACGCTGGCAAGGGACAAGCACCTGGAGTTCTCTTCACTCCGAAG AGCCCAGTGGTCCACCATGTGCATGCTGGTGGAGCTGCACACGCAGAGCCAGGACTACCCGTACGACGTTCCG GACTACGCTTCTTGA

Table 2.3: UAP1 target exon /flanking intron sequence and sgRNAs

Annotations included are: exon(orange), intron (blue), sgRNA DNase hypersensitive target region (highlighted). Only a portion of 5' flanking intron is shown.

Construct	Sequence
UAP1 sgRNA S1	CCACATAGCAGTGCTACAAA
UAP1 sgRNA S2	CACATAGCAGTGCTACAAAT
UAP1 sgRNA AS1	CATTTGTAGCACTGCTATGT
UAP1 sgRNA AS2	CCATTTGTAGCACTGCTATG
UAP1 pre-mRNA	CCTTATTTATTCCCACATAGCAGTGCTACAAATGGGAAGTCAGAGACCATC ACAGCTGATGTCAATCACAA

3 Inducible RNAi systems in microgravity

Portions of this chapter have been published in:

Kitto, R. Z.*, Dhillon, Y. S.*, Bevington, J., Giegé, P., Guo, H. S., Chua, N., McKay, C. & Hammond, M. C. Demonstration of a CubeSat growth system for inducible protein expression in plants under microgravity conditions. *Manuscript in preparation*.

3.1 Introduction

Plant growth in a variety of systems and species has been studied in great detail both on the International Space Station (ISS) and in simulated microgravity.[71–87] Food crops, such as lettuce or barley, are grown under different conditions to produce optimal plant size and nutritional content.[76,77,86,87] Samples from plants grown on the ISS are also shipped back to Earth and analyzed to determine if higher levels of radiation affect gene expression and nutritional content. Transgenic plants have been grown on the ISS, but primarily to support studies on growth optimization. One example is a recent experiment that focused on using a transgenic line of *Arabidopsis thaliana* to express GFP and track root development in microgravity.[78]

Current efforts focus heavily on improving plant growth and verifying that samples are safe for human consumption. As a result, there is no precedent for testing transgenic plants that are engineered to provide additional resources, such as pharmaceutical or fuel compounds.[88] One could envision that a plant could fulfill its typical role, recycling carbon dioxide, until a particular protein was required for medical treatment. At that point, a crew member could turn on protein production in the plant tissue, ideally through a non-invasive method such as light induction. The crew member could then consume the necessary drug compound with his/her typical plant-based meal. This inducible expression system would allow missions to transport a single transgenic plant species onboard with the potential to produce a variety of compounds when triggered with the correct inducer.

While the eventual goal of transgenic plant growth in space is on-demand protein expression, there are currently no published studies that have validated traditional chemically-induced expression in microgravity. However, experiments have shown that plant gene expression profiles are markedly different in a space environment and stress-related pathways are upregulated in many species.[72,73,81] A typical protein induction experiment on Earth often requires highly optimized, stress-free conditions to produce maximum protein production. Therefore, the logical first step was to develop an experiment to test whether a plant-based system can undergo inducible protein expression in a space environment.

The ultimate goal of this project was to develop a self-contained, remotely accessible growth apparatus that could fit easily on the back of a lunar or planetary rover. The unit would be sealed against the vacuum of a space environment, limiting plant growth to a short window where carbon dioxide levels were sufficient for development. Our team developed an airtight, autonomously-run system that could track plant growth and inducible protein expression in real time. The CubeSat system was the ideal hardware for our purpose, providing a microenvironment for the plants to grow while allowing customization of the lid to install camera and lighting components for remote monitoring.

Here we report the development and optimization of a CubeSat system for tracking transgenic and wild-type *Arabidopsis thaliana* growth in the presence of a chemical inducer under spaceflight conditions. A chemical induction method was chosen due to the fact that these systems are frequently used for high levels of protein expression *in planta* and are well-

characterized under normal surface conditions. This system was launched on SpaceX CRS-16 from the Kennedy Space Center on December 5th, 2018. During the course of an experiment that ran for 14 days on the ISS, we showed through real-time observation of physical characteristics and later analysis of genetic elements that one can trigger inducible protein expression in a transgenic plant line under the stresses of microgravity. This result will hopefully pave the way for future development of transgenic plants that can serve a practical function aboard the ISS or exploratory spacecraft. It also demonstrates the functionality of the CubeSat hardware for future plant experiments, potentially in extraterrestrial habitats.

3.2 Results

The initial focus of the project was developing a functional set of hardware that could support plant growth for at least 14 days in an airtight chamber. Previous plant growth hardware designs on the ISS are typically either open-air or arranged to allow gas exchange through a permeable membrane.[89] As a result, our team conducted extensive materials testing to select hardware that would not inhibit healthy plant growth.

The overall CubeSat design consisted of a plant growth cassette, metal chamber, and lid (Figure 3.1). Stage 0 testing was with the plant cassette alone to select a compatible material for direct contact with the seedlings. Next, Stage 1 development tested a covered system with the plant cassette in an aluminum chamber covered with a clear polyvinyl chloride (PVC) lid. Growth lighting was placed outside the chamber and pictures were taken from above the PVC lid. Stage 2 further refined this design by incorporating an interior camera and LED light set for growth and image capture. This design was not fully sealed/airtight and required image storage on an outside device and outside water delivery. Stage 3 was the final, fully-integrated CubeSat design with an interior lighting, camera, and water delivery system. Electronics for data storage, liquid media release, and lighting components were located in the lid atop the chamber that provided a hermetic seal.

Before beginning hardware material screening, we worked to find an appropriate line of transgenic *A. thaliana* that could produce a clear phenotype change upon exposure to an inducer chemical. Three candidate transgenic lines of *Arabidopsis thaliana* that could produce either a photobleached (white) or yellowed phenotype in response to an inducer chemical were screened for activity in the chamber.[42,90,91] After calculating a maximum growth time of two weeks in carbon dioxide-limited CubeSat system, the pX7-AtPDSi line from Guo *et al.* was selected for its ability to display protein expression within 6 days of germination.[91] These seeds were then grown alongside wild-type Columbia-0 ecotype *A. thaliana* in the presence of a chemical inducer to test the CubeSat hardware through various stages of development.

The pX7-AtPDSi *A. thaliana* line (pX7) stably expresses a synthetic estradiol-inducible XVE promoter. This promoter triggers expression of a Cre/lox recombination system, which activates an RNAi pathway to silence an endogenous phytoene desaturase pathway. The XVE system was shown to be tightly regulated, having no “leaky” expression of the recombinase without estradiol

present, making it a perfect choice for our model system.[91] The Chua and Guo labs generously shared seeds from the pX7-AtPDSi line with us, and we verified the inducible transgenic response on phytoagar plates (Figure 3.2). Stage 0 testing verified that the seedlings display a white, photobleached phenotype within 6 days of setup when grown on phytoagar with 0.5X MS salts and 2 μM estradiol.

Initial Stage 1 testing focused on the very basics of plant growth in the CubeSat setup. Collaborators at NASA Ames had previously used Delrin plastic as a scaffold for plant growth, but only in an open-air setting with water. Due to the narrow time frame of the CubeSat experiment, we tested wild-type seeds in water versus liquid growth media (MS salts) to determine which would support a faster rate of growth (Figure 3.3A). Although the liquid media produced more biomass in a 7-day time frame, it also introduced the risk of contamination. This required the materials that could be easily cleaned and sterilized to prevent unwanted fungal or bacterial growth. Unfortunately, Delrin plastic significantly inhibited plant growth when autoclaved or surface-sterilized. As a result, we decided to switch to an aluminum-based cassette since this material showed robust plant growth after autoclave sterilization (Figure 3.3B).

After running several rounds of experiments to test various material candidates (Table 3.1), the team decided to continue with an aluminum-based plant growth unit. The unit consists of a cylindrical aluminum chamber, a flat growth cassette disk with multiple troughs, and a combination of filter paper and aluminum mesh spacers to create rows for individual seedlings (See Materials and Methods for details). Several biocompatible plastics were tested for the growth cassette and mesh spacers, but these materials exhibited reduced growth efficiency, especially after both autoclave and surface sterilization (Figure 3.3). Although high levels of aluminum have been shown to have a negative effect on plant growth in other systems,[92] we found that this material consistently performed better than plastic alternatives in our setup.

Interestingly, we saw that the pX7-AtPDSi line photobleached and eventually died after 8-10 days of exposure to 2 μM estradiol, likely due to its inability to photosynthesize after this point. Some plants were able to survive and recover if they displayed partial photobleaching, most likely due to consumption or degradation of the estradiol.[91] The commonly observed response in each experiment was a row of healthy, wild-type plants and a row with growth-stunted, dead transgenic plant tissue after 8-14 days (Figure 3.3B)

Once the final materials were selected for the growth scaffolding, we began testing the stage 2 prototype chamber with manually activated blue/red growth LEDs at an optimal photosynthetic photon flux density (PPFD) of 140-150 μmol m^{-2} s^{-1}. Although we attempted to manufacture an airtight environment, the prototype chamber was not yet fitted with an approved hermetic seal for the LED wire passthrough, so there was likely minor gas exchange with the outside environment. Initial experiments were promising, showing clear differences in plant growth between Col-0 wild-type plants and the pX7-AtPDSi line when grown with inducer chemical (Figure 3.4). Wild-type plants that were allowed to grow beyond the 14-day experiment duration

and actually bolted and flowered within the chamber space – an impressive result for such a compact experimental setup.

In addition to normal tests with freshly-sterilized seeds and newly prepared 2 μM estradiol in growth media, we conducted "aging" tests to see how seeds would perform after a 30-day storage period in the dark at room temperature. These tests were meant to simulate the approximate 30-day dormancy period between hardware assembly and activation on the ISS. Fortunately, the wild-type plants still showed robust growth after dormancy with both fresh and aged media + estradiol. Transgenic pX7-AtPDSi photobleached and tests were performed on an increasing number of plants and it was determined that roughly 20-30 seeds was the maximum volume for ideal growth in the chamber (Figure 3.4B). Adding more seeds reduced the plant growth rate and final size. Experiments also showed varied growth between neighboring plants, suggesting that light access could be obstructed due to overcrowding. As a result of these findings, plants would later be distributed across all rows with gaps between individual seeds to prevent crowding and resource competition. Additional tests were also performed to see if the amount of chemical inducer could be increased to prevent "escape" and recovery of partially photobleached pX7-AtPDSi after 10-14 days in the chamber. It was also speculated that increased amounts of estradiol might prevent loss of photobleaching response in the plants due to degradation of the chemical during launch and prior to activation on the ISS. However, using reported conditions of 10 μM estradiol[91] resulted in poor plant growth in our set up, this idea was discarded in favor of the previously validated 2 μM estradiol conditions (Figure 3.4C).

Although plant growth was robust in the stage 2 setup, there were still some issues that required troubleshooting before moving on to a fully automated stage 3 setup. Although wild-type plants were able to continue growing and even bolt after 28 days in the chamber, small regions of contamination would intermittently appear in the experiments despite rigorous seed and hardware sterilization (Figure 3.4D, left panel). After consulting with R. Bowman, a plant scientist at NASA Ames, it was decided that the issue of contamination was most likely due to damaged seeds that were not fully sterilized during the preparation process. Damage in the seed coat can allow fungal spores to enter the interior of the seed and our surface-sterilization protocol cannot penetrate deeply enough to destroy them. To solve this problem, rounds of seed winnowing were performed to sort out damaged and misshapen seeds and ensure that only intact seeds were surface sterilized. This change in protocol resolved the contamination issue moving into Stage 3.

Another hardware-based issue was seen with the color-coded strips placed inside the chamber to determine the internal temperature. These strips acted as a cheap alternative to more complicated temperature-sending hardware, but unfortunately were not waterproof and leaked dye into the chamber (Figure 3.4D, left panel). Once exposed the water, the strips also bled dye into every temperature indicator slot, rendering them useless for this experimental setup (Figure 3.4D, right panel). As a result, plans for the manufacturing of the Stage 3 hardware were updated to include an electronic temperature sensor within the chamber interior.

Final testing in Stage 3 required further material compatibility checks due to the presence of automated, internal electronics. While liquid media was added manually in Stages 0 through 2, Stage 3 was constructed with a spring-loaded water delivery system that could be activated through an electronic valve. The close proximity of the media to the circuit board posed a risk for short-circuiting and early termination of the experiment aboard the ISS. To prevent hardware malfunction, two different types of coatings were applied to the electronics and tested for biocompatibility with the Col-0 and pX7-AtPDSi lines. The first was an epoxy-based resin, which was applied and then allowed to off-gas in a chemical fume hood for 24 hours prior to the start of the experiment. The second was a conformal coating, which was applied and hardened over another 24-hour period.

Although both coatings have been used in previous plant growth experiments,[89] they have never been tested in airtight growth environment nor with the pX7-AtPDSi line. As one might expect, both coatings had significant effects on plant growth (Figure 3.5). The epoxy coating showed significantly reduced plant size but a clear difference in color between the Col-0 and pX7-AtPDSi lines. In contrast, the conformal coating showed greatly improved Col-0 growth, but inhibited pX7-AtPDSi germination. Close examination of the deconstructed experiment (Figure 3.5B) shows that the pX7-AtPDSi seeds have no tissue, photobleached or green, and the aluminum is therefore unsuitable for our CubeSat protocol.

Due to time constraints in the SpaceX launch schedule, no other coatings were tested and the epoxy coating was chosen for the final flight and ground control experiments. Although the plants were stunted in size, there was a clear difference in color between the wild-type and transgenic lines and a visible amount of tissue when viewed with the high-resolution camera. Although there were some hardware issues with poor in-chamber camera resolution (Figure 3.5, panel 1), the hardware was re-calibrated to ensure that we could visualize stunted seedling growth. To mitigate the effects of the epoxy coating, the final flight and ground control hardware were "baked out" at an elevated temperature (38 °C) after application to accelerate off-gassing of any harmful chemicals.

The final ground control and flight experiments (otherwise known as Hydra1) were set up using identical batches of sterilized seeds and growth media. Fresh estradiol stocks were prepared and the hardware was assembled as described (see Materials and Methods) before shipping the flight unit to the Kennedy Space Center (See Figure 3.6 for detailed timeline). The flight unit launched successfully on December 5th, 2018 as part of the SpaceX CRS-16 mission to the ISS. Experimental activation took place on December 14th, 17 days after initial assembly on the ground. A lag period of two days was left between the flight and ground control experiments to allow a controlled response to unexpected changes in the flight environment (e.g. temperature changes, power outages, etc.). Images of both experiments were taken over a two week period and representative images are shown below (Figure 3.7).

The ground control showed acceptable germination rates (for both plant lines (68% for Col-0, 91% for pX7) and clear photobleaching response in the pX7-AtPDSi line. Estradiol levels were

depleted after approximately one week, leading to a recovery in some of the partially-bleached pX7-AtPDSi lines. In contrast, the flight unit showed reduced germination and very pale coloration in both wild-type and transgenic lines. Over the course of the two week period, the flight unit had staggered germination of seedlings and a few late-sprouting Col-0 seeds displayed slightly darker coloration. The differences between the flight and ground control unit are most likely due to the stresses of microgravity and temperature differences between the ground and flight unit. The ground control was run between 23°C - 25°C and the flight unit was situated in an area of the ISS at 27°C . The ideal growth temperature range for *A. thaliana* is between 23°C - 26°C , so it is likely that this temperature difference contributed to the difference in growth.

Although the flight unit produced with overall lighter color and reduced biomass, we were able to quantify the difference in pigmentation between Col-0 and pX7-AtPDSi for both experiments. Using image analysis software, collaborators at ISU outlined the area of each leaf in a specific color channel and calculated the pixel intensity over the 19-20 day growth period (Figure 3.8A). In addition to image analysis, plant biomass was collected and the genomic DNA of the pX7 and Col-0 lines were tested via PCR to determine if protein expression had been triggered on the ISS. Primers from the original pX7 development paper were used and revealed that correctly sized DNA sequence was present, indicating that induction had taken place. These product bands were also sequenced for further confirmation (Figure 3.8B). Taken together, these data show that we have successfully demonstrated inducible gene expression in a transgenic plant line under microgravity conditions.

3.3 Discussion

These experiments present the first known instance of chemically inducible protein expression in a transgenic plant line in a space environment. Now that these inducible systems are known to function in microgravity, we hope that future experiments can refine our method to demonstrate inducible plant production of compounds that are relevant to future spaceflight and colonization efforts. The simple transgenic photobleaching response could be translated into multi-component gene-regulation systems within a single plant line. A simple lettuce leaf could one day contain the genetic information to produce multiple compounds on demand, depending on the inducer molecule that one applies.

Past experiments on the ISS have mainly focused on optimizing plant growth to produce large, healthy crops with minimal genetic variation under microgravity conditions. In contrast, this project was centered around validating a synthetic biological pathway in a transgenic plant line. As such, we did not design the CubeSat system for prolonged, robust plant growth; the CubeSat hardware could not support life beyond a 3-4 week window and seedlings at the two-leaf stage were sufficient to prove our hypothesis. However, we hope that the system can be further optimized to extend the growth window and sustain autonomous plant experiments for extended periods in microgravity.

Although the immediate goal of this project was to validate inducible protein expression on the ISS, the long-term vision was to develop a functional set of hardware for future experiments. After many rounds of plant growth optimization and gradual refinement of the chamber composition, the CubeSat system can now support autonomous plant growth and withstand the challenges of space launch and orbit aboard the ISS (Figure 3.9). One future goal for this hardware is to optimize plant growth under conditions that would simulate the lunar solar cycle, consisting of 14-day periods of light and darkness. The small size and autonomous design of the CubeSat make it the perfect system to attach to a lunar rover, allowing scientists to observe how different plant lines respond to such extreme growth conditions. Hopefully the protocol and hardware that we have developed can also prove useful for other experiments with similarly specialized growth system requirements.

3.4 Materials and Methods

3.4.1 Materials

Chamber components

- CubeSat aluminum chamber
- Aluminum mesh “combs”
- Whatman filter paper (17 CHR, 70 x 90 mm)
- Gauze sponge (Absorbent Cotton Co.)
- 1% guar gum
- Chamber lid with Raspberry Pi and hermetic wire passthrough
- Metal camera stand with LED red/blue light array and valve
- Sterile water bag, prepared for water delivery
- Sterile water indicator spots
- Rubber o-ring
- Aluminum screws (for chamber lid)

Media components

- 0.5X Murashige & Skoog (MS) Salts, pH 5.8
- 2 mM estradiol stock in ethanol
- MilliQ filtered water

Seed sterilization materials

- Metal tweezers
- Arabidopsis thaliana pX7-AtPDSi seeds
- Arabidopsis thaliana Columbia-0 seeds
- Filter-sterile 70% ethanol + 0.1% Tween-20
- Filter-sterile 100% ethanol
- Autoclaved Whatman filter paper 113 (125 mm Wet Strengthened Circles)
- Aluminum foil

Chamber assembly materials

- 30 mL sterile luer-lock syringe, in packaging
- Sterile needle in packaging (BD Precisionglide 18G pink needle preferred)
- Sterile petri dish in packaging
- Autoclaved glass microscope slides
- Autoclaved 1% guar gum (w/v) in a sterile tube
- Micropipettors, P1000 and P200
- Sterile barrier pipette tips
- Several pieces of autoclaved aluminum foil
- Sterile 50 mL conical tube

See Figure 3.10 for representative pictures of chamber and interior.

3.4.2 Hardware Design

The growth chamber was fashioned from aluminum to produce a cylindrical interior. A circular aluminum cassette containing 8 rows for plant growth and a circular opening for water delivery was machined to fit within the chamber without lateral movement. The interior camera and LEDs are supported on a circular PCR placed on an aluminum stand to sit atop the growth cassette. A steel spring is fitted atop the stand and places pressure on an aluminum plate, which contacts the water bag containing the plant growth media. This water delivery system is triggered through the release of an electronic valve, allowing the spring pressure to deliver media through a tubing system that directly contacts a piece of gauze below the growth cassette. This method allows media delivery directly to the rows of filter paper holding the seeds in the absence of gravity.

The chamber is sealed with a box-shaped lid outfitted with a Raspberry Pi containing the operating system for the experimental commands (LED lights on/off, valve on/off, camera on/off). Hermetic wire passthroughs allow communication between the chamber and the Raspberry Pi without exposing the electronics to condensation from the growth media. All exposed electronic components in the chamber were coated with epoxy resin to prevent short-circuiting and premature termination of the experiment.

3.4.3 Experimental Setup

Pre-assembly

Growth Chamber

Filter paper pieces and aluminum combs were pre-measured and cut according to growth cassette dimensions (see Table 3.2). Seed placement was marked between comb sections with a pencil using the metal end of a mechanical pencil to create a pronounced indent that could be viewed from the side. Growth cassette was assembled by filling each row with a “sandwich” of filter paper (seed markings), aluminum mesh, and filter paper (spacer, no seed markings). The seed markings were arranged to face the center of the cassette. A piece of folded gauze was placed at the bottom of aluminum growth chamber. Gauze was centered to avoid contact with the edge of chamber and prevent blockage. Metal tweezers or forceps were used to grasp filled growth cassette and lower into the bottom of the chamber. Cassette was turned so rows were

aligned to suit camera angle and cassette and gauze were pressed down to be flush with bottom of chamber.

Entire chamber was covered with aluminum foil and autoclaved at 121°C or 134°C. Once cycle was complete, exterior was wiped with 70% ethanol and transferred to biohood. All metal tweezers and aluminum screws were autoclaved in a covered container or aluminum foil. Once the cycle was complete, materials were sprayed thoroughly with 70% ethanol and container transferred to biohood. All other materials were sprayed thoroughly and place in biohood to dry under air flow for at least 30 minutes. Electronic components were not autoclaved. Instead, these were sprayed with 70% ethanol and dried for at least 30 minutes. Electronic components must be completely dry prior to assembly to avoid short-circuiting and equipment damage.

Media / Guar Gum / Seeds
A 1% w/v guar gum solution was prepared with sterile water and autoclaved at 121°C or 134°C on liquid cycle. 2 mL sterile liquid aliquots were prepared in the biohood using the upper clear portion of the guar gum solution. A 50-100mL container of 0.5X MS Salts, pH 5.8 was autoclaved and transferred to biohood. Seeds were sterilized according to NASA Ames protocol -- 5 min hand swirling with 70% ethanol + 0.1% Tween-20 followed by 2X rinses with 100% ethanol 10 second hand swirling. Seeds were then gently suspended in 100% ethanol and ejected onto sterile Whatman filter paper circle to air dry completely. Sterile seeds were collected in autoclaved tubes and stored in a dark drawer until use.

Assembly of Growth Chamber
This assembly was performed in the biohood with pre-sterilized materials – biohood was sprayed down with 70% ethanol and allow it to dry at least 15 minutes before starting work. Autoclaved aluminum foil was also sprayed down and used as a work surface in the hood as a form of secondary containment. Sterile growth cassette was unwrapped in hood and the blunt end of sterile tweezers was used to gently press each piece of filter paper down to ensure contact with lower gauze layer. A small amount of seeds was tapped out of the tube (20-25 seeds for full experiment) in a sterile petri dish or glass slide. A minimal amount of guar gum was added to coat the seeds.

Forceps were used to gently lift (do not squeeze or seed may be damaged) a single seed and place it within the indent in filter paper. Seeds were placed in every other indent, with pX7 seeds in the 4 rows closer to the water delivery opening. Rows were filled until desired seed number was reached – usually, add to every other indent for 4 rows with pX7 and Col-0, reaching a rough total of ~40 seeds. The final flight experiment used seeds in every single slot. A water indicator dot was added to the top edge of a piece of filter paper using forceps.

A total of 30 mL of 0.5X MS Salts was poured into a sterile 50 mL Falcon tube. 30 μL of 2 mM estradiol in ethanol (1:1000 dilution) was added for a final concentration of 2 μM estradiol in MS salts. A sterile syringe + needle was filled with the media, 28 mL total. The media was added to the sterile water bag through the clear tubing attachment (the needle should fit within the tubing

and NOT puncture either tubing or bag). Excess air was squeezed from the bag and tubing was attached to one of the valve openings. Tubing can be heated with flame or hot water to provide a better fit. A second piece of free tubing was inserted so that the opening faced the bottom of the cassette and passed through the hole at one at end. The tubing was connected to the other port on the valve.Metal camera stand was added to the chamber and valve was placed below to prevent lid from jamming. Components were added on top in this order: metal spring, metal plate, water bag, lid. Wiring was arranged to ensure that tubing would not be pinched shut.

A thin layer of silicone grease was applied to the o-ring in the biohood, then the lid was lowered onto the chamber. Significant force was required to keep the lid in place – always check that everything is neatly aligned before applying force to prevent broken tubing / wiring connections. Chamber was closed with aluminum screws / washers. Screws were first tightened by hand, then chamber was taken outside of biohood and screws were tightened more fully with an Allen wrench.

Growth chamber was attached to a power source in lab space using wiring from the hermetically sealed passthrough. Connection to the chamber Raspberry Pi was established through a laptop ethernet connection. Connection was tested by running CaptureImage and ActivateGrowthLED commands.Once tests were complete, the ActivateValve command was used to deliver media. After 5 minutes, the DeactivateValve was entered and the experiment was started by entering the ActivateHydra1 command. Monitor experiment through captured images over the next few weeks.

3.5 Figures

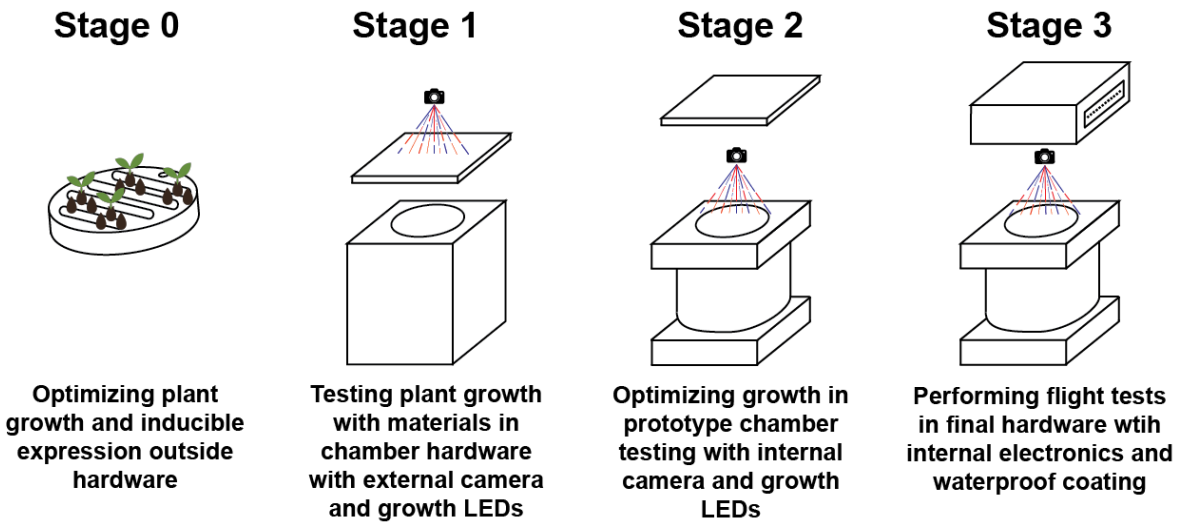


Figure 3.1 Development of the CubeSat plant growth system.
Plant growth was tested during each phase of development to determine the optimal conditions for spaceflight. From left to right: **Stage 0** testing of plant growth in the cassette outside of the chamber, **Stage 1** testing of the chamber with exterior lighting and image capture, **Stage 2** testing of the chamber with partially-integrated interior electronics for lighting and image capture, **Stage 3** testing the fully-integrated and autonomous chamber with interior electronics for lighting and image capture.

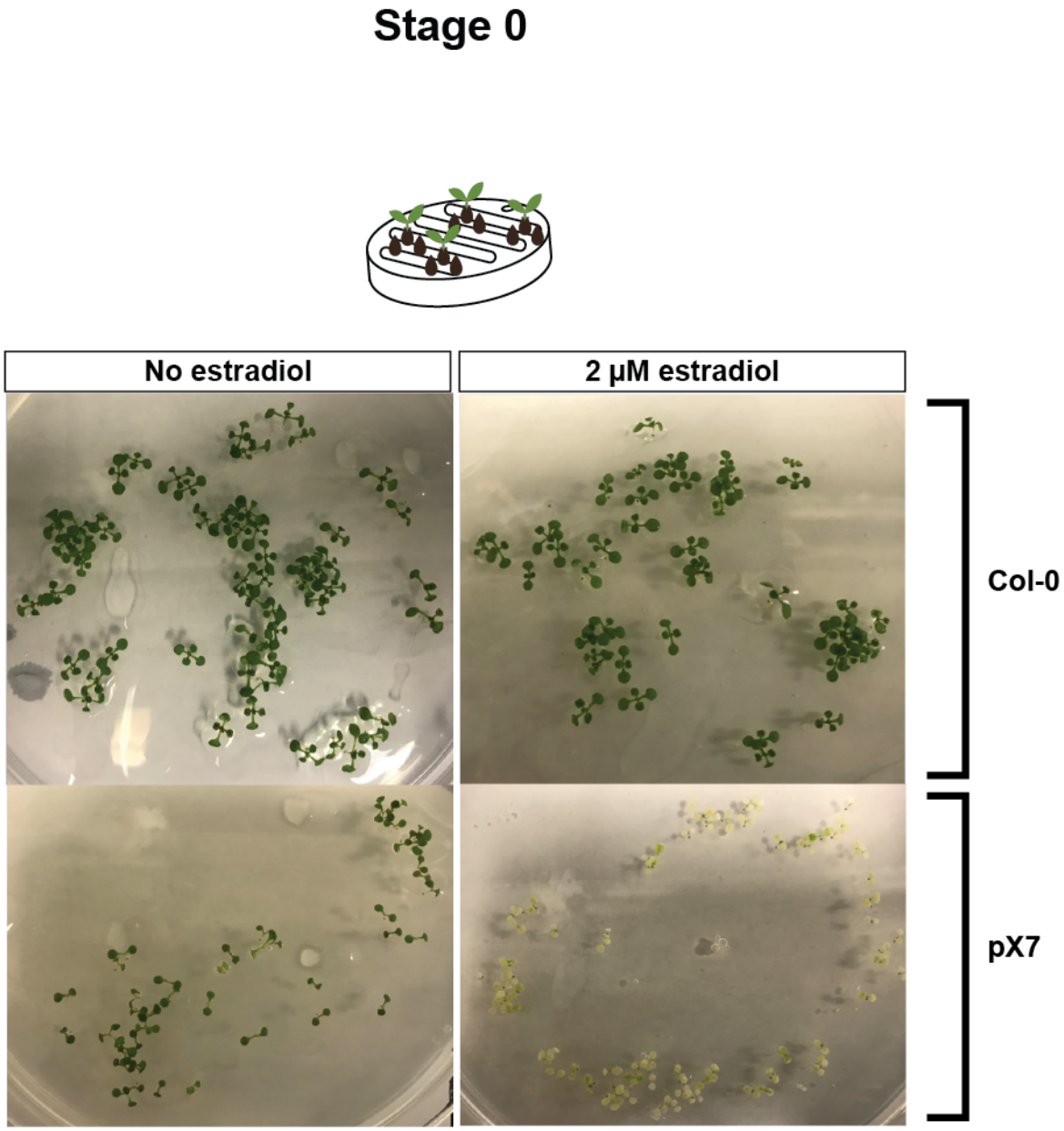


Figure 3.2 Verifying pX7-AtPDSi inducible protein expression.
Growing pX7-AtPDSi and Col-0 lines with and without 2 µM estradiol on phytoagar plates to verify inducible genetic response. Only the pX7-AtPDSi line exhibits photobleaching in the presence of estradiol. **(A)** Col-0, no estradiol, **(B)** Col-0, 2 µM estradiol, **(C)** pX7-AtPDSi, no estradiol, **(D)** pX7-AtPDSi, 2 µM estradiol.

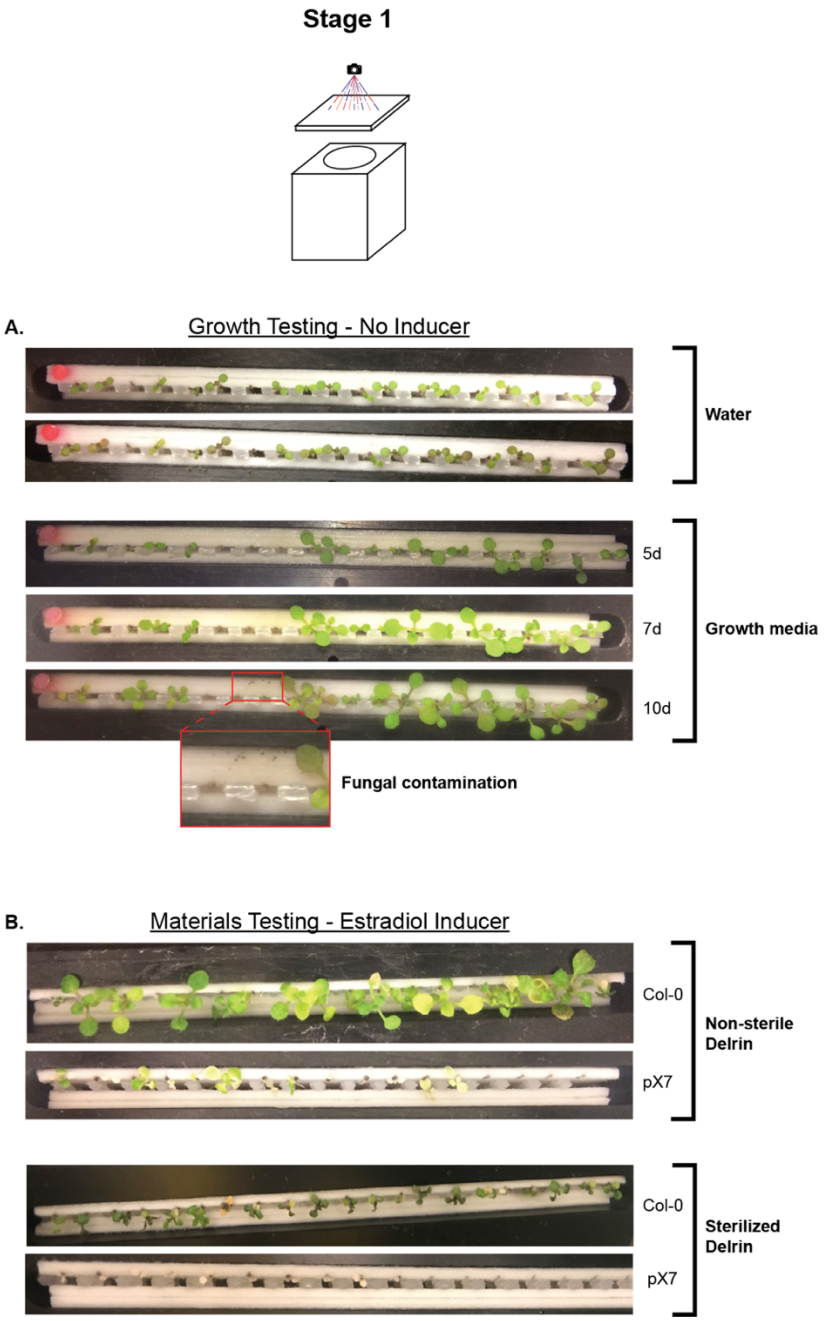


Figure 3.3 Liquid media accelerates growth but requires sterilization.
(A) Comparison of Col-0 growth in Delrin cassette with water and growth media (0.5X MS Salts, pH 5.8). The growth media produces accelerated plant development at 7 days compared to water, but non-sterile materials show fungal contamination at 10 days (red box). **(B)** Sterile Delrin vs. aluminum cassette growing Col-0 in growth media at 8 days. Sterile aluminum shows superior growth and no contamination and was selected for the final growth material.

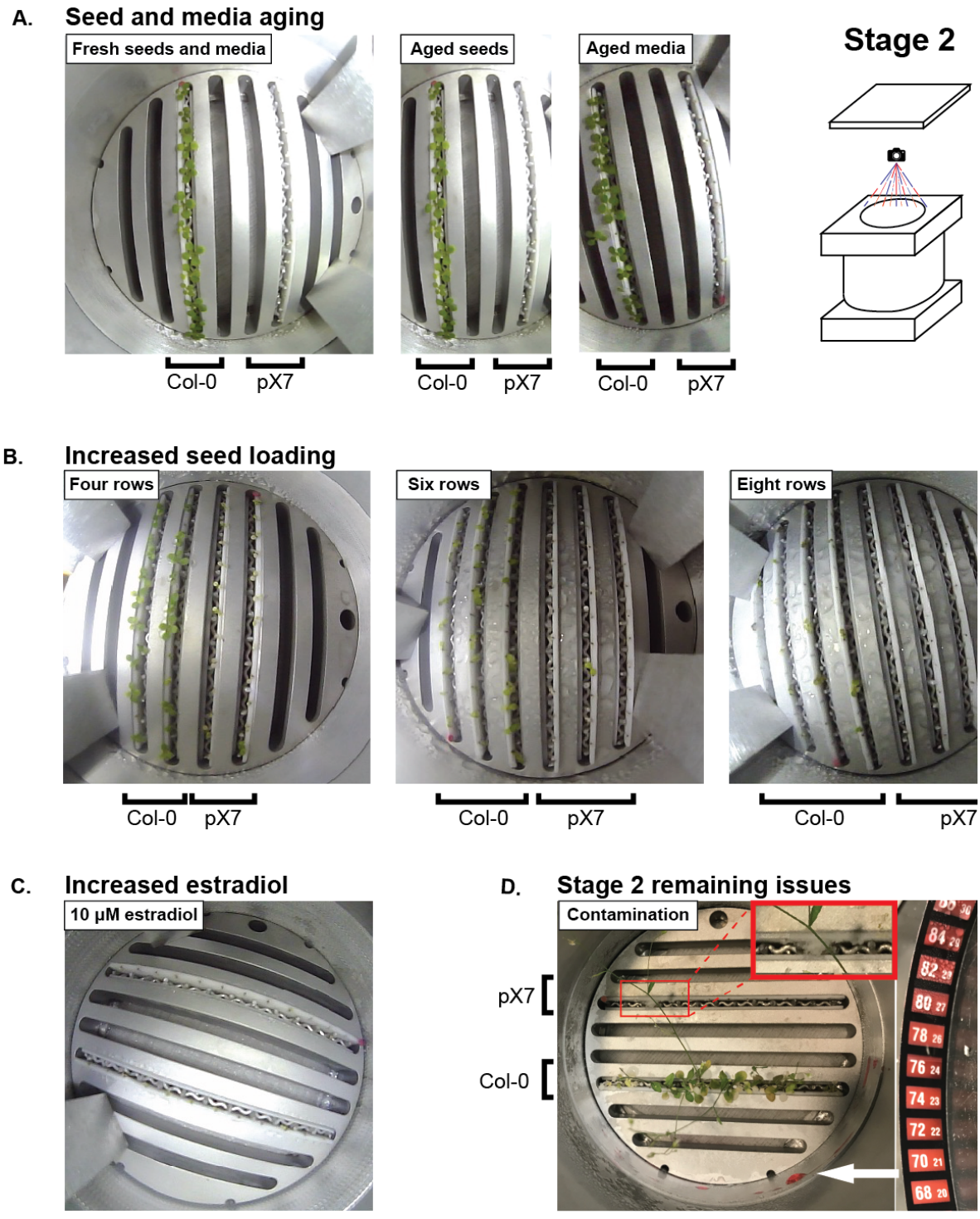


Figure 3.4 Plant growth optimization in Stage 2 design.
(A) Aluminum cassette testing of Col-0 (left) and pX7-AtPDSi (right) at 8 days for fresh seeds and 2 µM estradiol in media, aged seeds and fresh 2 µM estradiol in media, and

aged seeds and aged 2 μM estradiol in media. Aging of media and seeds took place at 25°C in the dark. **(B)** Testing increasing numbers of seeds in the growth chamber, images captured at 8 days for 4 rows, 6 rows, and 8 rows (full chamber). Reduced plant size and health seen as the number of plants increases. **(C)** Chamber image at 8 days with Col-0 (bottom) and pX7-AtPDSi (top) with 10 μM estradiol in media. Little to no growth was observed. **(D)** Left panel – deconstructed chamber with healthy Col-0 (bottom) and photobleached pX7-AtPDSi (top) at 28 days. Fungal contamination (red box) and residual dye from temperature strip (white arrow) were observed. Right panel – soaked and non-functional temperature strip removed from the chamber.

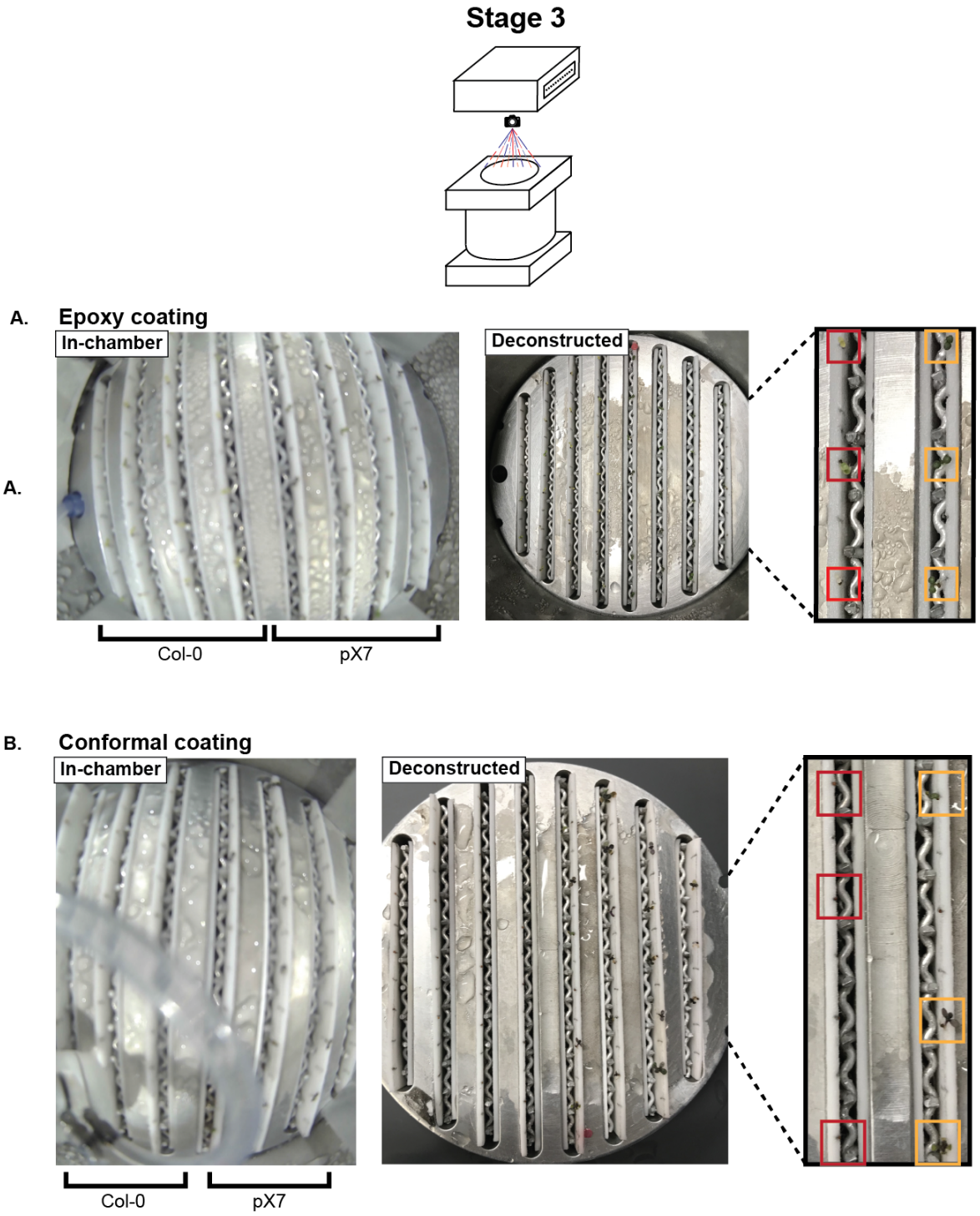


Figure 3.5 Plant growth optimization in Stage 3 design.
Aluminum cassette testing of pX7-AtPDSi (left) and Col-0 (right) with growth media and 2 μM estradiol after 7 days for: **(A)** Chamber with 6 grams of epoxy coating, images from interior camera (1), after deconstruction (2), and zoomed image of pX7-AtPDSi

(red boxes) and Col-0 (orange boxes) seedlings after deconstruction. Epoxy coating produces small seedlings but clear color difference after induction. **(B)** Chamber with conformal coating, images from interior camera (1), after deconstruction (2), and zoomed image of pX7-AtPDSi (red boxes) and Col-0 (orange boxes) seedlings after deconstruction. Conformal coating produces 4-leaf Col-0 seedlings but inhibits germination of pX7-AtPDSi seedlings.

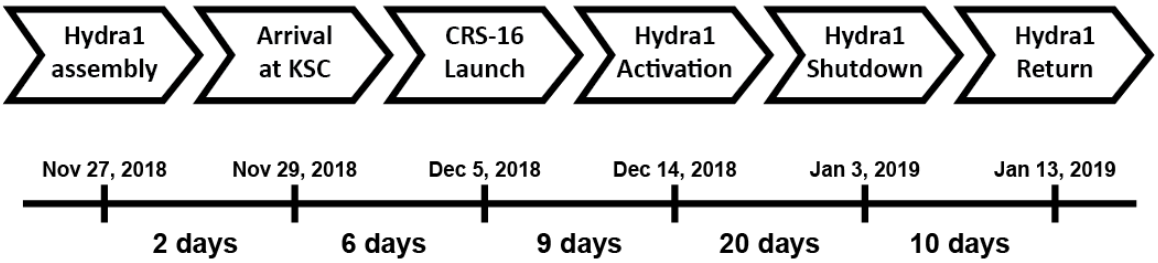


Figure 3.6 Timeline for Hydra1 experiment.
Time lapse between each step from Hydra1 final hardware assembly to eventual return to Earth. Tissue collection and data analysis began shortly after return. The ground control experiment was monitored on earth with a 2-day lag period to allow time to adjust for sudden changes in conditions on the ISS. (KSC: Kennedy Space Center).

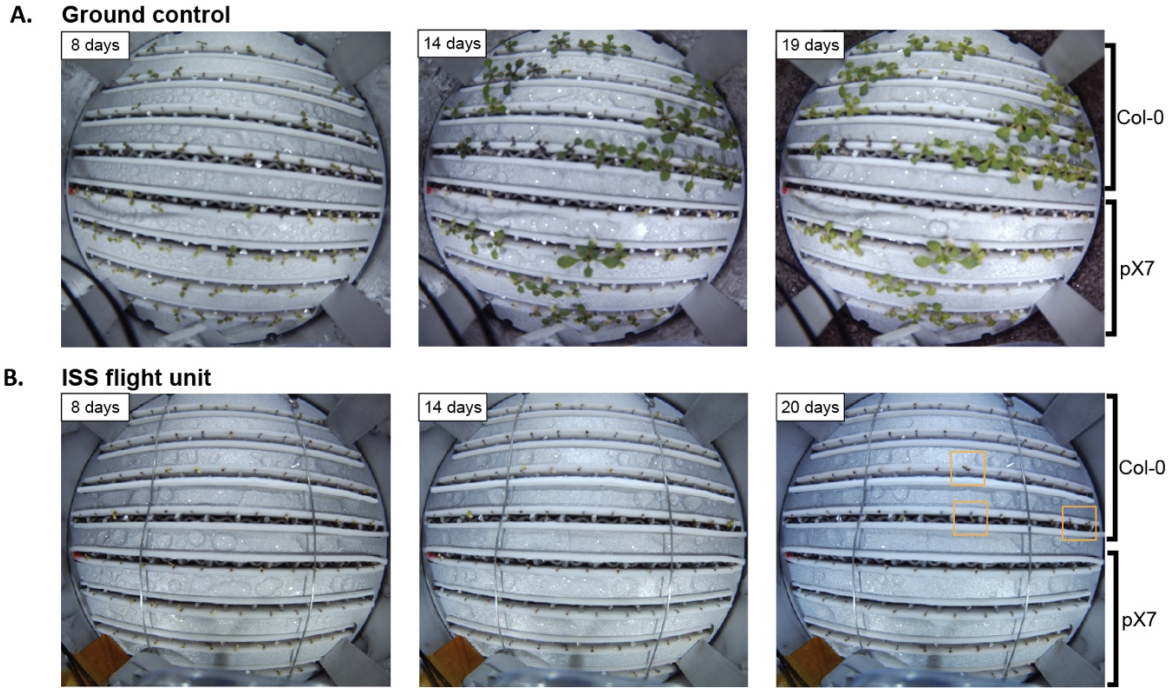


Figure 3.7 Plant growth in the ground control and flight units.
Images and PCR analysis of ground control and flight units **(A)** Ground control at 8, 14, and 19 days (final set was experiment decommission). Clear photobleaching response seen in pX7-AtPDSi samples (bottom four rows) with some seedlings recovering after consuming available estradiol. **(B)** ISS flight unit at 8, 14, and 20 days (final set was experiment decommission). Seedlings are much smaller than expected with poor germination rates for wild-type. Staggered germination (orange boxes) and slightly darker green coloration for Col-0 plants.

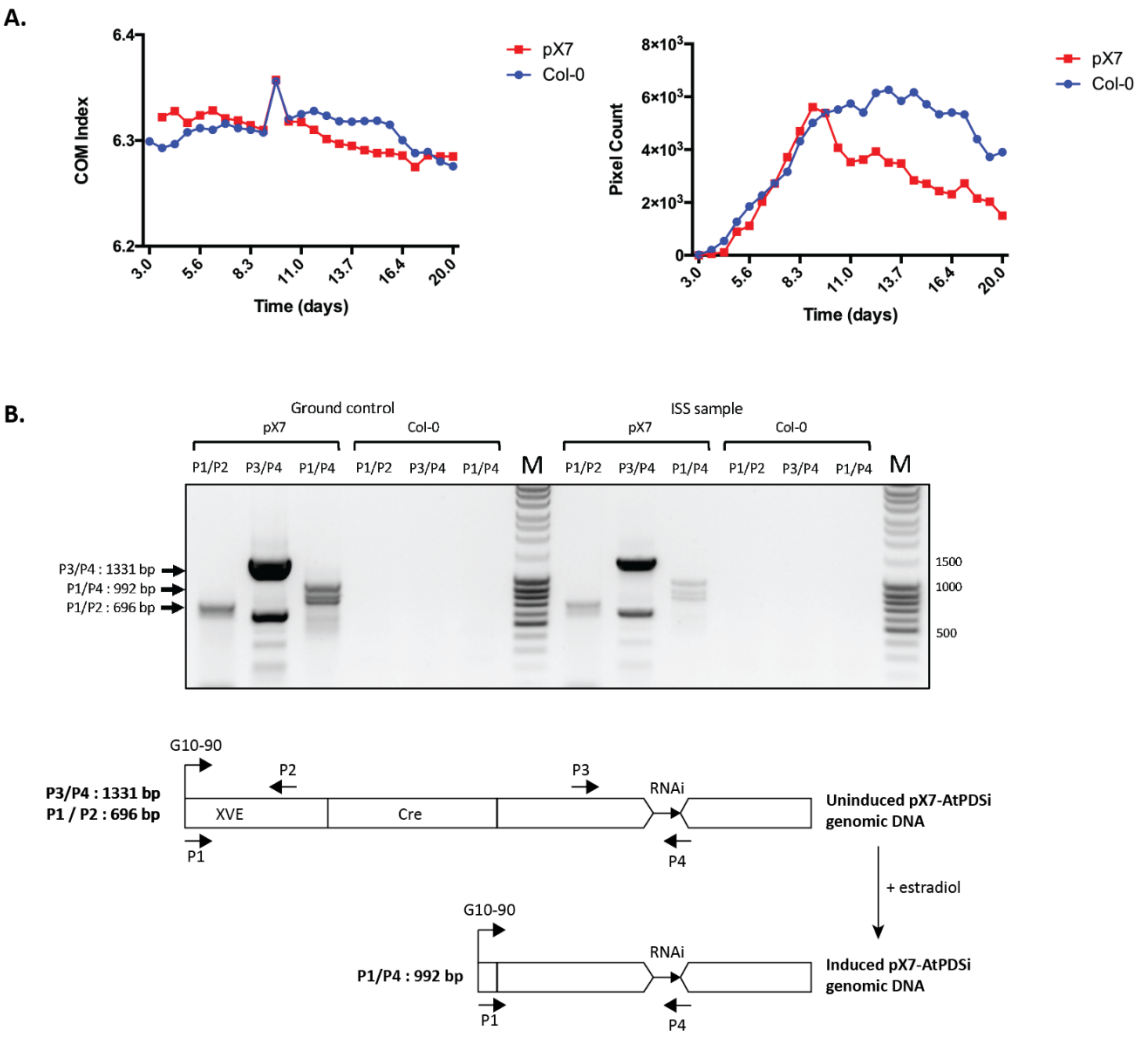


Figure 3.8 Verifying chemical induction through PCR and image analysis.
(A) Quantification of images from ISS, tracking greenness (COM index) and overall size (Pixel count) of plants over course of experiment. This method was used since the plants were much smaller than previous, ground-based runs and the color change in plants was difficult to discern by eye. The wild-type Col-0 plants displayed a slightly darker color after the 8-day timepoint until plants began to die. Overall plant mass was also higher for the Col-0 plants after 8 days, suggesting a loss of color and size for pX7, corresponding with protein induction and subsequent death of these transgenic seedlings. **(B)** PCR analysis of genomic DNA from ground control and ISS flight unit. Primer pairs P1/P2 and P3/P4 amplify unaltered DNA from uninduced pX7, while P1/P4 only amplifies the chemically-triggered, transposed DNA sequence. Both ground control and flight unit showed clear genetic evidence of chemical induction. All bands from P1/P4 PCR were extracted and Sanger sequenced and top band was confirmed to be the correct product. Side product bands were due to *Mesorhzobium* gDNA, seen here because whole-plant extracts were collected rather than leaf extracts. Data was collected by Philippe Giegé at IBMP.

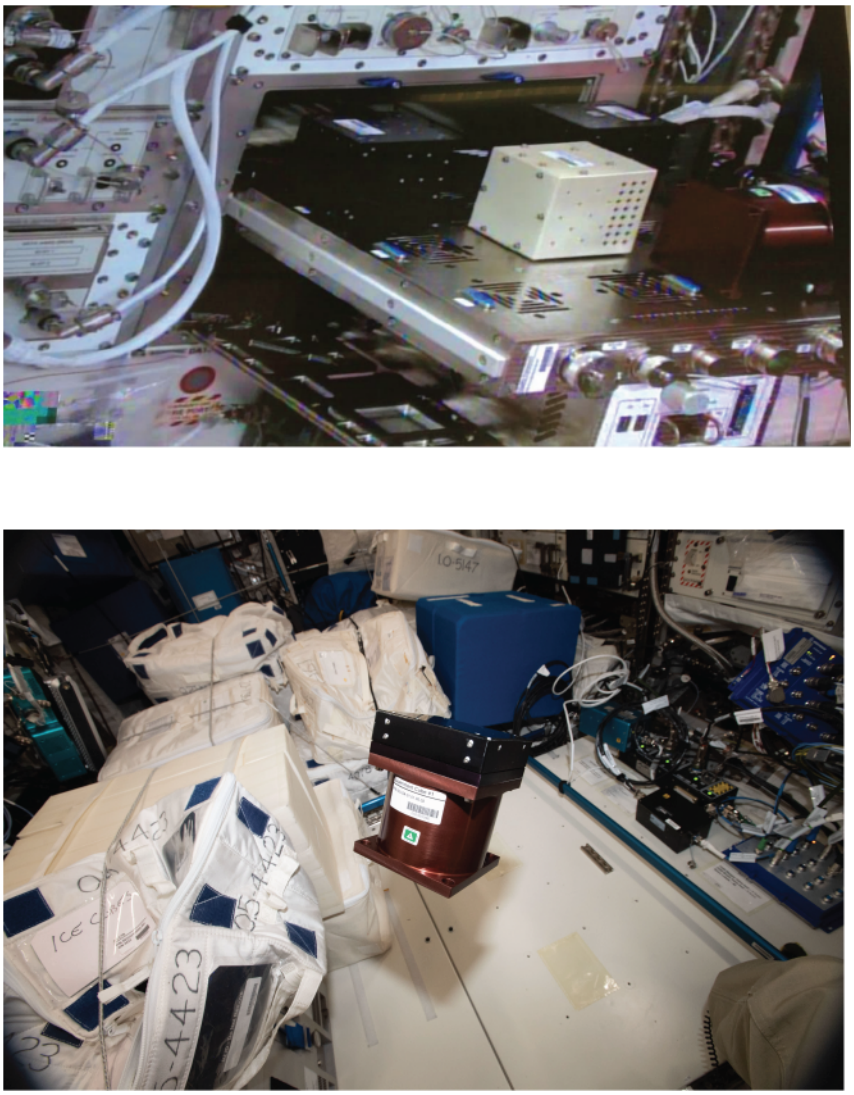


Figure 3.9 Hydra1 CubeSat unit on the ISS.
Images of the Hydra1 unit loaded into its rack on the ISS (top panel, bronze cylinder) and floating free just prior to packaging and shipment back to Earth's surface (bottom, bronze cylinder). Photo taken by ISS crew member, provided by Chris Welch of ISU.

A.

B.

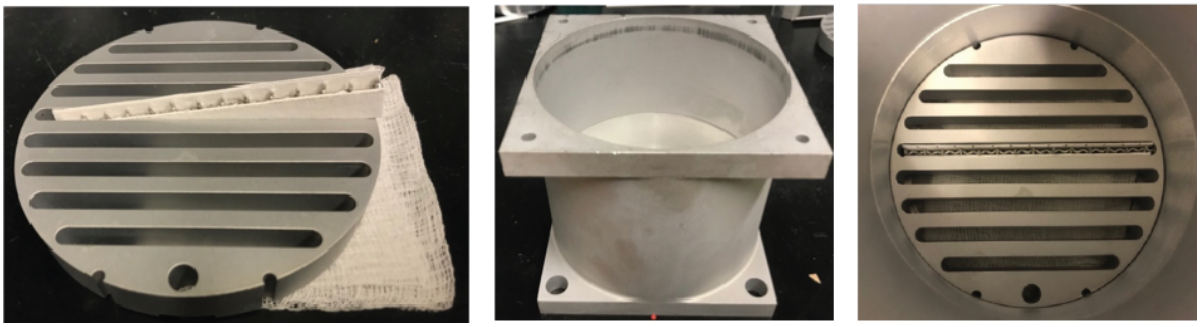


Figure 3.10 Pictures of plant growth hardware.
(A) Filter paper with aluminum mesh and markings/divots for seed placement from top and side view. **(B)** Growth cassette shown with filter paper + mesh partially inserted and on top of gauze layer (left), side view of plant growth chamber (middle), and growth cassette with one row filled inside of the plant growth chamber (right).

3.6 Tables

Table 3.1: Materials tested for growth apparatus

Chamber Component	Testing Stage	Material	Yes/No
Growth media	0	Water	No
		0.5X MS Salts	Yes
Spacers	1	Plastic	No
		Aluminum	Yes
Cassette	1	Delrin	No
		Aluminum	Yes
Temperature Sensor	2/3	Plastic indicator	No
		Electronic sensor	Yes
Electronic coating	3	Conformal coating	No
		Epoxy	Yes

Table 3.2: Filter paper and aluminum mesh dimensions

Row Number	Length (in)	Paper Width (in)	Mesh Width (in)
1	2.1875	0.5 seeds, 0.375 spacer	0.375
2	2.875	0.5 seeds, 0.375 spacer	0.375
3	3.25	0.5 seeds, 0.375 spacer	0.375
4	3.4375	0.5 seeds, 0.375 spacer	0.375

NOTE: These dimensions were measured by hand and small variations will not affect overall plant growth. “Seeds” dimension refers to paper with seeds affixed and “spacer” to paper with no seeds affixed for each row.

4 Development of genetically encodable mRNA and sRNA sensors

Portions of this chapter have been published in:

Hallberg, Z. F., Su, Y., Kitto, R. Z. & Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem* **86,** 515–39 (2017).

Kitto, R. Z. & Hammond, M. C. Development of RNA-based fluorescent biosensors for bacterial sRNAs. *Manuscript in preparation*.

4.1 Introduction

4.1.1 Biosensor engineering

The centrality of RNA in gene regulation is due not only to its role as messenger RNA but also to the regulatory roles of many new classes of noncoding RNAs that have been discovered. Some noncoding RNA classes function in the context of ribonucleoprotein complexes, including microRNAs, which typically target eukaryotic transcripts for translational repression,[93] and clustered regularly interspaced short palindromic repeat (CRISPR) RNAs, which target phage DNA for cleavage as a bacterial defense mechanism[94]. Other noncoding RNA classes function independently, without any protein partner. Riboswitches are a remarkable example class, as they are cis-regulatory structured RNA elements capable of controlling expression of downstream genes by direct response to a small molecule ligand. Riboswitches comprise a ligand-sensing domain, termed an aptamer, and a regulatory domain, termed the expression platform. Aptamer binding to the ligand compound stabilizes the aptamer structure and causes a conformational change or other activation mechanism in the expression platform domain, which mediates gene regulation. In the past 2 decades, more than 20 distinct riboswitch–ligand pairs have been discovered[95] that showcase the capability of these natural aptamers for selective, tight molecular recognition of diverse compounds in cells. In addition, the study of natural expression platforms has revealed several different mechanisms for ligand-dependent gene regulation by riboswitches in bacteria, plants, and fungi. These findings reinforce that nature makes broad use of riboswitches for gene regulation (although mostly in bacteria and conspicuously absent in animals), which has fueled researchers' interest in engineering riboswitch-like systems for their own in vivo applications.

Quantitative, real-time detection and tracking of small molecules in living cells have traditionally relied upon the use of protein-based biosensors.[96] Riboswitch-based biosensors present an appealing alternative for several reasons. First, unlike proteins, these RNAs can be synthesized in vitro from commercial DNA templates using a single enzyme, T7 RNA polymerase, which accelerates high-throughput screening for biosensor development. Second, riboswitch aptamers are naturally evolved for in vivo function with high ligand specificity, affinity, and allosteric regulation in response to ligand. These qualities enable RNA biosensors to produce a high fluorescence turn- on in the presence of their target molecule, with excellent discrimination against other cellular metabolites. Third, Spinach-based biosensors employ a fluorophore dye molecule that enables cellular imaging under anaerobic conditions.[97] In contrast, protein-based sensors generally require extensive engineering to optimize signal output and affinity, and fluorescent proteins in the GFP family require oxygen for chromophore maturation.[98] In this section, we review the applications of riboswitch-based fluorescent biosensors for monitoring metabolites and small molecule signals in bacteria.

Natural riboswitch aptamers exhibit high affinity for and dynamic response to a target ligand, making them privileged scaffolds for biosensor development. A biosensor binds specifically to a biological molecule of interest to generate a signal that can be detected by a noninvasive method such as fluorescence. Biosensors allow for the real-time imaging of biologically relevant molecules in vivo. General features of RNA-based fluorescent biosensors include: (a) a

recognition domain for binding the target ligand (e.g., riboswitch aptamer or hybridization sequence), which is fused to (b) a transducer module that communicates ligand binding to activate (c) a signaling domain that binds the dye to produce a fluorescent output.[99] The modular design of these sensors makes them easily adaptable for sensing a wide range of metabolites. The signaling domain for RNA-based biosensors consists of an *in vitro* selected aptamer sequence with high affinity to a profluorescent dye molecule. The first aptamer used as a signaling domain was the malachite green (MG) aptamer, which exhibits >2,000-fold fluorescence turn-on.

Spinach, Spinach2, Broccoli, Mango, and SRB-2 aptamers have been fused to riboswitch aptamers to generate ligand-responsive biosensors for live-cell imaging of target metabolites or sequences (Figure 4.1). The key challenge to biosensor engineering is the design of the transducer module, which must give allosteric regulation of the signaling domain by the recognition domain. An RNA-based biosensor that exhibited 25-fold fluorescence turn-on with SAM was constructed using the SAM-III riboswitch aptamer fused to 10 candidate transducer modules, which were stem sequences predicted to have weak thermodynamic stabilities.[100] The same strategy was applied to generate an adenosine diphosphate (ADP) biosensor.[100] A FRET-based system using Spinach and Mango took advantage of similar structural changes in the transducer region but for the purpose of triggering FRET through increased proximity between the two dye-binding pockets.[101]

For natural riboswitches, their regulatory mechanisms often involve ligand binding affecting the P1 stem, so we expected that a transducer module involving the natural P1 stem would be effective for allosteric regulation. This P1–P2' strategy (Figure 4.2a) was used to construct an RNA- based fluorescent biosensor capable of visualizing the signaling molecule cdiG in live bacterial cells.[102] With the P1–P2' strategy, it was possible to screen a library of riboswitch aptamers fused to Spinach, in which the natural riboswitch sequences were identified by bioinformatics analysis of bacterial genomes. The screen resulted in a suite of four biosensors for cdiG that exhibited a broad range of affinities, including one with Kd < 5 nM, and improved fluorescent turn-on, with a maximal fluorescence signal that was brighter than the parent Spinach2 aptamer in vivo and useful in flow cytometry assays[97]. The P1–P2' strategy also has been employed to generate biosensors for cdiA [103] and 2',3'-cGAMP.[104]

Although the conventional riboswitch–Spinach fusion design is effective for appending most riboswitch aptamers to a Spinach aptamer, it is problematic for sequences that contain a terminal pseudoknot structure, such as the one found in the 3' portion of the SAH riboswitch aptamer. To solve this problem and generate a biosensor that exhibits>11-fold fluorescence turn-on in response to SAH, we instead fused the riboswitch through the natural P2 stem to a circularly permuted Spinach2 (cpSpinach2) aptamer (Figure 4.2b).[105] The design of cpSpinach2 was necessary to keep the biosensor as one contiguous RNA sequence. Similar to the second-generation cdiG biosensors, highly functional SAH biosensors were identified in one round of phylogenetic screening. There is also preliminary data suggesting that this cpSpinach2 strategy is generalizable to other terminal pseudoknot-containing riboswitches (P. Sander &M.C. Hammond, unpublished data).

An alternate strategy to engineering Spinach-based biosensors involves the incorporation of Spinach as an expression platform–like element, separate from the riboswitch aptamer (Figure 4.2c). Similar to the antiterminator–terminator structure switch in transcriptionally regulated riboswitches, the ligand-free riboswitch aptamer interacts with part of the Spinach aptamer and blocks DFHBI binding, but upon ligand binding, Spinach is restored to its functional state. These Spinach fusions provide fluorescence turn-on for detection of guanine, adenine, TPP, and SAM[106]. Similar construction principles were used to produce a turn-off glycine biosensor, which assumes the functional Spinach conformation in the absence of ligand.[107]

Although many RNA-based fluorescent biosensors are constructed using natural or engineered riboswitches to detect small molecules, alternative designs have been employed to sense different targets. Recent work has shown that engineered RNA sensors can detect RNA sequences of interest both *in vitro* and *in vivo*. These sensors make use of fluorescent aptamers that are misfolded until target regions hybridize to a specific RNA molecule, leading to reformation of the dye-binding pocket and fluorescent turn-on.[108–112] Overall, the mechanism is very similar to natural riboswitch-based sensors, but makes use of an engineered platform with direct base-pairing to a target sequence rather than nucleotide interactions with a small molecule. RNA-based RNA sensors are structurally diverse and often designed through the use of software to model folding patterns and response to target RNA.[110]

These different design strategies offer a generalizable approach for the engineering of fluorescent biosensors to visualize a variety of small molecule ligands in vivo. The natural stem approach has proven effective for generating sensors that exhibit greater sensitivity and fluorescence turn-on in response to natural targets. This strategy has also proven useful when engineering aptamers selective for new target ligands, as evidenced by the development of our cGAMP- and cAG-responsive biosensors.[104,113] The development of new fluorophore–aptamer pairs in vitro bodes well for the expansion of RNA-based biosensors into further wavelengths for orthogonal imaging.[114–116] The future is bright for RNA-based fluorescent biosensors.

To date, riboswitch-based biosensors have been developed for in vivo fluorescent imaging of metabolites SAM, ADP, TPP, and SAH in bacteria.[100,105,106] For example, the TPP biosensor was used to monitor TPP biosynthesis in E. coli after addition of thiamine, detecting different rates of accumulation within the cell population.[106] A more stringent example for in vivo metabolite detection is the SAH biosensor, which must detect SAH selectively because the related SAM cofactor has been measured to be in ∼300-fold excess in E. coli[117]. It has been shown that the SAH biosensor has higher selectivity than commercial monoclonal antibodies and can be used to monitor the chemical inhibition of 5′-methylthioadenosine nucleosidase (MTAN), an endogenous enzyme involved in SAH turnover, in live E. coli cells using flow cytometry.[105] Another potential application of SAH biosensors is to detect methyltransferase activity in vivo, although product inhibition of this enzyme class by SAH is common and will require a more sensitive biosensor to detect low turnover enzyme activity. These initial applications do illustrate the potential for using riboswitch-based biosensors to study real-time dynamics of specific metabolic pathways in bacteria.

RBF (riboswitch-based fluorescent) biosensors have also been developed for in vivo imaging of small molecule signals cdiG, cdiA, and cAG in bacteria.[97,102,103,113] The biosensors have been applied toward studying the activity of the corresponding signaling enzymes within cells. We employed the cdiA biosensor in a flow cytometry assay to validate the activity of a putative archaeal diadenylate cyclase, providing the first experimental evidence for cdiA as a signaling molecule in this domain of life.[103] A high-throughput flow cytometry screen of 29 candidate GGDEF enzymes (named after five conserved residues in the catalytic domain) from *Geobacter sulfurreducens* using our cdiG and cAG biosensors recently revealed a novel subclass of hybrid promiscuous (Hypr) GGDEF enzymes that produce cAG.[118] This discovery breaks the long-standing paradigm that GGDEF enzymes make only cdiG, and the distribution of Hypr GGDEFs shows that cAG signaling is more widespread in bacteria than previously understood.

In addition to these small-molecule detection, RBF sensors have been used to image nucleic acids. Unlike natural riboswitch designs, RNA-detecting RBF biosensors rely on direct base-pairing interactions with the target sequence and require extensive engineering. Fluorescence detection has been demonstrated using designs incorporating the Spinach, Broccoli, and SR-DN RNA aptamers for applications in bacteria[109,112] and mammalian tissue.[110,111] These sensors can respond to native miRNAs and sRNAs as well as longer, non-native RNA sequences. Depending on the nature of their design, the sensors can respond dynamically to changes in target levels or provide cumulative signal increase after catalytic interaction with a "trigger" sequence. Although catalytic designs provide high levels of fluorescence *in vivo*, these methods cannot visualize cellular processes in real-time and do not provide a ratiometric readout.[112] In contrast, dynamic sensor designs generally suffer from a lower signal-to-noise ratio and require alternative design strategies such as multiplexing (multiple sensor copies in tandem) for cellular imaging.[110]

In this section, we report our efforts towards optimizing riboswitch-based RBF sensors and designing a new suite of biosensors for dynamic RNA detection both *in vitro* and *in vivo*. For the traditional riboswitch designs, we show that sensor activity can be improved through alterations to sequence space distal to the ligand-binding pocket. Our work on the RNA-detecting RBF sensors has shown that a previously published design method[110] can be applied towards detection of both mRNA and sRNA sequences. We have also shown that these engineered sequences can be modified for improved signal in bacteria.

4.2 Results

***Methods for improved RBF biosensor signal* in vivo**

The aim of this project was to determine if the activity of various traditional Riboswitch-Spinach biosensors developed by our lab could be improved through changes to non-ligand binding regions. When these RBF biosensors are used in bacteria, they are generally appended to the anticodon stem of a human methionine tRNA sequence (Figure 4.3).[97,103] This tRNA "scaffold" is known to stabilize expression of structured RNA for cellular experiments, but the exact

mechanism of assistance is unknown[119]. Our lab has speculated that there are several possible explanations for this stabilization: 1) Improved RNA folding, either co-transcriptionally or post-transcriptionally due to the stable, highly evolved structure of the tRNA, 2) Binding to tRNA cofactors in the cell that improve stability and/or longevity of the transcript, or 3) tRNA "masking" of the sensors as functional RNA to avoid degradation.

Two lab alumnae designed Spinach2-tRNA chimeric constructs using tRNA scaffolds from a diverse set of organisms and calculated the predicted mean free energy of the sequences. They then tested each of them in *E. coli* to determine if the RNA folding stability had a positive correlation with the fluorescence intensity of the aptamer. Surprisingly, two scaffolds outperformed the conventional *H. sapiens* $tRNA^{Lys}$ sequence, *O. sativa* $tRNA^{Ala}$ and *C. korarchaeum* $tRNA^{Val}$, despite having a higher predicted mean free energy.[120] These data suggest that folding stability alone is not responsible for the improved fluorescent output of the RBF sensors with a tRNA scaffold.

Spinach2 appended to the *O. sativa* scaffold showed greater than 3-fold improvement in fluorescence compared to the *H. sapiens* sequence, so the logical next step was to test this alternative scaffold with a traditional RBF biosensor design (Figure 4.2A). The Pl-A and Pl-B cdiG biosensors[97] were cloned into the scaffold and tested for their response with a cdiG synthase (WspR) versus an inactive enzyme (WspR G249A). The *O. sativa* scaffold did indeed improve performance and signal for the sensors (Figure 4.4A), though it was unclear how the scaffold was causing this effect. With these data in hand, the next step was to test the scaffold with an RBF biosensor that contained a non-traditional architecture; in this case, a circularly permuted design (Figure 4.2b). We used a cdiA sensor that was developed by visiting student Philipp Sander while he was working in the lab, dubbed BsuO P6-4. Surprisingly, this sensor did not show similar improvements in signal with the *O. sativa* scaffold (Figure 4.4B).

These unexpected results led us to delve further into the sequences and compare the two scaffolds in greater detail. It was revealed that the *O. sativa* constructs contained the minimal version of the T7 promoter, while the *H. sapiens* sequences had an extended T7. As a result, a control test was performed with the two versions of the promoter and each of the scaffolds appended to Spinach2. Interestingly, the signal improvement was only seen for the minimal T7 promoter with the *O. sativa* scaffold attached to the aptamer (Figure 4.4C). This experiment was repeated several times and the signal difference was found to be statistically significant. Next, the full and minimal versions of the promoter with the BsuO P6-4 cdiA sensor were tested to see if either could recapitulate the improved signal that was evident for the Pl-A and Pl-B cdiG sensors. Unfortunately, the results indicated that there was little change between the minimal and extended T7 promoter for BsuO P6-4 with the *O. sativa* scaffold (Figure 4.4D).

To determine if the fluorescence increase was due to an increase in sensor transcript levels, the bacterial RNA was isolated and visualized on a 1% agarose gel stained with EtBr (ethidium bromide, stains all RNA) or DFHBI-1T. The gels displayed characteristic *E. coli* rRNA (ribosomal RNA) bands and a smaller band with EtBr staining. The lower band was determined to be sensor

/ aptamer RNA, confirmed through DFHBI-1T incubation and fluorescent signal at the same molecular weight for the Spinach2 aptamer samples. The sensor RNA did not show signal because target ligands were not present in these gel conditions. The ratio between the rRNA standards and the biosensor bands did not change significantly between the two tRNA scaffolds, suggesting that RNA levels are not responsible for the *O. sativa* signal increase (Figure 4.5A). Further tests were conducted to determine if different promoter / scaffold combinations could change overall RNA levels for the Spinach2 aptamer. Once again, total RNA analysis was performed on-gel and it was found that relative sensor levels were not significantly different for the two scaffolds even with different promoters (Figure 4.5C).

Hoping to uncover the mechanism behind these unpredictable signal changes, we began testing the tRNA scaffold with sensors and aptamers *in vitro*. A simple screen was conducted with Spinach2, Pl-A, and Pl-B appended to either scaffold. The overall signal for all designs was lower, with no clear difference between either scaffold (Figure 4.6A). Next, melting curve analysis was performed for the same constructs to determine if improved RNA folding stability could explain the signal change *in vivo*. However, once again the two scaffolds performed very similarly, with the *O. sativa* melting temperature actually measuring lower than that of *H. sapiens* (Figure 4.6B).The sensor signal accumulation during T7 transcription was then measured to determine if any co-transcriptional effects *in vitro* could account for the improved performance of the *O. sativa* sequences. However, the *O. sativa* scaffold actually showed lower overall signal and a reduced transcription rate compared to the *H. sapiens* scaffold (Figure 4.6C).

These data suggest that the effects of the *O. sativa* tRNA scaffold are somehow dependent on both the promoter and adjacent RBF sensor sequences but are not due to increased sensor RNA levels or folding stability. However, we do acknowledge that our *in vitro* testing may not provide a true representation of the complex cellular environment – it is entirely possible that cofactors within the cell are preferentially binding the *O. sativa* scaffold to improve stability, increase longevity, and/or change co-transcriptional folding. Because the exact mechanism of the *O. sativa* scaffold remains unclear, it cannot be added into any biosensor construct to create an improvement in signal output. However, we do believe that the potential for a >2-fold increase in MFI justifies testing to see if a particular RBF sensor sequence will be responsive to this added feature. Hopefully future investigation can provide more insight into the exact reason for these drastic changes in fluorescent signal and allow us to predict when and where the *O. sativa* scaffold may be successfully applied.

Designing RNA-based fluorescent biosensors for alternatively spliced mRNA

Employing riboswitches as the platform for sensor design has proven to be an effective strategy for creating selective, functional RBF biosensors. However, these sensors are limited by the range of targets that can be bound by native (or engineered) riboswitches. Because their target ligands are small molecules, traditional riboswitch-based designs cannot be easily adapted to detect nucleic acid targets. As a result, a number of different strategies have emerged that use artificially engineered sequences adapted for RNA detection. At the beginning of this project, two research groups had recently developed orthogonal designs that could detect miRNA targets in

mammalian cells[110,111]. Our interest in alternative splicing originally led us to propose adapting these sensors for detection of mRNA sequences *in vivo*. Current methods for studying splicing processes rely on quantification through total RNA extraction or tag-based based tracking, but these techniques cannot provide ratiometric image data *in vivo*. We envisioned that these RBF sensors could be engineered to detect alternative exon junctions, allowing one to visualize exon-included versus excluded RNA products in real time within the cell for the first time (Figure 4.7).

Two previously published miRNA sensors with orthogonal dye-aptamer pairs were chosen: the FASTmiR[110] and SR systems.[111] The FASTmiR method used a modeling-based approach for sensor design, while the SR system simply contains a short hybridization region within a standard backbone sequence (Figure 4.8). FASTmiR contains a modified version of the Spinach aptamer, dubbed modSpinach, while the SR system contains the SRB-2 aptamer. Because the two designs bind to different dye molecules with different excitation and emission wavelengths, they were an ideal starting point to create sensors for alternatively spliced transcripts. Sensor candidates were created using both design methods and were targeted to a well-studied alternative exon region in eukaryotes. The target exon was in the SRSF3 (also known as SRP20) gene and this sequences was of particular interest because the alternative exon is well-annotated[121,122] and splicing patterns may change throughout the course of development. We hoped that our two-color imaging system could provide real-time information about these programmed switching events.

First, the FASTmiR architecture was adapted into a FASTmRNA design for the SRSF3 target using the previously described ViennaRNA cofold software.[123] Through trial and error, a number of candidate sensors were generated for 21 and 23 nucleotide target regions across the SRSF3 splice junction (Table 4.2). These candidates were chosen based upon three criteria emphasized in the original design paper: 1) Disruption of the modSpinach aptamer G-quadruplex in the "OFF" configuration, 2) Presence of toehold binding regions complementary to the target RNA, 3) Energetically favorable "ON" configuration upon binding to target sequence (Figure 4.9). In order to attain sequences with these desired features, we chose to use the same 5' and 3' overhangs described in the original paper.[110] Candidates were chosen after manual inspection of ViennaRNA secondary structures for several hundred sequences.

The FASTmRNA sensors were transcribed and tested *in vitro* in plate reader format (see Materials and Methods). All sequences used modSpinach (modified version of the Spinach aptamer) and the FASTmiR171 sensor from Huang et al. was transcribed in parallel as a control. Sensors were refolded and tested under conditions described in the FASTmiR paper (10 mM Tris, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, 5 µM-10 µM sensor and 5 µM-15 µM target RNA) [110] because FASTmiR171 was non-functional under other, previously published RBF biosensor screening conditions. Several 200 nucleotide "mock" mRNA targets were used to screen for sensors hits (Figure 4.10). After screening, several designs were found to be responsive to their target RNA, although the overall brightness was significantly reduced compared to the parent aptamer and FASTmiR171. Two of these sensors (FAST 21U and FAST23U) targeted the upstream junction of the exon-included form of SRSF3, while one targeted the exon-skipped junction of the same gene (FAST

21S). This was exciting to see, as we now had sensors for both the exon-included and skipped forms of this mRNA target.

After validating these three RBF sensors, further *in vitro* characterization was performed to determine if they could respond to varying target lengths and concentration under different conditions (Figure 4.11A-C). The sensors were specific to their target sequences, but only showed modest improvement in signal after changing the sensor-target concentration ratios. Encouragingly, the sensors showed similar signal for 200 nt and 23 nt target RNAs and performed well during a total RNA dope-in experiment (Figure 4.11D).

These *in vitro* results suggested that these FAST designs would perform well in *E. coli* and we decided to test the brightest sensor, FAST 21, in a pETDuet expression system along with its target RNA. Unfortunately, this system only produced a modest change in cellular fluorescence, possibly due to the low copy number of the pETDuet vector (Figure 4.12). We were poised to begin *in vivo* optimization, but were forced to put a halt to this project due to issues with the orthogonal SRB-2 sensor design. Unfortunately, we were unable to recapitulate the SR-21 fluorescence signal *in vitro* for control or SRSF3-responsive designs. These failed experiments meant that we no longer had the capability to image mRNA splicing with our planned two-color system. As a result, we decided to transfer the FAST design method that we had validated to a new class of RNA targets: bacterial sRNAs.

Designing RNA-based fluorescent biosensors for bacterial sRNAs

Bacterial sRNAs are essential for regulation of normal cellular processes, including gene expression. Certain sRNAs in pathogenic strains are also known to control crucial functions during infection of a host organism. Our initial interest in detection of sRNAs was motivated by recent work that has shown transcript levels of the sRNA PinT fluctuate during stages of *S. enterica* infection in the human gut.[124] The ultimate goal of this project was to create a dynamic, selective RNA-based sensor for detection of PinT and other sRNAs to study processes such as human infection in real time (Figure 4.13). Initial development of these so-called FASTsRNA sensors focused on three model targets: PinT from *Salmonella enterica* and RhyB and SgrS from *E. coli*. RhyB is involved in iron metabolism and produces a spike in transcript levels after addition of an iron chelator,[125–127]while SgrS is regulates sugar metabolism and spikes when glucose is added during certain growth phases.[22,112,128] The native *E. coli* sRNAs could be used to validate sensor response to normal changes in native transcript levels, while a sensor for PinT could be optimized in *E. coli* and eventually ported into *S. enterica* for live-cell monitoring during infection.

The process of designing these FASTsRNA sensors was very similar to the FASTmRNA case, beginning with modeling cofolding in ViennaRNA software and choosing sequences that met the established criteria. Sensors were also designed for a 21-nucleotide portion of the target sRNA, chosen for minimal off-target activity within the *E. coli* or *S. enterica* genome. While the FASTmRNA designs met all three design standards using overhang regions described in the original FASTmiR publication, the FASTsRNA sequences required some modification. Our previous experience with the low signal of the FASTmRNA designs led us to incorporate the

cpSpinach2 aptamer into several candidates rather than the previously described modSpinach sequence. In addition, the new sensors needed different 5' and 3' overhangs from those used for FASTmRNA and FASTmiR171 in order to meet the original design criteria. The final designs were ordered, transcribed, and tested *in vitro* for their activity in the presence of target sRNA (Figure 4.14).

For each sensor, two modSpinach- and cpSpinach2-based sequences were tested along with a control sequence with the original 5' and 3' overhangs (Table 4.3). The controls did not met the cofolding criteria and were meant to provide a comparison for our designs with newly created overhangs. Interestingly, although the majority of candidate sensors showed a response to target sRNA, two of the control designs also showed a statistically significant signal change. These results suggest that the ViennaRNA folding predictions may have falsely predicted the major folding patterns for the control sequences. In the end, a total of six sensors (two for each target) were selected for further testing *in vitro*. The hits showed clear, target-specific response to increasing levels of sRNA target and were brighter on average than previous FASTmRNA designs (Figure 4.9 and Figure 4.15). Although the folding characteristics of the hits were compared to the rest of the library, no clear trends emerged based on the ViennaRNA characteristics (Table 4.4). This design tool is useful for generating candidate sequences, but it seems that actual screening is still a necessary final test of the folding program's accuracy.

Although these sensors performed well *in vitro*, there were some initial difficulties involved in transferring them into *E. coli* for *in vivo* flow cytometry testing. Pilot studies with the P4 sensor (PinT, Design 4) showed minimal signal change when expressed in a pET vector system. After confirming by RT-PCR that both sensor and target were indeed expressed, we decided to try using higher copy number plasmids to boost the sensor signal further. Interestingly, a pET-pET sensor-target expression system showed only modest signal changes, while a pRSET-pET system boosted the same signal fold change significantly. (Figure 4.16).

4.3 Discussion

Alternative tRNA scaffolds

Testing tRNA scaffolds from different species *in vivo* highlighted the importance of both the RBF sensor and promoter sequences for improved fluorescent output. The promoter sensitivity was an unexpected result and does suggest that the tRNA scaffold effect could influence co-transcriptional folding or cofactor binding, rather than simply "stabilizing" the RBF sensor. We have shown that the scaffold does not increase relative RNA lifetime either, with RNA levels remaining similar between sequences with the *H. sapiens* or *O. sativa* tRNA. However, there are still many tests that remain before any decisive conclusion can be made about the tRNA mechanism. It is particularly interesting that these effects are only seen *in vivo*, requiring some element of the cellular machinery to function.The addition of the minimal T7 promoter and *O. sativa* scaffold does provide a significant signal boost for certain RBF sensors (>2-fold), and these sequence features are certainly worth testing with new designs as an alternative method to

improve turn-on. Hopefully future work can further elucidate the exact processes that create this effect and allow us to predict when this signal boost will occur.

RBF biosensors for mRNA and sRNA targets

Inspired by the FASTmiR design method, we were able to successfully create RBF biosensors for mRNA and sRNA targets. Although there were some exceptions to the rule, following the three design criteria for sensor folding in ViennaRNA yielded a >50% success rate after screening. The resulting sensors are selective for their target sequence and display clear turn-on both *in vitro* and *in vivo*. These designs are currently limited to *E. coli* expression, but we hope to transfer them to mammalian cells and pathogenic bacteria to monitor alternative splicing and sRNA expression in real time. There is some precedent for the FAST designs in mammalian tissue culture,[110] and we have had success with our traditional RBF sensor designs in other bacteria in previous studies. However, before our two-color splice imaging system can be tested, further work needs to be done to validate an orthogonal mRNA sensor. It was very unfortunate that the SR-21 sensor containing the SRB-2 aptamer could not be recapitulated using reported paper methods, but it is likely that another design could take its place one day. Future work to optimize a FAST design with an alternative aptamer could very well provide this desired color option.

Although the cellular expression of our FAST sensors showed modest improvement through sequence modification, these sensors are limited by the fact that they are artificially designed. Unlike natural riboswitch-based designs, we cannot access the phylogenetic sequence space to easily screen for more responsive variants. Instead, extensive engineering would be required to try to improve the sensor architecture and increase fluorescent output. Unfortunately, the tried-and-true tRNA scaffold method for bacterial expression cannot be used for this RBF sensor due to the fact that the 5' and 3' ends are the sensory domains in this design. It is therefore likely that a split tRNA scaffold would prevent the necessary exposure of the toehold region and no signal change would be seen in the presence of target RNA. We are currently testing several alternative strategies to boost signal *in vivo*, including multiplexing and terminal tRNA scaffolds (Figure 4.17). Initial results for both strategies with the FASTsRNA sensors are promising, and we hope that they may one day be used for expression in different bacterial systems such as *S. enterica* to visualize sRNA levels during infection.

4.4 Materials and Methods

General reagents and oligonucleotides

Oligonucleotides for tRNA scaffold cloning were ordered from ElimBio (Elim Biopharmaceuticals) The FASTmRNA and FASTsRNA library oligonucleotides were ordered in plate format from IDT (Integrated DNA Technologies). Complementary strands were annealed and used as a starting point for cloning in sensory domains onto the cpSpinach2 and modSpinach backbones. DFHBI and DFHBI-1T were synthesized as previously described or ordered from Tocris Biosciences. Non-library oligonucleotides for FASTsRNA cloning were ordered form the University of Utah Core Laboratories.

ViennaRNA modeling

RNA modeling was performed using ViennaRNA fold and cofold software packages. 21 nucleotide fragments of the sRNA target were generated using a Python script and checked on NCBI BLAST to ensure that they would not have off-target regions with high complementarity in the *E. coli* (or *S. enterica* for PinT) genome. Sequences were then ported into Aquamacs and manually checked for desired folding criteria with and without target. For FASTmRNA, overhangs from the original FASTmiR design[110] were kept in place and criteria were met. FASTsRNA sequences did not display proper folding with original overhangs, so manually inserted overhangs were used to obtain the desired folding patterns. Overhang-induced folding changes were difficult to predict and required testing of 20-25 sensor sequences with >10 overhangs for each target. Final hits were taken forward for cloning design.

Molecular cloning

For tRNA scaffold testing, biosensors and aptamers were flanked by the split scaffold sequence and cloned into a pET31b(+) plasmid as previously described.[102,113,129] These sequences were preceded by either a full or minimal T7 promoter (Table 4.1). CDN (cyclic dinucleotide) synthases were expressed on the pCOLADuet vector for the cdiG and cdiA biosensors. The two plasmids were co-expressed in *E. coli* BL21 (DE3) Star cells (MacroLab) for flow cytometry analysis. For *in vitro* testing, sequences were PCR-amplified to include T7 promoter and constructed through run-off transcription.

FASTmRNA and FASTsRNA sequences were constructed using primer overhang addition of sensory regions onto a modSpinach or cpSpinach2 backbone sequences. Complete primer list for each sensor and for construction of backbone sequences are included below (Table 4.5 and Table 4.6). FASTmRNA sensor 21U was cloned into pETDuet along with the target SRSF3 exon included upstream sequence using Gibson assembly. FASTsRNA sensors were cloned into both pET31b(+) and pRSET vectors through Gibson assembly and co-expressed with pET24a containing the target sRNA sequence. pRSET was eventually chosen for the final expression system since it displayed more stable expression with the pET24a vector and had a higher copy number.

In vitro fluorescence assays

All sensors were prepared as previously described.[130] Briefly, templates were amplified from plasmid sequences to contain T7 promoter + sensor. Transcription was then performed with T7 RNA polymerase and the products were purified using either a denaturing 6% Urea-PAGE gel or the Zymo RNA Clean and Concentrator kit, depending on product purity after transcription. RNA was then purified and quantified through thermal hydrolysis[131].

Fluorescence assays for the tRNA scaffold experiments were performed as previously described[130]. Each well contained RNA, ligand, and DFHBI-1T in a binding buffer (40 mM HEPES, 125 mM KCl, and 3 mM $MgCl_2$, pH 7.5). DFHBI-1T was used to mimic *in vivo* conditions since these experiments were for direct comparison. RNA was refolded in buffer by heating to 70°C and slowly cooling to room temperature before setting up the assay. The reaction plate was then pre-incubated at the read temperature and measured using the i3X plate reader (Molecular Devices)

at 448 nm excitation and 506 nm emission. Reads were taken every 5 minutes over a 1-hour period and the final data analysis was performed on measurements after sensor equilibration The FASTmRNA and FASTsRNA sensors were placed under similar conditions, but the refolding protocol and buffer were slightly different (10 mM Tris-HCl, 100 mM NaCl, 1 mM $MgCl_2$, pH 7.5 with 5-10 µM sensor and 5-10 µM target RNA, refolded slowly after 95°C denaturation for 10 min). These sensors were tested with DFHBI, rather than 1T.

Flow cytometry analysis of in vivo fluorescence

Flow cytometry experiments were conducted as previously described.[113] Cultures containing co-expressed sensor + target plasmids were grown overnight in ZYP-5052 autoinduction media for 16 hours.[132] Cells were then diluted 1:50 into 1X PBS, pH 7.4 with 100 – 500 µM DFHBI-1T and added to a plate. Samples were then measured on an Attune NxT (Thermo Fisher) capturing at least 20,000 events with a 488 nm laser and 530 nm filter. Mean fluorescence intensity for each population was quantified and graphed using FlowJo software.

4.5 Figures

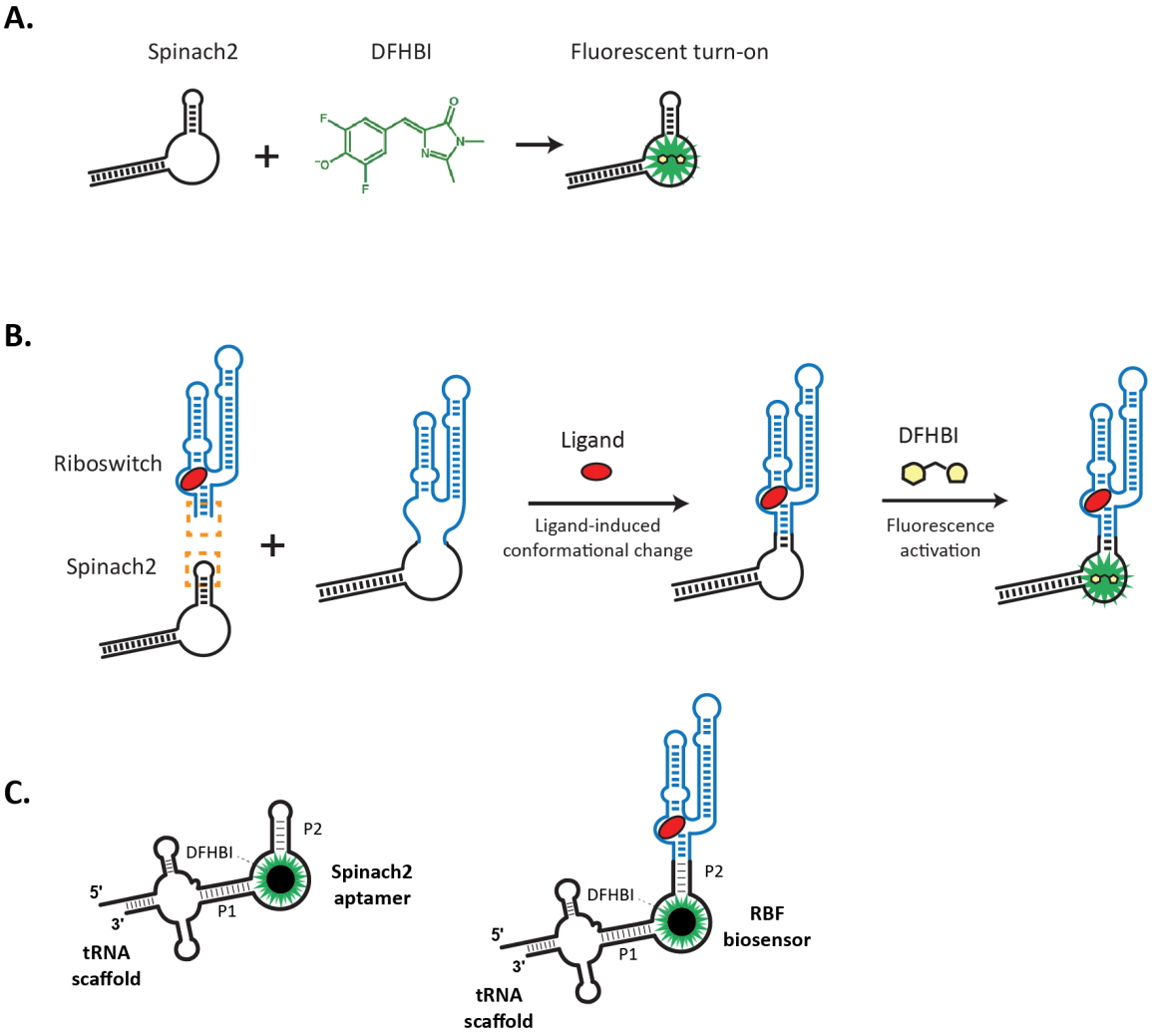


Figure 4.1 Spinach2 aptamer and RBF-biosensor general schematic.
Schematics illustrating function of Spinach2 aptamer and RBF biosensors. **(A)** Interaction between the Spinach2 aptamer and profluorescent dye molecule DFHBI to produce fluorescent turn-on. **(B)** General example of an RBF that is responsive to a target ligand. **(C)** Spinach2 and RBF biosensor appended to a tRNA scaffold to produce greater signal output *in vivo*.

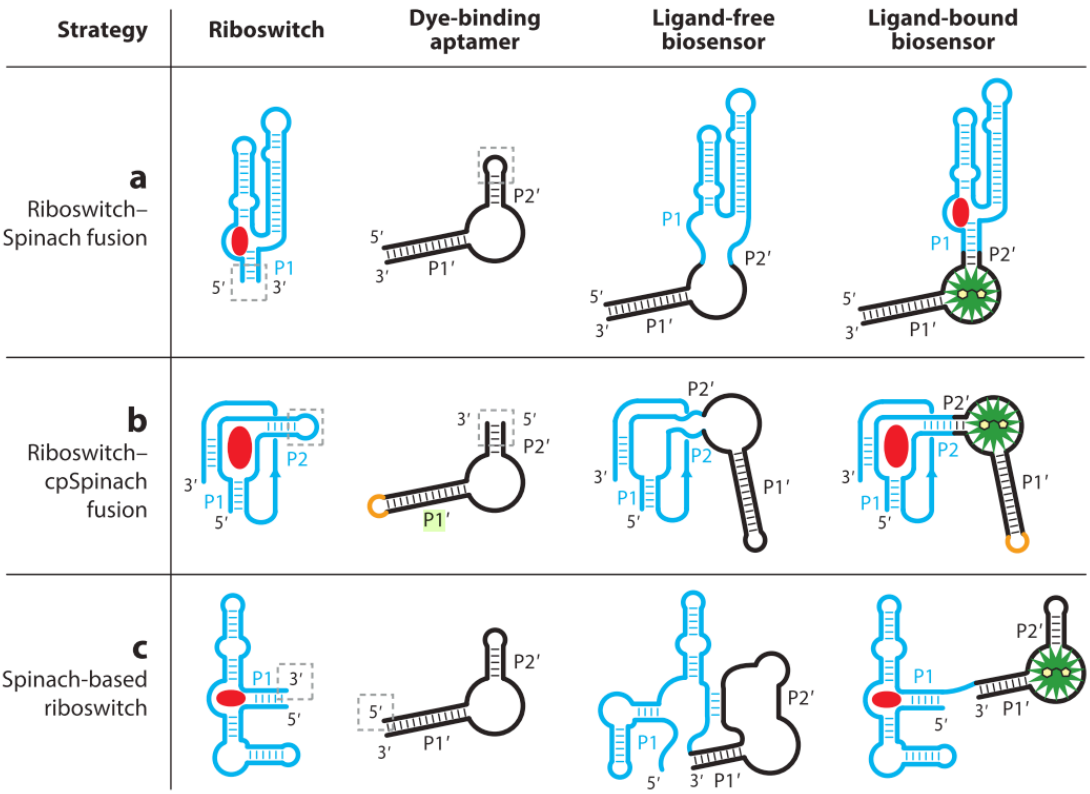


Figure 4.2 Riboswitch-based fluorescent biosensor structures.
Figure adapted with permission from Hallberg et al[18]. Table shows three main classes of RBFs and their structural features.

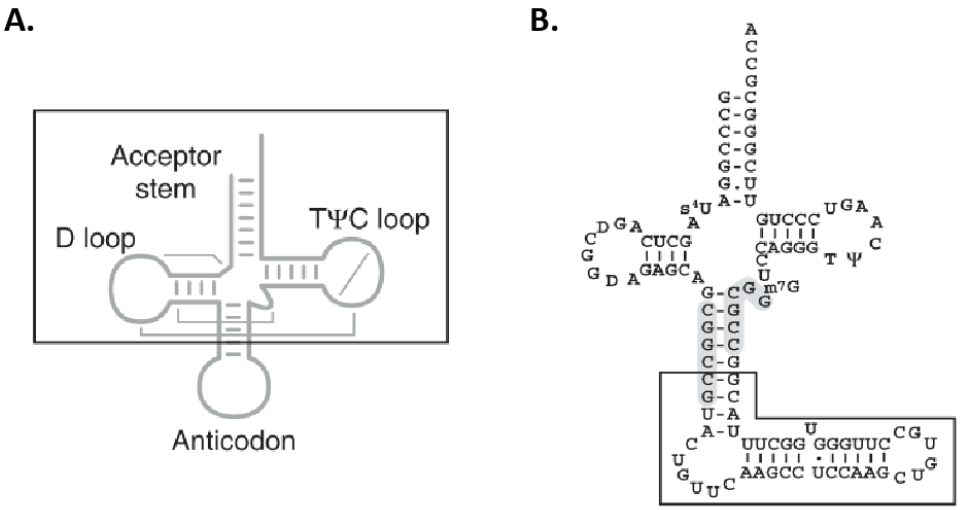


Figure 4.3 tRNA scaffold schematic.
Figure adapted with permission from Ponchon et al[119]. **(A)** Cartoon of general tRNA structure and key features. **(B)** Secondary sequence of tRNA scaffold with an RNA aptamer inserted into the anticodon loop region (black box).

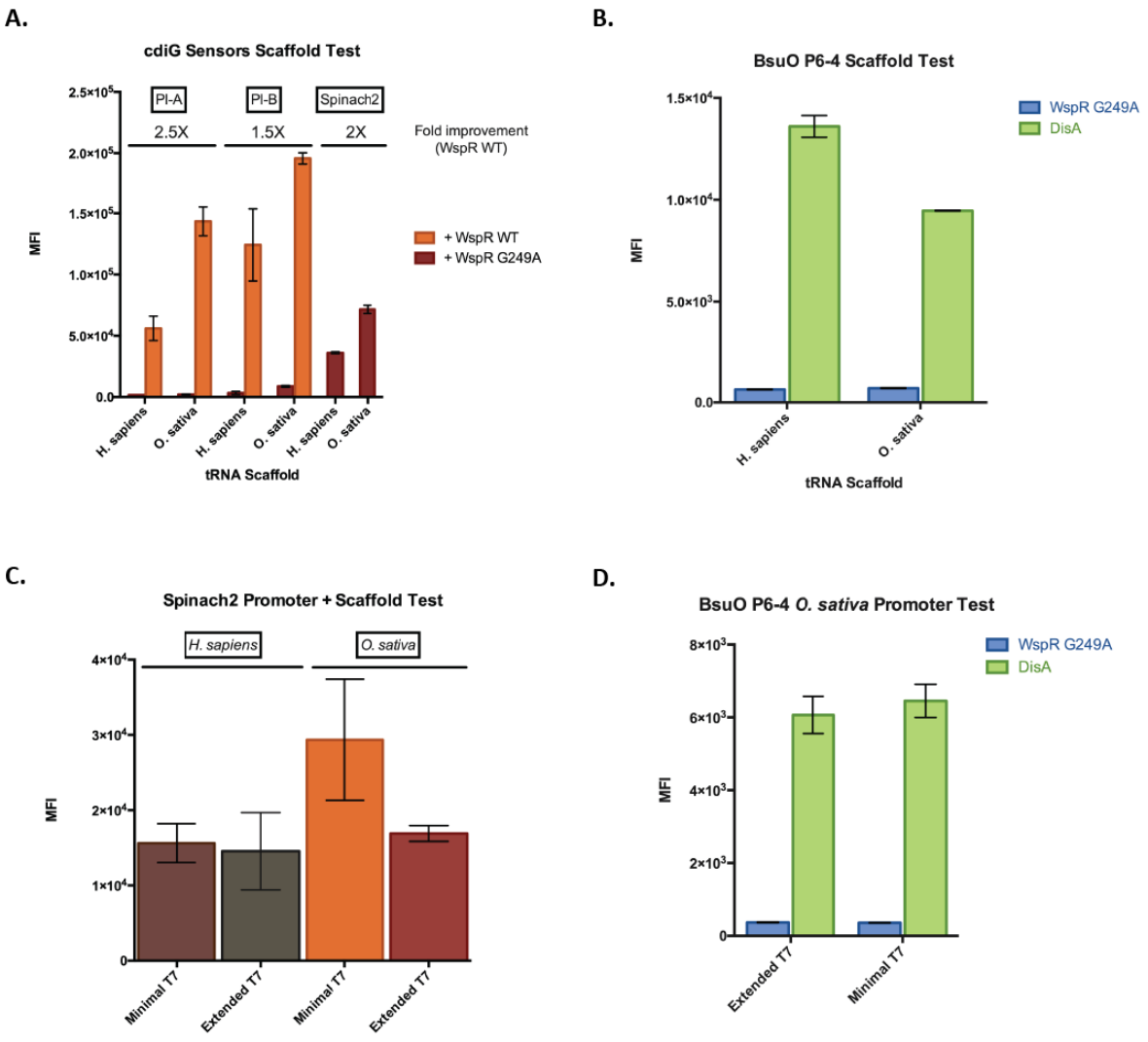


Figure 4.4 Testing tRNA scaffolds with RBF sensors and aptamers.
Flow cytometry data for different promoter-sensor-scaffold combinations. **(A)** Traditional Riboswitch-Spinach sensors for cdiG with two different tRNA scaffolds. Unknown at the time, but all had the T7 minimal promoter. The WspR WT enzyme produces cdiG and the WspR G249A enzyme is a negative control. **(B)** Circularly permuted / Riboswitch-cpSpinach sensor BsuO P6-4 for cdiA with two different tRNA scaffolds. DisA is a cdiA synthase and WsPR G249A is a negative control. Unknown at the time, but both had the T7 extended promoter. **(C)** Comparison of Spinach2 aptamer with two different tRNA scaffolds and two different T7 promoters. **(D)** The BsuO P6-4 sensor with the *O. sativa* tRNA scaffold and two different T7 promoters.

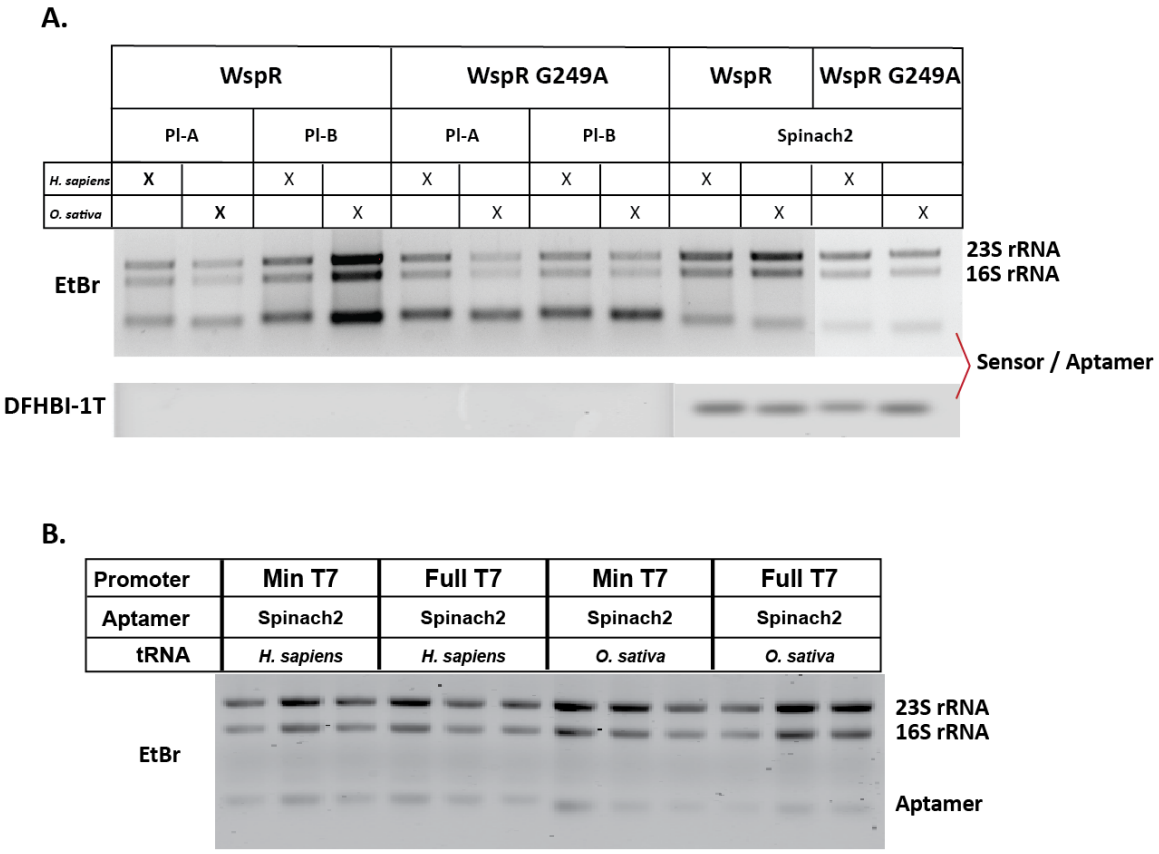


Figure 4.5 tRNA scaffold effects on RNA levels.
(A) Total RNA from *E. coli* used in flow analysis, run on a 1% agarose gel. Staining with Ethidium bromide and DFHBI-1T showed that the bottom band corresponds to expressed sensor/aptamer RNA. Relative RNA levels are similar (with respect to rRNA controls) for each scaffold, despite 2-fold differences in flow cytometry MFI. **(B)** Relative RNA levels for Spinach2 with different T7 promoters. Again, there is no clear 2-fold difference in RNA levels for the different cases.

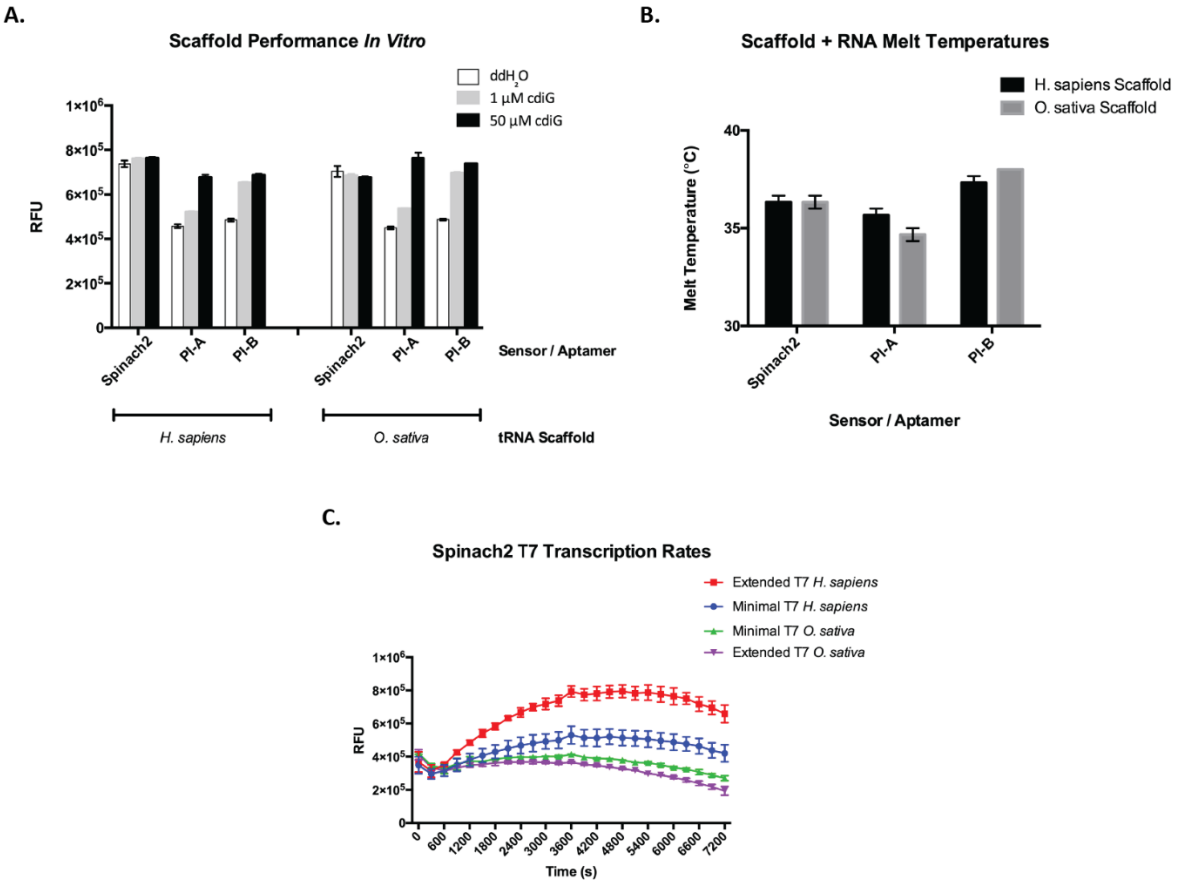


Figure 4.6 tRNA scaffold performance *in vitro*.
(A) Spinach2, Pl-A, and Pl-B sequences with corresponding tRNA scaffold tested *in vitro*. The difference in signal seen during flow cytometry experiments is not evident here. **(B)** Relative Spinach2 fluorescence when transcribed with the two promoters and tRNA scaffolds. Relative *in vitro* signal does not correlate with previously described *in vivo* results. **(C)** Relative melting temperature for sensor / aptamer with tRNA scaffold. The two scaffolds have similar melting temperatures, with *O. sativa* slightly lower, implying that relative RNA stability does not account for improved signal *in vivo*.

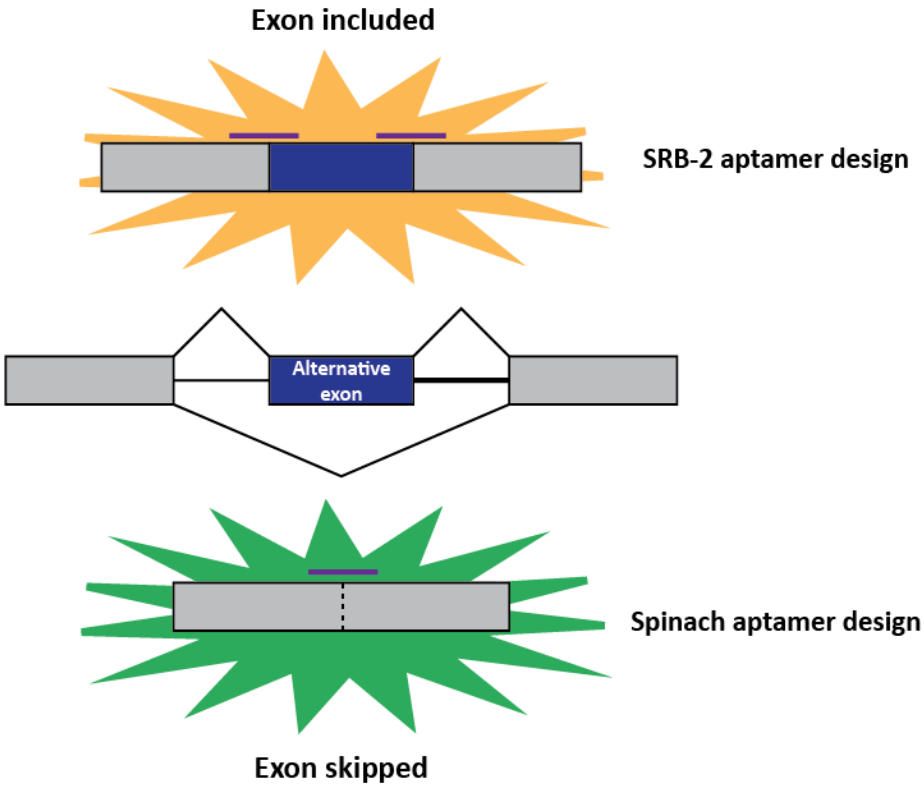


Figure 4.7 Imaging mRNA splicing patterns with RBF sensors.
Representative cartoon showing alternative exon (blue) between constitutive exon (grey) and intron regions (black lines). The exon-included case will have unique upstream and downstream junction sequences (purple) compared an exon-skipped sequence. Different RBF sensors can be targeted to these sequence junctions, producing a yellow-orange or green fluorescent signal depending on the parent RNA aptamer.

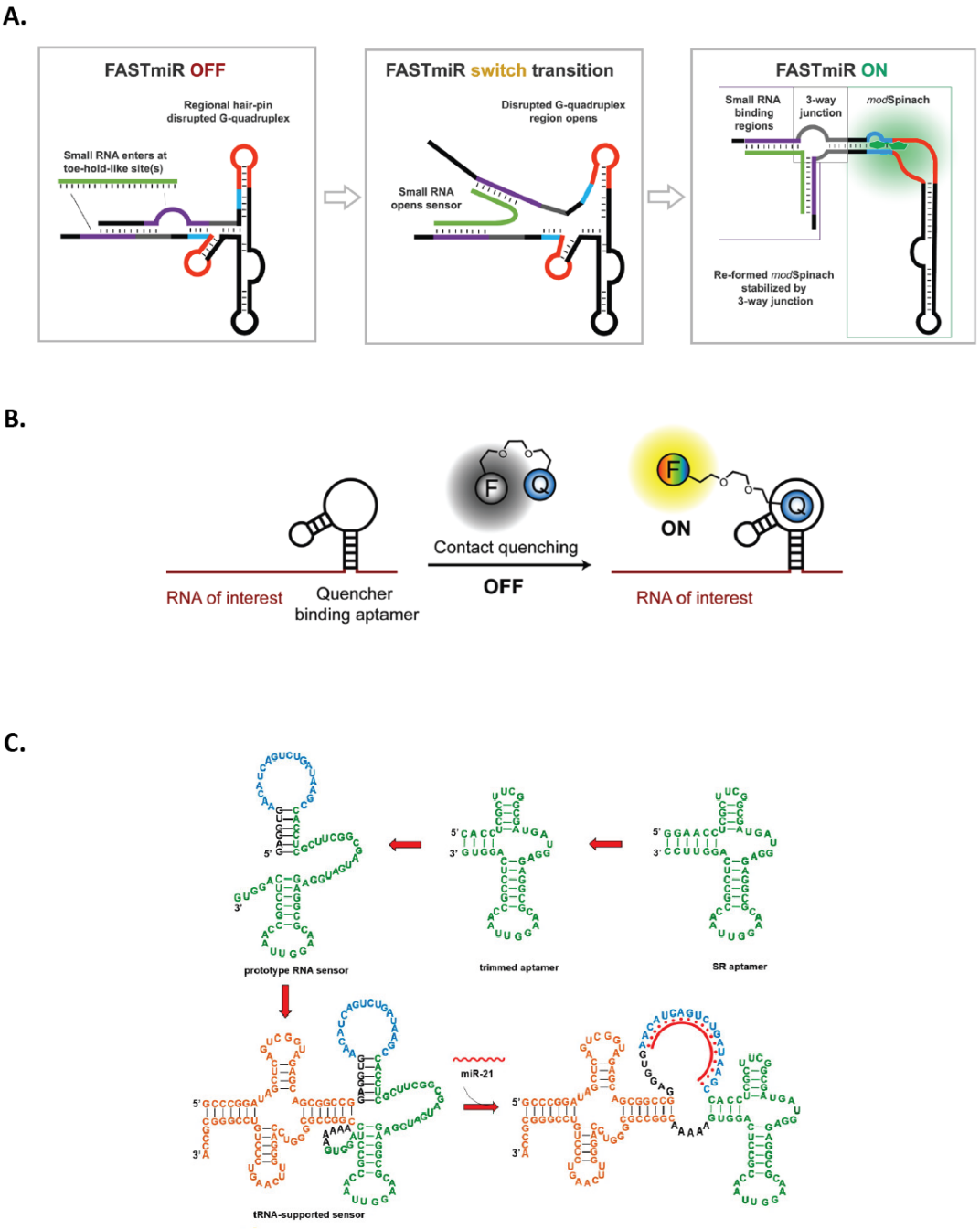


Figure 4.8 FASTmiR and SR-21 miRNA sensors.
(A) FASTmiR sensor system in the OFF configuration contains a toehold-like site where target miRNA binds, triggering a transition into the ON state with a refolded modSpinach (modified Spinach) aptamer. DFHBI binds and the sensor produces fluorescent signal[110]. **(B)** Illustration of SRB-2 aptamer, which binds the dinitroaniline quencher molecule, freeing the paired fluorescent dye after binding[133]. **(C)** SR-21

sensor with tRNA scaffold responds to miRNA target and reforms the SR (SRB-2) aptamer domain, binding to the quencher-dye pair and allowing fluorescent turn-on.[111]

Part (A) is adapted with permission from Huang et al. Part (B) is adapted with permission from Arora et al. Part (C) is reprinted with permission from Ying, Z. M., Wu, Z., Tu, B., Tan, W. & Jiang, J. H. Genetically Encoded Fluorescent RNA Sensor for Ratiometric Imaging of MicroRNA in Living Tumor Cells. *J. Am. Chem. Soc.* **139,** 9779–9782 (2017). Copyright 2017 American Chemical Society.

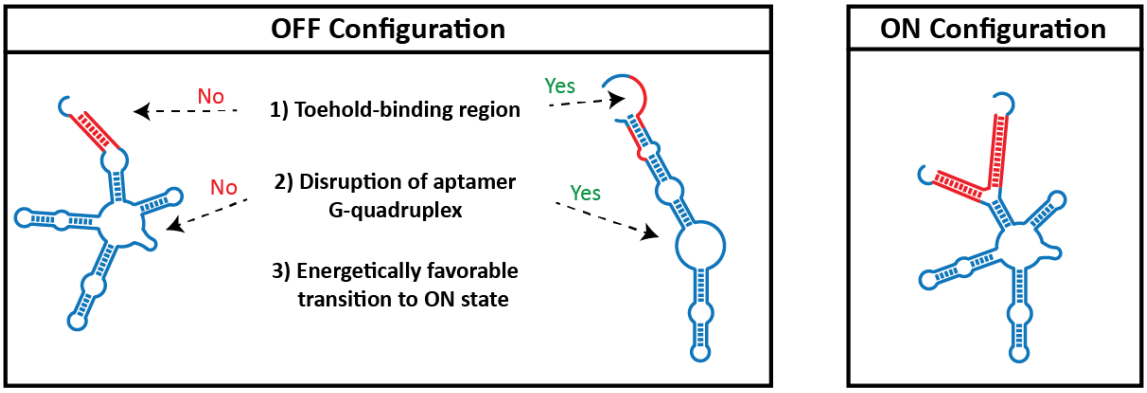


Figure 4.9 Design criteria for FAST sensors

Left panel: necessary design criteria for OFF configuration of sensors, with representative secondary structure for a design that does not meet criteria (left) and one that does meet criteria (right). The toehold binding region is the exposed portion of the sequence that hybridizes to the target (red) and should be accessible for RNA binding. Overhang regions are shown in blue at the terminal ends of these sequences. Right panel: the aptamer G-quadruplex is shown in the ON configuration in the main blue sequence and the red sequences are the target RNA bound to the hybridization region of the sensor. Resemblance to this structure in the OFF configuration was considered a failure to meet criterion 2.

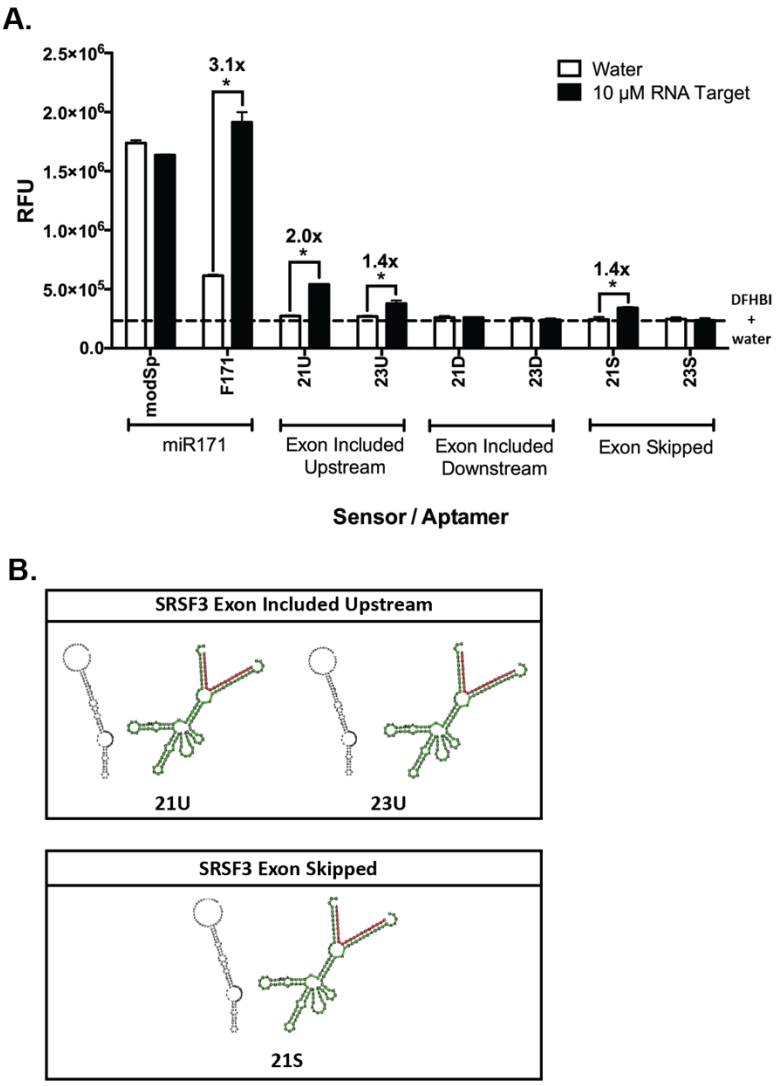


Figure 4.10 Initial screening of FASTmRNA designs.
Screen performed in buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$) with 5 µM sensor and 10 µM target RNA at 25°C. **(A)** Fluorescence data from plate reader screen of SRSF3 sensor sequences. modSpinach and FASTmiR171 are included as controls, with miR171 used as the RNA target for these sequences. Three sensors showed a statistically significant change in fluorescence upon addition of the 200 nt mock mRNA target. **(B)** Structural representation of the OFF (left) and ON (right) configuration for the active sensor designs.

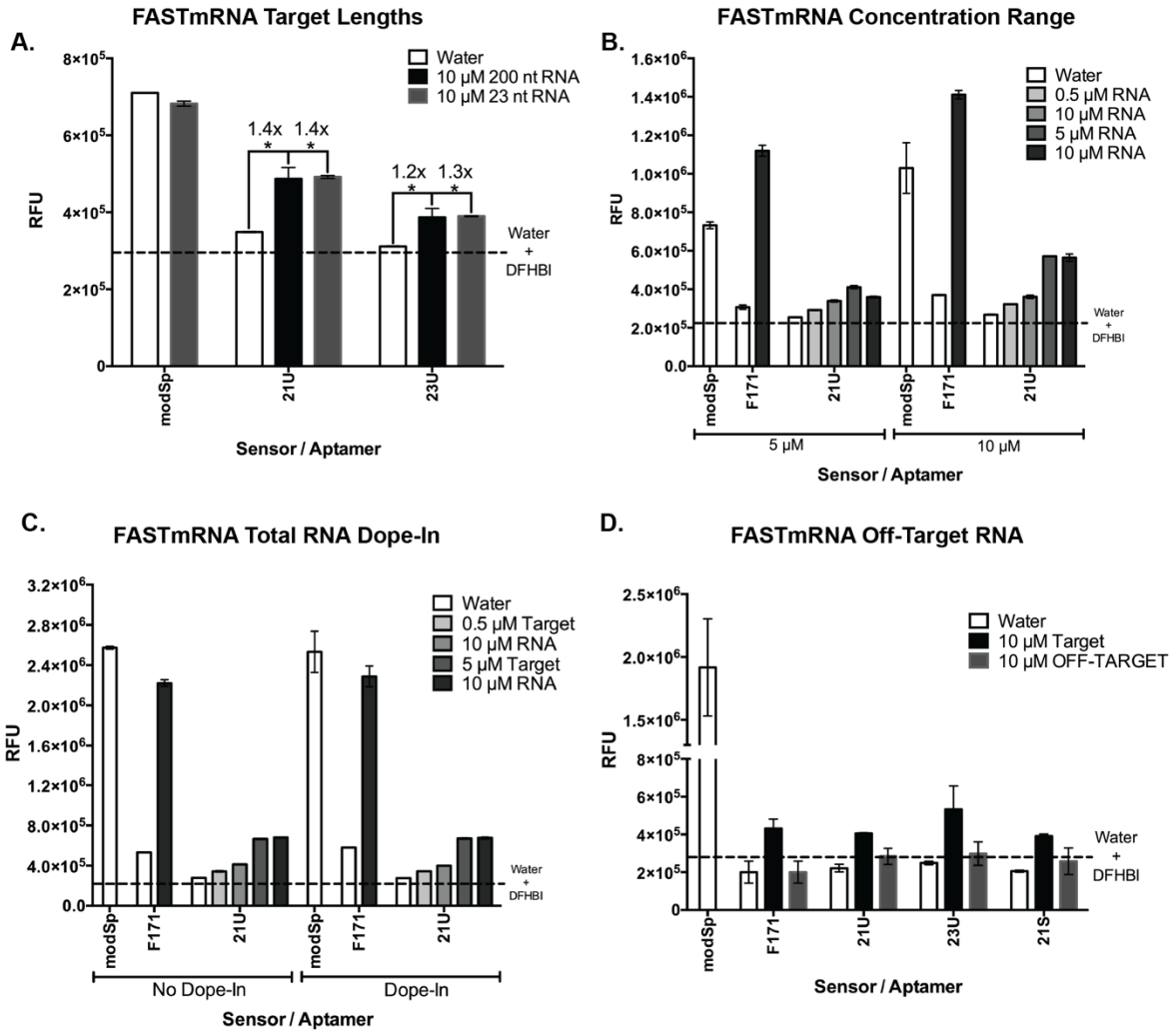


Figure 4.11 Characterizing FASTmRNA sensors *in vitro*.
Fluorescence plate reader data for FASTmRNA sensors under a number of different conditions. Screen performed in buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$) with 5-10 µM sensor and 5-15 µM target RNA at 25°C. **(A)** Mock mRNA targets of different len were compared to determine if 200 nt targets showed reduced signal. The fluorescent output was very similar for both cases. **(B)** Modified screening conditions comparing 2X original sensor/aptamer concentration to see if this would boost maximum signal. There was a slight improvement with the doubled concentration, so these conditions were used moving forward. **(C)** Testing top sensor FAST 21U for fluorescence signal with a total RNA dope-in. *E. coli* BL21 Star total RNA was doped in at a concentration of 1 ng / 30 µL to predict whether cellular conditions would reduce sensor signal. **(D)** Testing sensor hits with 10 µM target RNA versus 10 µM off-target RNA. Off-target sequence for FAST21U and FAST23U was exon-skipped SRSF3, while for FAST21S the exon-included upstream SRSF3 was used.

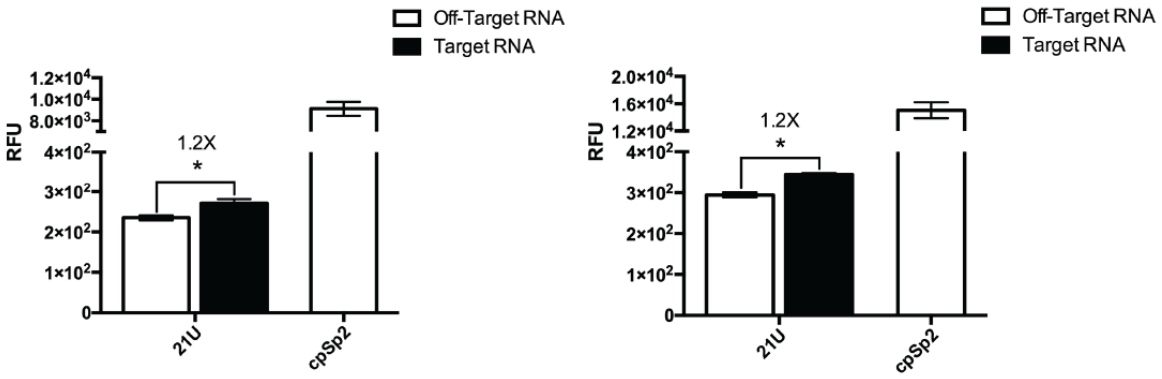


Figure 4.12 FASTmRNA sensor 21IU *in vivo* testing.
Fluorescence plate reader data for FASTmRNA sensor 21U. Sensor was expressed on pETDuet with and without target SRSF3 mRNA sequence. For signal comparison and as an expression control, pET31b cpSpinach2 was used here. The sensor displays a consistent, statistically significant increase in signal, but the overall change is low.

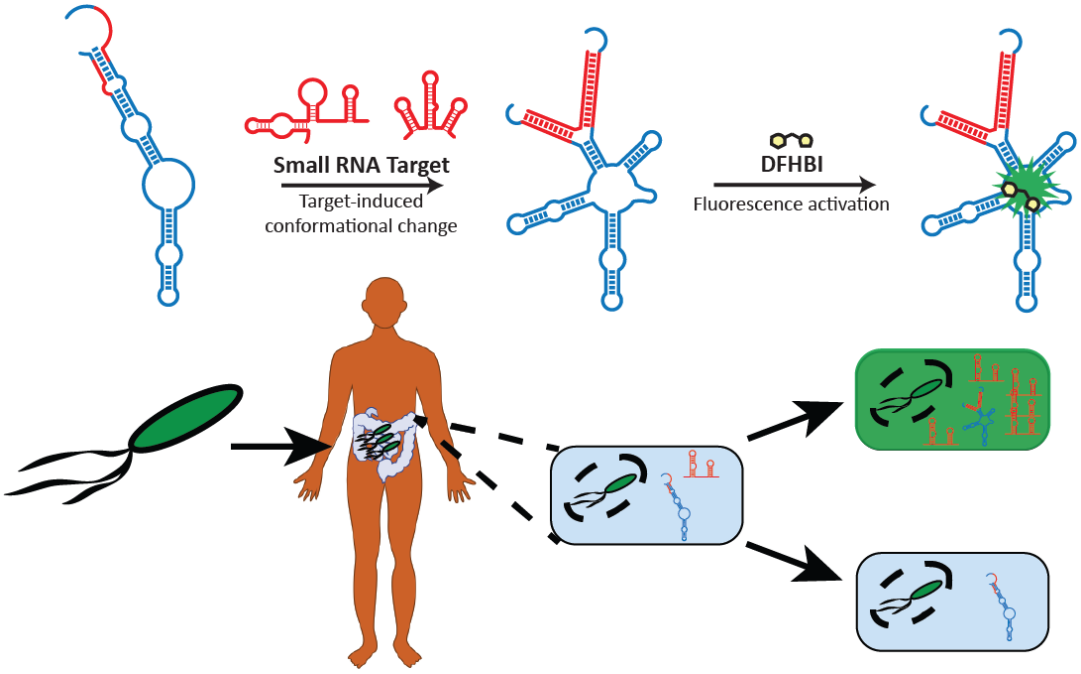


Figure 4.13 FASTsRNA sensor application.
Cartoon illustrating FASTsRNA sensor application to study sRNA infection in the human gut. Bacterial invasion of gut cells could produce high or low levels of sRNA relevant to infection. The FASTsRNA system could provide a fluorescent readout of these levels and monitor changes throughout the course of infection (high signal in green, low signal in light blue).

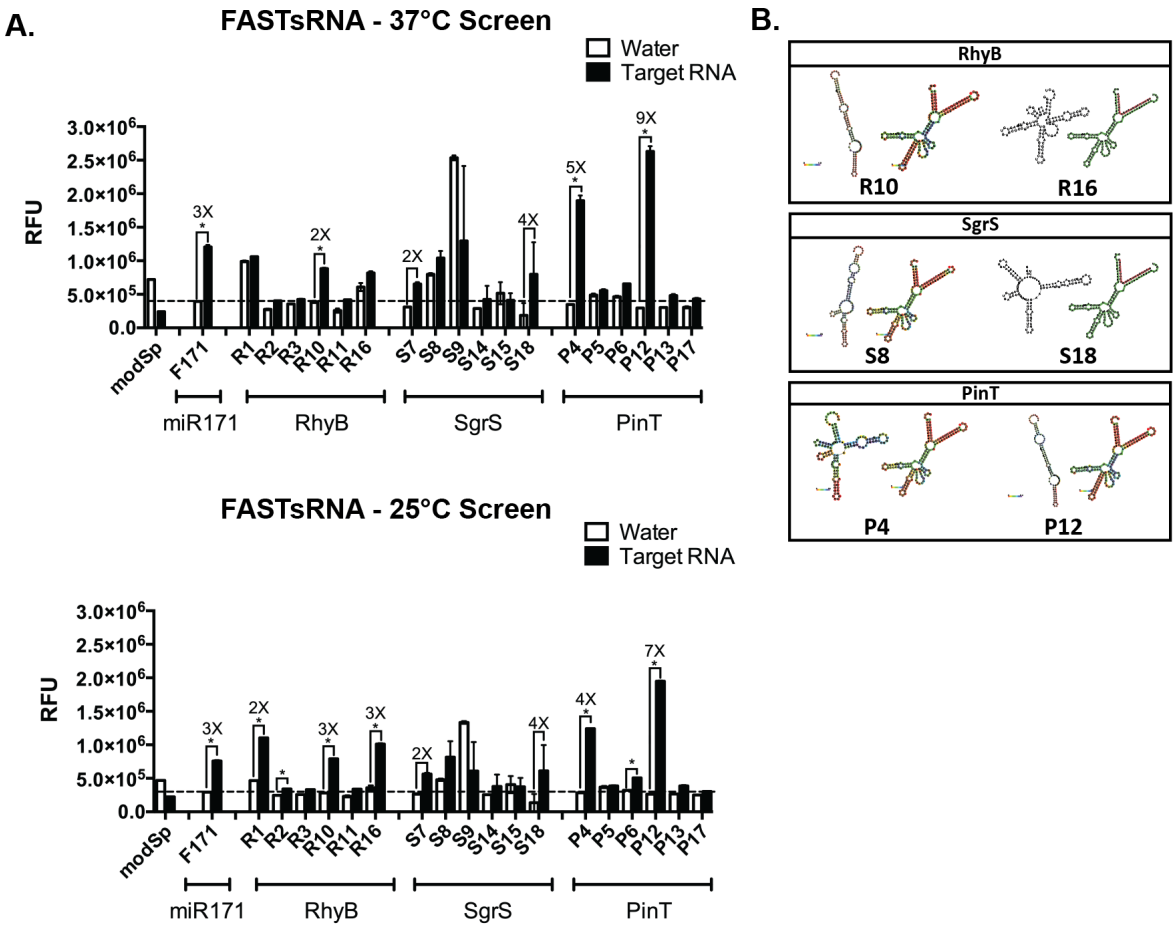


Figure 4.14 Screening FASTsRNA designs *in vitro*.
Fluorescence plate reader data for FASTSRNA sensors at 25°C and 37°C. Screen performed in buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$) with 5 µM sensor and 10 µM target RNA. Several sensors only showed clear turn-on at the higher temperature, and hits were selected from the 37°C data. The chosen designs are shown on the right, with OFF and ON configurations and probability heat mapping at each nucleotide positions. Two sensors without heat mapping (R16 and S18) were predicted to show no turn on as they did not meet the initial design criteria, but surprisingly performed well in the fluorescence assay.

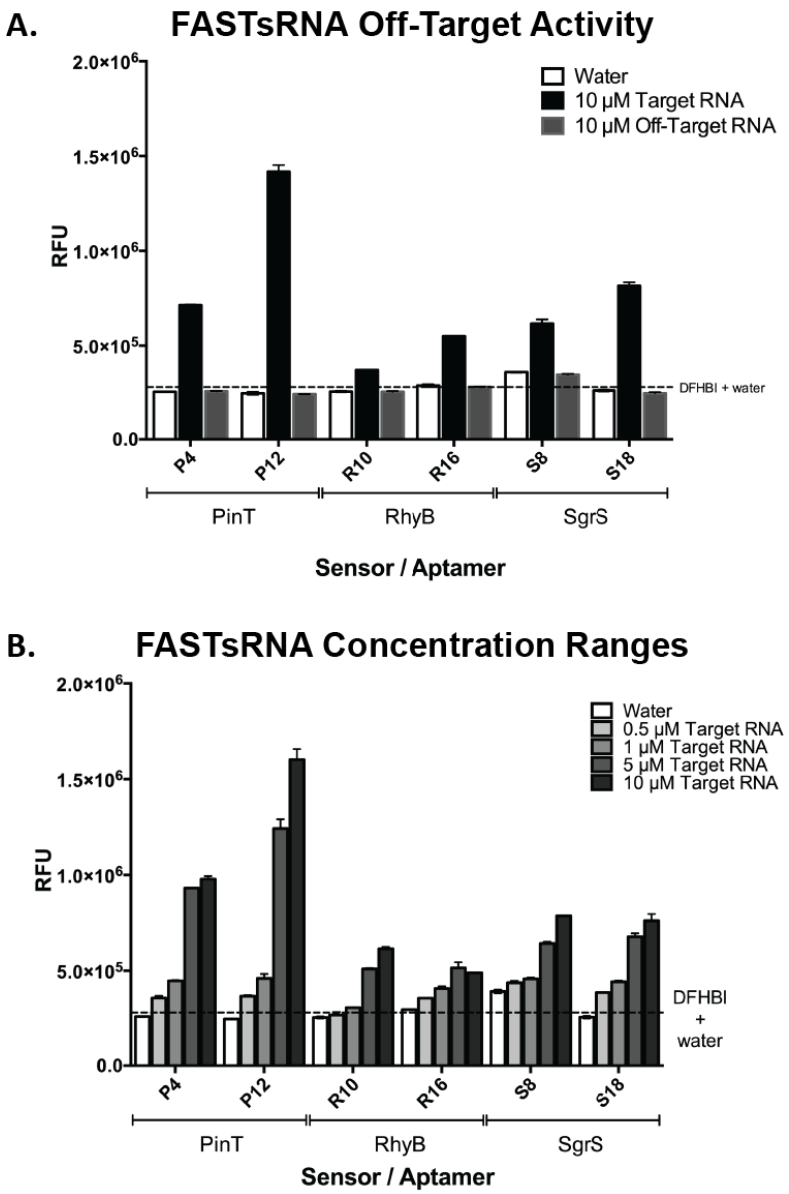


Figure 4.15 Characterizing FASTsRNA sensor hits.
Fluorescence plate reader data for FASTsRNA sensors testing off-target activity and response to different target sRNA concentrations. Screen performed in buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$) with 5 µM sensor and 5-10 µM target RNA at 37°C. **(A)** FASTsRNA designs tested against on- and off-target sRNA for turn-on. PinT sensors were tested with RhyB, RhyB sensors with SgrS, and SgrS sensors with PinT for off-target screen. Fluorescent signal was negligible for off-target sequences (below background, similar to water). **(B)** Sensors were incubated with 5, 10, and 15 µM target sRNA to determine approximate working range. All sensors displayed a response at 5 µM, but some designs were saturated at 10 µM target levels.

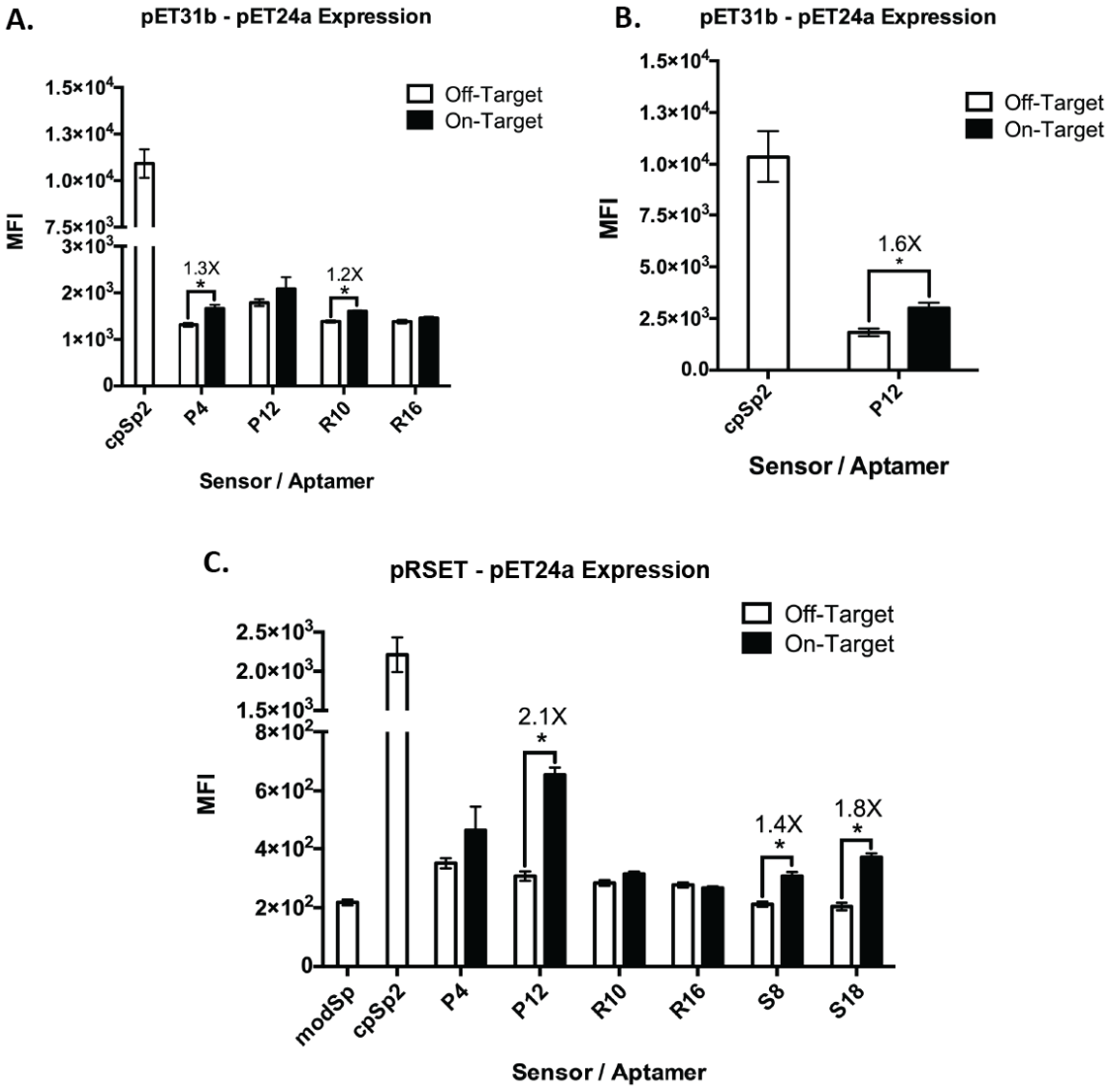


Figure 4.16 FASTsRNA sensor performance *in vivo* under different expression systems.

Flow cytometry data for FASTsRNA sensors under different plasmid expression systems. In all panels, target sRNA is expressed on the pET24a plasmid. **(A)** pET31b expression of sensors showed modest turn-on for a few designs, but required a large number of replicates to prevent error from obscuring the signal change. There was large variability in this experimental setup, as shown in **(B)** where sensor P12 was run on a separate day and produced statistically significant response to target sRNA. **(C)** To prevent the error seen in a pET-pET expression system, sensors were transferred to the pRSET vector. Signal change was more consistent between runs and modest, significant signal change was seen for a few designs. The top hit *in vitro*, P12, also showed the greatest signal change *in vivo*.

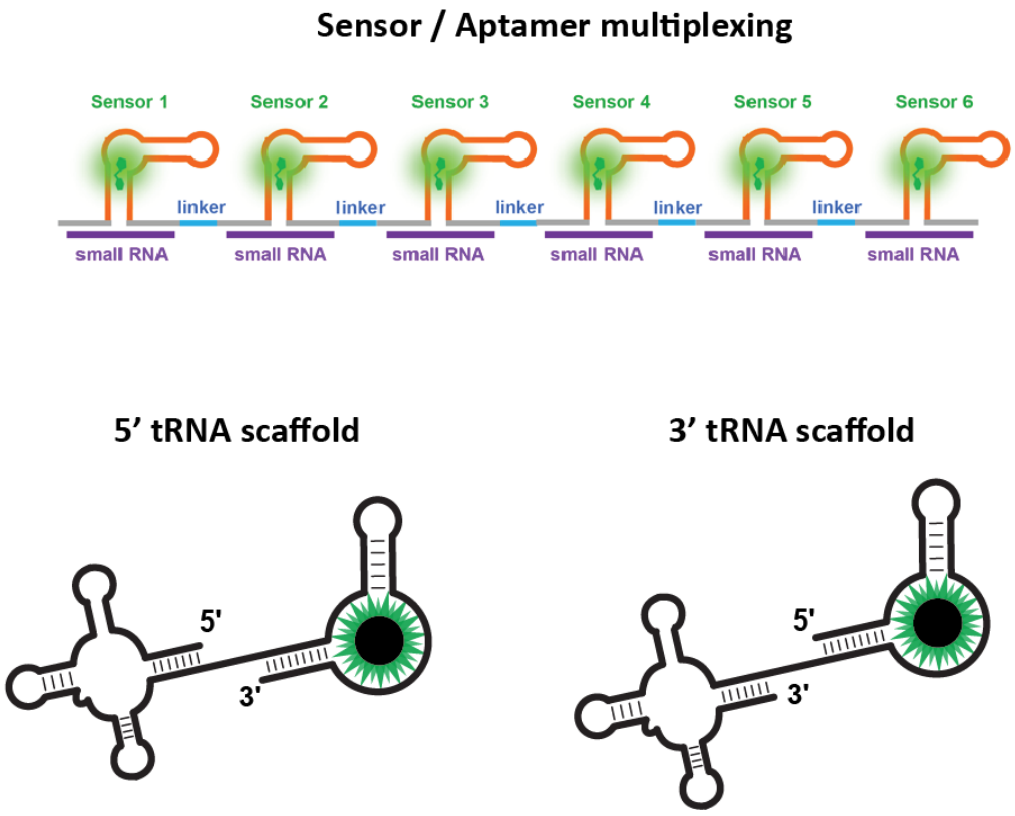


Figure 4.17 Alternate strategies for improved FASTsRNA signal.
Alternative design strategies to improve FASTsRNA sensor performance *in vivo*. The multiplexing approach shown in the top panel (figure adapted with permission from Huang et al.) was shown to improve sensor performance for imaging in eukaryotes[110]. The bottom two diagrams represent alternative tRNA linkage strategies to improve signal without occluding the sRNA toehold binding region at the terminal ends of the sensors. All representative images shown are with aptamers, but these strategies would be employed for both aptamers and sensors.

4.6 Tables

Table 4.1: Promoter, tRNA Scaffold, and RBF biosensor sequences.
Annotations included are: Insertion site for sensor (/-/).

Construct	Sequence (5' -> 3)
T7 Minimal Promoter	TTAATACGACTCACTATAGGG
T7 Extended Promoter	CGATCCCGCGAAATTAATACGACTCACTATAGGG
H. sapiens $tRNA^{Lys}$	GCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGG /-/ CGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA
O. sativa $tRNA^{Ala}$	GGGGGATGTAGCTCAAATGGTAGAGCGCTCGC /-/ GCGAGAGGCACGGGGTTCGATCCCCCGCATCTCCA
Pl-A	GATGTAACTAGTTACGGAGCTCACACTCTACTCAACAACCCTTCGGTAGCCAG GCTGCCATTGACGGTTCATCAAGAAGGCCCTGTGGCTTTGCGTCCTTATCTTT CGACAAGTTTGCCGTTATCGAAGAGTGGACCCGTCCTTCACCATTTCATTCAG TTACATC
Pl-B	GATGTAACTAGTTACGGAGCTCACACTCTACTCAACAACTTCGGTAGCCAGGC TGCCATTGACGGTTCATCAAGAAGGCCCTGTGGCTTTGCGTCCTTATCTTTCG ACAAGTTTGCCGTTATCGAAGTGGACCCGTCCTTCACCATTTCATTCAGTTACA T

Table 4.2: FASTmRNA Sequences.
Sensor-specific regions are in lower case while aptamer backbone is in upper case.200 nt RNA targets are shown below. Please note that these are all DNA templates for final RNA sequences.

Construct	Sequence (5' -> 3')
21U	GGGgatggtgacAAAGGGTACTTGTTGAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCTT CGGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGTGCCTAtctgcgacgaggAAGCT
23U	GGGtgatggtgacAAAGGGTACTTGTTGAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCT TCGGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGTGCCTAtctgcgacgaggtAAG CT
21D	GGGtccttcttgAAAGGGTACTTGTTGAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCTTC GGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGTGCCTAgagattagttttAAGCT
23D	GGGctccttcttgAAAGGGTACTTGTTGAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCTT CGGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGTGCCTAgagattagttttgAAGCT
21S	GGGttggagatcAAAGGGTACTTGTTGAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCTT CGGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGTGCCTAtgcgacgaggtgAAGCT
23S	GGGctccttcttgAAAGGGTACTTGTTGAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCTT CGGACGCAACTGAATGAAATGGTGAAGGACGGGTCCAGTGCCTAgagatctgcgacAAGCT
SRSF3 Exon Included Upstream Fragment	CGAATGGTGAAAAAAGAAGTAGAAATCGTGGCCCACCTCCCTCTTGGGGTCGTCGCCCT CGAGATGATTATCGTAGGAGGAGTCCTCCACCTCGTCGCAGAGTCACCATCATGTCTCTT CTCACCACCCTCTGAATCTGCATTAGCCAGTCAACTAGCCCTTTCAGCGTCATGTGACCAG CGCGCCCCATTCAGCTTGGC
SRSF3 Exon Included Downstream Fragment	TCCAACGCAACATCTGGCAAAACCTTTTCAGCAAATTCTTCCTGGCCGTCAGTCCGGCAG CCTCACCTCACCATTTCTAGCTTGTTGAAACCCAAAACTAATCTCCAAGAAGGAGAAGCTT CTCTCGCAGCCGGAGCAGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGAT CAGCCTCGACTGTGCCTTCT
SRSF3 Exon Skipped Fragment	CGAATGGTGAAAAAAGAAGTAGAAATCGTGGCCCACCTCCCTCTTGGGGTCGTCGCCCT CGAGATGATTATCGTAGGAGGAGTCCTCCACCTCGTCGCAGATCTCCAAGAAGGAGAAG CTTCTCTCGCAGCCGGAGCAGGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCT GATCAGCCTCGACTGTGCCTTCT

Table 4.3: FASTsRNA sequences.

For sensors, upper case sequences are parent aptamer, cpSpinach2 for sequences preceded with 'c' and modSpinach for all others. Lower case regions contain sensory domains and nonspecific overhang regions required for proper folding. Red sequences are controls that did not meet folding criteria in ViennaRNA. Bolded sequences are selected hits from initial screen. sRNA targets are shown at bottom. Please note that these are all DNA template for RNA sequences.

Construct	Sequence (5' -> 3')
R1	agaaatgtcgtAAAGGGTACTTGTTGAGTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGAC GCAACTGAATGAAATGGTGAAGGAtaatccgttactttctt
R2	cactggaagcaaTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATG GTGAAGGAgtgtgagcaatgtaaggg
R3	acttaccggctggcTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAAT GGTGAAGGAtaagtaatactgaaa
P4	gattaaccacaaagTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAAT GGTGAAGGAtaatccgttactttctt
P5	cggctactacgctTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATG GTGAAGGAgccgcaaatatgaaagct
P6	gggatatgaggaTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATG GTGAAGGAggacaattaccgttcc
S7	gttgcaggcaagtTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATG GTGAAGGAcaactttcagaaaagct
S8	cgcgaacttcgcTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATGG TGAAGGAtgtcgcggtaaaaaaa
S9	tggttgggacgctTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATG GTGAAGGAtaaccaacgcaaaaa
R10	gggtggaagcaaAGAGTGTGAGCTCCGTAACTAGTTACATCGCAAGATGTAACTGAATGAAATGG TGAgtgtgagcaatgttttgc
R11	gggctggctaagtaataAGAGTGTGAGCTCCGTAACTAGTTACATCGCAAGATGTAACTGAATGAAA TGGTGActggaagttac
P12	gattaaccacaaagAGAGTGTGAGCTCCGTAACTAGTTACATCGCAAGATGTAACTGAATGAAATG GTGAtaatccgttactttctt
P13	gggatatgaggaAGAGTGTGAGCTCCGTAACTAGTTACATCGCAAGATGTAACTGAATGAAATGG TGAggacaattaccgttcc
S14	aaatgggacgctAGAGTGTGAGCTCCGTAACTAGTTACATCGCAAGATGTAACTGAATGAAATGGT GAtaaccaacgcaaagcg
S15	gcgacaacttcgcAGAGTGTGAGCTCCGTAACTAGTTACATCGCAAGATGTAACTGAATGAAATGG TGAtgtcgcggtaaaaaaa
R16	gggaatgtcgtTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATGGT GAAGGAgctttcaggttctaagct
P17	gggactacgctTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATGG TGAAGGAgccgcaaatatgaaagct
S18	gggcaggcaagtTAGAGTGTGAGCTCCGTAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATG GTGAAGGAcaactttcagaaaagct
PinT	AGTAACGGATTACTTTGTGGTGTAGCGTAACGGTAATTGTCCTCCTCATATTTGCGGCAGCGTA GTCTGCCGCTTTTTTT
RhyB	GCGATCAGGAAGACCCTCGCGGAGAACCTGAAAGCACGACATTGCTCACATTGCTTCCAGTAT TACTTAGCCAGCCGGGTGCTGGCTTTTTTTT

SgrS	GATGAACAAGGGGGTGCCCCATGCGTCAGTTTATCAGCACTATTTTACCGCGACAGCGAAGTT GTGCTGGTTGCGTTGGTTAAGCGTCCCACAACGATTAACCATGCTTGAAGGACTGATGCAGTG GGATGACCGCAATTCTGAAAGTTGACTTGCCTGCATCATGTGTGACTGAGTATTGGTGTAAAAT CACCCGCCAGCAGATTATACCTGCTGGTTTTTTTT

Table 4.4: Free energy and predicted folding frequency of FASTsRNA sensors
ViennaRNA values for predicted change in free energy upon binding to target sRNA and relative frequency of the predicted OFF folding conformation for each FASTsRNA sensor design. No clear trends were obvious for top sensor hits. Top sensors that performed well at 25°C and 37°D are highlighted in blue, while those that only showed signal change at 37°C are highlighted in green.

Sensor	*ΔG (kcal/mol)*	*Frequency of ΔG*
R1	-25.55	0.087
R2	-25.78	0.039
R3	-28.91	0.138
P4	-20.96	0.081
P5	-21.39	0.102
P6	-23.84	0.069
S7	-22.32	0.084
S8	-19.09	0.065
S9	-22.82	0.060
R10	-21.70	0.021
R11	-21.92	0.026
P12	-18.24	0.042
P13	-17.79	0.043
S14	-23.46	0.034
S15	-15.54	0.035
R16	-24.70	0.091
P17	-24.51	0.109
S18	-23.15	0.107

Table 4.5: Oligonucleotide library for FASTmRNA and SRB-2 sensor designs
Lower-case regions are sensor-specific sequences while upper case are the common modSpinach or SRB-2 aptamer backbone. SRB-2 sensors ('MBS') were cloned but never tested due to issues with control aptamers tests.

Primer Name	Sequence (5' -> 3')
21U Fwd	ttaatacgactcactatagggGATGGTGACaaagggtacttgttgag
21U Rev	agcttCCTCGTCGCAGAtaggcactggacccgtcc
23U Fwd	ttaatacgactcactatagggTGATGGTGACaaagggtacttgttgag
23U Rev	agcttACCTCGTCGCAGAtaggcactggacccgtc
21DFwd	ttaatacgactcactatagggTCCTTCTTGaaagggtacttgttgag
21D Rev	agcttAAAACTAATCTCtaggcactggacccgtcc
23D Fwd	ttaatacgactcactatagggCTCCTTCTTGaaagggtacttgttgag
23D Rev	agcttCAAAACTAATCTCtaggcactggacccgtcc
21S Fwd	ttaatacgactcactatagggTTGGAGATCaagggtacttgttgAG
21S Rev	agcttCACCTCGTCGCAtaggcactggacccgtcc
23S Fwd	ttaatacgactcactatagggCTCCTTCTTGaaagggtacttgttgag
23S Rev	agcttGTCGCAGATCTCtaggcactggacccgtcc
MBS3IU_21_6_F	GATGGTGACTCTGCGACGAGGcacctcgcttcggcgat
MBS3IU_21_6_R	CCTCGTCGCAGAGTCACCATCcacctccggccgctgct
MBS3IU_23_7_F	TGATGGTGACTCTGCGACGAGGTcacctcgcttcggcgat
MBS3IU_23_7_R	ACCTCGTCGCAGAGTCACCATCAcacctccggccgctgct
MBS3ID_21_10_F	TCCTTCTTGGAGATTAGTTTTcacctcgcttcggcgat
MBS3ID_21_10_R	AAAACTAATCTCCAAGAAGGAcacctccggccgctgct
MBS3ID_23_11_F	CTCCTTCTTGGAGATTAGTTTTGcacctcgcttcggcgat
MBS3ID_23_11_R	CAAAACTAATCTCCAAGAAGGAGcacctccggccgctgct
MBS3S_21_4_F	TTGGAGATCTGCGACGAGGTGcacctcgcttcggcgat
MBS3S_21_4_R	CACCTCGTCGCAGATCTCCAAcacctccggccgctgct
MBS3S_23_11_F	CTCCTTCTTGGAGATCTGCGACGcacctcgcttcggcgat
MBS3S_23_11_R	CGTCGCAGATCTCCAAGAAGGAGcacctccggccgctgct
FASTmiR Oligo Set 2_1	aaagggtacttgttgAGTAGAGAGTGAGCTCCGTAACTAGTCGCGTCTTCGGACG
FASTmiR Oligo Set 2_2	CGTCCGAAGACGCGACTAGTTACGGAGCTCACTCTCTACTcaacaagtacccttt
FASTmiR Oligo Set 3_1	TAACTAGTCGCGTCTTCGGACGCAACTGAATGAAATGGTGAAGgacgggtccagtgccta
FASTmiR Oligo Set 3_2	taggcactggacccgtcCTTCACCATTTCATTCAGTTGCGTCCGAAGACGCGACTAGTTA

SRB-2 Backbone 1	TTAATACGACTCACTATAGGGGCCCGGATAGCTCAGTCGGTAGAGCAGCGGCCGGAGGTG
SRB-2 Backbone 2	CACCTCCGGCCGCTGCTCTACCGACTGAGCTATCCGGGCCCCTATAGTGAGTCGTATTAA
SRB-2 Backbone 3	CACCTCGCTTCGGCGATGATGGAGAGGCGCAAGGTTAACCGCCTCAGGTGAAAAACGGCC
SRB-2 Backbone 4	GGCCGTTTTTCACCTGAGGCGGTTAACCTTGCGCCTCTCCATCATCGCCGAAGCGAGGTG
SRB-2 Backbone 5	TAACCGCCTCAGGTGAAAAACGGCCGCGGGTCCAGGGTTCAAGTCCCTGTTCGGGCGCCA
SRB-2 Backbone 6	TGGCGCCCGAACAGGGACTTGAACCCTGGACCCGCGGCCGTTTTTCACCTGAGGCGGTTA

Table 4.6: Oligonucleotide library for FASTsRNA designs

For sensors, upper case sequences are parent aptamer, cpSpinach2 for sequences preceded with 'c' and modSpinach for all others. Lower case regions contain sensory domains (underlined) and nonspecific overhang regions required for proper folding. Red sequences are controls that did not meet folding criteria in ViennaRNA. Bolded sequences are selected hits from initial screen. sRNA targets are shown at bottom.

Primer Name	Sequence (5' -> 3')
R1 Fwd	TTAATACGACTCACTATAGagaaatgtcgtAAAGGGTACTTGTTGAG
R2 Fwd	TTAATACGACTCACTATAGcactggaagcaaAAAGGGTACTTGTTGAG
R3 Fwd	TTAATACGACTCACTATAGacttaccggctggcAAAGGGTACTTGTTGAG
P4 Fwd	TTAATACGACTCACTATAgattaaccacaaagAAAGGGTACTTGTTGAG
P5 Fwd	TTAATACGACTCACTATAGcggctactacgctAAAGGGTACTTGTTGAG
P6 Fwd	TTAATACGACTCACTATAgggatatgaggaAAAGGGTACTTGTTGAG
S7 Fwd	TTAATACGACTCACTATAgttgcaggcaagtAAAGGGTACTTGTTGAG
S8 Fwd	TTAATACGACTCACTATAGcgcgaacttcgcAAAGGGTACTTGTTGAG
S9 Fwd	TTAATACGACTCACTATAGtggttgggacgctAAAGGGTACTTGTTGAG
R10 Fwd	TTAATACGACTCACTATAgggtggaagcaaAAAGGGTATTGTTGAGT
R11 Fwd	TTAATACGACTCACTATAgggctggctaagtaataAAAGGGTATTGTTGAGT
P12 Fwd	TTAATACGACTCACTATAgattaaccacaaagAAAGGGTATTGTTGAGT
P13 Fwd	TTAATACGACTCACTATAgggatatgaggaAAAGGGTATTGTTGAGT
R14 Fwd	TTAATACGACTCACTATAGaaatgggacgctAAAGGGTATTGTTGAGT
S15 Fwd	TTAATACGACTCACTATAgcgacaacttcgcAAAGGGTATTGTTGAGT
R16 Fwd	TTAATACGACTCACTATAgggaatgtcgtAAAGGGTACTTGTTGAG
P17 Fwd	TTAATACGACTCACTATAgggactacgctAAAGGGTACTTGTTGAG
S18 Fwd	TTAATACGACTCACTATAgggcaggcaagtAAAGGGTACTTGTTGAG
R1 Rev	ttgttagaacctgaaagcTAGGCACTGGACCCGTCC
R2 R	cccttacattgctcacacTAGGCACTGGACCCGTCC
R3 Rev	tttcagtattacttaTAGGCACTGGACCCGTCC
P4 Rev	aagaaagtaacggattaTAGGCACTGGACCCGTCC
P5 Rev	agctttcatatttgcggcTAGGCACTGGACCCGTCC
P6 Rev	ggaacggtaattgtccTAGGCACTGGACCCGTCC
S7 Rev	agcttttctgaaagttgTAGGCACTGGACCCGTCC
S8 Rev	ttttttttaccgcgacaTAGGCACTGGACCCGTCC
S9 Rev	tttttgcgttggttaTAGGCACTGGACCCGTCC
R10 Rev	gcaaaacattgctcacacTAGGCATGGACCCGTCCT
R11 Rev	gtaacttccagTAGGCATGGACCCGTCCT
R12 Rev	aagaaagtaacggattaTAGGCATGGACCCGTCCT
P13 Rev	ggaacggtaattgtccTAGGCATGGACCCGTCCT
R14 Rev	cgctttgcgttggttaTAGGCATGGACCCGTCCT
15 Rev	ttttttttaccgcgacaTAGGCATGGACCCGTCCT

R16 Rev	agcttagaacctgaaagcTAGGCACTGGACCCGTCC
P17 Rev	agctttcatatttgcggcTAGGCACTGGACCCGTCC
S18 Rev	agcttttctgaaagttgTAGGCACTGGACCCGTCC

5 References

1. Leavitt, S. A. Deciphering the Genetic Code: Marshall Nirenberg. *Office of NIH History* (2010).
2. Weinberg, R. & Penman, S. Metabolism of small molecular weight monodisperse nuclear RNA. *BBA Sect. Nucleic Acids Protein Synth.* **190**, 10–29 (1969).
3. Salditt-Georgieff, M. & Darnell, J. E. Further Evidence That the Majority of Primary Nuclear RNA Transcripts in Mammalian Cells Do Not Contribute to mRNA. *Mol. Cell. Biol.* **2**, 701–707 (2015).
4. Paul, I. J. & Duerksen, J. D. Chromatin-associated RNA content of heterochromatin and euchromatin. *Mol. Cell. Biochem.* **9**, 9–16 (1975).
5. Reddy, A. S. N., Marquez, Y., Kalyna, M. & Barta, A. Complexity of the alternative splicing landscape in plants. *Plant Cell* **25**, 3657–83 (2013).
6. Filichkin, S., Priest, H. D., Megraw, M. & Mockler, T. C. Alternative splicing in plants: Directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* **24**, 125–135 (2015).
7. Lee, Y. & Rio, D. C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.* 1–33 (2015). doi:10.1146/annurev-biochem-060614-034316
8. Clark, M. B. *et al.* The Reality of Pervasive Transcription. *PLoS Biol.* **9**, e1000625 (2011).
9. van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R. Most 'dark matter' transcripts are associated with known genes. *PLoS Biol.* **8**, (2010).
10. Guil, S. & Esteller, M. Cis-acting noncoding RNAs: Friends and foes. *Nat. Struct. Mol. Biol.* **19**, 1068–1075 (2012).
11. Lam, J. K. W., Chow, M. Y. T., Zhang, Y. & Leung, S. W. S. siRNA Versus miRNA as Therapeutics for Gene Silencing. 1–20 (2015). doi:10.1038/mtna.2015.23
12. He, L. & Hannon, G. J. MicroRNAs: Small RNAs with a big role in gene regulation. *Nature Reviews Genetics* **5**, 522–531 (2004).
13. Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**, 57–65 (2004).
14. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).
15. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology* **5**, 451–463 (2004).
16. Serganov, A. & Patel, D. J. Metabolite Recognition Principles and Molecular Mechanisms Underlying Riboswitch Function. *Annu. Rev. Biophys.* **41**, 343–370 (2012).
17. Sherwood, A. V. & Henkin, T. M. Riboswitch-Mediated Gene Regulation: Novel RNA Architectures Dictate Gene Expression Responses. *Annu. Rev. Microbiol.* **70**, 361–374 (2016).
18. Hallberg, Z. F., Su, Y., Kitto, R. Z. & Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem* **86**, 515–39 (2017).
19. Carrier, M., Lalaouna, D. & Massé, E. Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action. *Annu. Rev. Microbiol.* **72**, 141–161 (2018).
20. Villa*, J. K., Su*, Y., Contreras, L. M. & Hammond, M. C. Synthetic Biology of Small RNAs and Riboswitches. *Microbiol. Spectr.* **6**, 25–32 (2018).

21. Westermann, A. J. *et al.* Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* **529**, 496–501 (2016).
22. Bobrovskyy, M. & Vanderpool, C. K. The small RNA SgrS: roles in metabolism and pathogenesis of enteric bacteria. *Front. Cell. Infect. Microbiol.* **4**, 1–8 (2014).
23. Rousseau, C., Gonnet, M., Le Romancer, M. & Nicolas, J. CRISPI: A CRISPR interactive database. *Bioinformatics* **25**, 3317–3318 (2009).
24. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 1–10 (2007).
25. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Current Opinion in Microbiology* **14**, 321–327 (2011).
26. Philippe, H. & Rodolphe, B. CRISPR/Cas, the immune system of bacteria and archaea. *Science (80-.).* **327**, 167–170 (2010).
27. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
28. Dicarlo, J. E. *et al.* Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
29. Cáceres, J. F. *et al.* Alternative splicing: multiple control mechanisms and involvement in human disease. *Nat. Biotechnol.* **22**, 186–193 (2002).
30. Gómez Acuña, L. I., Fiszbein, A., Alló, M., Schor, I. E. & Kornblihtt, A. R. Connections between chromatin signatures and splicing. *Wiley Interdiscip. Rev. RNA* **4**, 77–91 (2013).
31. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell* **43**, 880–891 (2011).
32. Crick, F. H. *On protein synthesis*. (Symp Soc Exp Biol, 1958).
33. Cáceres, J. F. & Kornblihtt, A. R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18**, 186–193 (2002).
34. Will, C. L. & Max, R. L. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3**, 2875–2881 (2002).
35. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* **18**, 437–451 (2017).
36. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).
37. Mata, M. De *et al.* A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Mol. Cell* **12**, 525–532 (2003).
38. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. *Annu. Rev. Biochem.* **84**, 165–198 (2015).
39. Cooper, D. N. *et al.* Human Gene Mutation Database. *Hum Mutat* 577–581 (2019).
40. Hammond, M. C., Wachter, A. & Breaker, R. R. A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. *Nat. Struct. Mol. Biol.* **16**, 541–549 (2009).
41. Hickey, S. F. *et al.* Transgene regulation in plants by alternative splicing of a suicide exon. *Nucleic Acids Res.* **40**, 4701–4710 (2012).
42. Gonzalez, T. L. *et al.* Tight regulation of plant immune responses by combining promoter

and suicide exon elements. *Nucleic Acids Res.* **43**, 7152–7161 (2015).
43. Guo, X. *et al.* Efficient RNA/Cas9-mediated genome editing in Xenopus tropicalis. *Development* **141**, 707–14 (2014).
44. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–6 (2013).
45. Liang, Y. *et al.* Endoribonuclease-Based Two-Component Repressor Systems for Tight Gene Expression Control in Plants. *ACS Synth. Biol.* **6**, 806–816 (2017).
46. Callis, J., Fromm, M. & Walbot, V. Introns increase gene expression in cultured maize cells. *Genes Dev.* **1**, 1183–1200 (1987).
47. Choi, T., Huang, M., Gorman, C. & Jaenisch, R. A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol.* **11**, 3070–3074 (2015).
48. Le Hir, H., Nott, A. & Moore, M. J. How introns influence and enhance eukaryotic gene expression. *Trends Biochem. Sci.* **28**, 215–220 (2003).
49. Gallegos, J. E. & Rose, A. B. The enduring mystery of intron-mediated enhancement. *Plant Sci.* **237**, 8–15 (2015).
50. Buggiano, V. *et al.* Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* **16**, 717–725 (2009).
51. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, 510–7 (2015).
52. Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
53. Thakore, P. I. *et al.* Highly specific epigenome editing by {CRISPR-Cas9} repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143–1149 (2015).
54. Cherblanc, F. L., Chapman, K. L., Brown, R. & Fuchter, M. J. Chaetocin is a nonspecific inhibitor of histone lysine methyltransferases. *Nat. Chem. Biol.* **9**, 136–7 (2013).
55. Pappano, W. N. *et al.* The Histone methyltransferase inhibitor A-366 uncovers a role for G9a/GLP in the epigenetics of leukemia. *PLoS One* **10**, 1–13 (2015).
56. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
57. Wang, Q. & Rio, D. C. JUM is a computational method for comprehensive annotation-free analysis of alternative pre-mRNA splicing patterns. *Proc. Natl. Acad. Sci.* **115**, E8181–E8190 (2018).
58. Zhang, F. CRISPR/guides selection server. Available at: http://crispr.mit.edu. (Accessed: 20th July 2006)
59. Kent, W. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
60. Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
61. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* **33**, 755–760 (2015).
62. Nihongaki, Y. *et al.* CRISPR-Cas9-based photoactivatable transcription systems to induce neuronal differentiation. *Nat. Methods* **14**, 963–966 (2017).
63. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing

and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
64. Wright, A. V *et al.* Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2984–9 (2015).
65. Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11**, 198–200 (2015).
66. Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
67. Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical control of CRISPR/Cas9 gene editing. *J. Am. Chem. Soc.* **137**, 5642–5645 (2015).
68. Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676.e14 (2018).
69. Jillette, N. & Cheng, A. W. CRISPR Artificial Splicing Factors. *bioRxiv* 431064 (2018). doi:10.1101/431064
70. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
71. Cowles, J. R., Scheld, H. W., Lemay, R. & Peterson, C. Growth and lignification in seedlings exposed to eight days of microgravity. *Ann. Bot.* 33–48 (1984).
72. Tripathy, B. C., Brown, C. S., Levine, H. G. & Krikorian, A. D. Growth and Photosynthetic Responses of Wheat Plants Grown in Space. *Plant Physiol.* **110**, 801–806 (2002).
73. Matía, I. *et al.* Plant cell proliferation and growth are altered by microgravity conditions in spaceflight. *J. Plant Physiol.* **167**, 184–193 (2010).
74. Paul, A.-L. *et al.* Parabolic Flight Induces Changes in Gene Expression Patterns in *Arabidopsis thaliana*. *Astrobiology* **11**, 743–758 (2011).
75. Millar, K. D. L., Johnson, C. M., Edelmann, R. E. & Kiss, J. Z. An Endogenous Growth Pattern of Roots Is Revealed in Seedlings Grown in Microgravity. *Astrobiology* **11**, 787–797 (2011).
76. Massa, G. D. & Levine, H. G. Veg-03 Experiment. (2019).
77. Yurkevich, O. Y. *et al.* Molecular Cytogenetics of Pisum sativum L. Grown under Spaceflight-Related Stress . *Biomed Res. Int.* **2018**, 1–10 (2018).
78. Ferl, R. J. & Paul, A.-L. The effect of spaceflight on the gravity-sensing auxin gradient of roots: GFP reporter gene microscopy on orbit. *npj Microgravity* **2**, 15023 (2016).
79. Zabel, P., Bamsey, M., Schubert, D. & Tajmar, M. Review and analysis of over 40 years of space plant growth systems. *Life Sci. Sp. Res.* **10**, 1–16 (2016).
80. Kordyum, E. L. Biology of Plant Cells in Microgravity and under Clinostating. 1–78 (2008). doi:10.1016/s0074-7696(08)62585-1
81. Ueda, J. *et al.* Growth and Development, and Auxin Polar Transport in Higher Plants under Microgravity Conditions in Space: BRIC-AUX on STS-95 Space Experiment. *J. Plant Res.* **112**, 487–492 (2006).
82. Kiss, J. Z., Brinckmann, E. & Brillouet, C. Development and Growth of Several Strains of Arabidopsis Seedlings in Microgravity . *Int. J. Plant Sci.* **161**, 55–62 (2002).
83. Braun, M., Buchen, B. & Sievers, A. Actomysin-Mediated Statolith Positioning in Gravisensing Plant Cells Studied in Microgravity. *J. Plant Growth Regul.* **21**, 137 (2002).
84. Soga, K., Wakabayashi, K., Kamisaka, S. & Hoson, T. Stimulation of elongation growth and

xyloglucan breakdown in Arabidopsis hypocotyls under microgravity conditions in space. *Planta* **215**, 1040–1046 (2002).
85. Porterfield, D. M. The biophysical limitations in physiological transport and exchange in plants grown in microgravity. *J. Plant Growth Regul.* **43**, 1067–1071 (2002).
86. Hoson, T. *et al.* Cell wall changes involved in the automorphic curvature of rice coleoptiles under microgravity conditions in space. *J. Plant Res.* **117**, 449–455 (2004).
87. Stutte, G. W. *et al.* Microgravity effects on leaf morphology, cell structure, carbon metabolism and mRNA expression of dwarf wheat. *Planta* **224**, 1038–1049 (2006).
88. Menezes, A. A., Montague, M. G., Cumbers, J., Hogan, J. A. & Arkin, A. P. Grand challenges in space synthetic biology. *J. R. Soc. Interface* **12**, (2015).
89. Bowman, R. N. Personal communication. (2018).
90. Wielopolska, A., Townley, H., Moore, I., Waterhouse, P. & Helliwell, C. A high-throughput inducible RNAi vector for plants. *Plant Biotechnol. J.* **3**, 583–590 (2005).
91. Guo, H. S., Fei, J. F., Xie, Q. & Chua, N. H. A chemical-regulated inducible RNAi system in plants. *Plant J.* **34**, 383–392 (2003).
92. Kochian, L. V. Cellular Mechanisms of Aluminum Toxicity and Resistance in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 237–260 (2003).
93. Wilson, R. C. & Doudna, J. A. Molecular Mechanisms of RNA Interference. *Annu. Rev. Biophys.* **42**, 217–239 (2013).
94. Wright, A. V, Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29–44 (2016).
95. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17–24 (2013).
96. Okumoto, S., Jones, A. & Frommer, W. B. Quantitative Imaging with Fluorescent Biosensors. *Annu. Rev. Plant Biol.* **63**, 663–706 (2012).
97. Wang, X. C., Wilson, S. C. & Hammond, M. C. Next-generation fluorescent RNA biosensors enable anaerobic detection of cyclic di-GMP. 1–10 (2016). doi:10.1093/nar/gkw580
98. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci.* **91**, 12501–12504 (1994).
99. Stojanovic, M. N. & Kolpashchikov, D. M. Modular aptameric sensors. *J. Am. Chem. Soc.* **126**, 9266–9270 (2004).
100. Paige, J. S., Nguyen-Duc, T., Song, W. & Jaffrey, S. R. Fluorescence Imaging of Cellular Metabolites with RNA. *Science (80-.).* **335**, 1194–1194 (2012).
101. Jepsen, M. D. E. *et al.* Development of a genetically encodable FRET system using fluorescent RNA aptamers. *Nat. Commun.* **9**, 1–10 (2018).
102. Kellenberger, C. A., Wilson, S. C., Sales-Lee, J. & Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc.* **135**, 4906–4909 (2013).
103. Kellenberger, C. A., Chen, C., Whiteley, A. T., Portnoy, D. A. & Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messenger Cyclic di-AMP. *J. Am. Chem. Soc.* **137**, 6432–5 (2015).
104. Bose, D., Su, Y., Marcus, A., Raulet, D. H. & Hammond, M. C. An RNA-Based Fluorescent Biosensor for High-Throughput Analysis of the cGAS-cGAMP-STING Pathway. *Cell Chem.*

Biol. **23**, 1–11 (2016).
105. Su, Y., Hickey, S. F., Keyser, S. G. L. & Hammond, M. C. In vitro and in vivo enzyme activity screening via RNA-based fluorescent biosensors for S-adenosyl-L-homocysteine (SAH). *J. Am. Chem. Soc.* jacs.6b01621 (2016). doi:10.1021/jacs.6b01621
106. You, M., Litke, J. L. & Jaffrey, S. R. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E2756-65 (2015).
107. Ketterer, S., Gladis, L., Kozica, A. & Meier, M. Engineering and characterization of fluorogenic glycine riboswitches. *Nucleic Acids Res.* **44**, gkw465 (2016).
108. Bhadra, S. & Ellington, A. D. A Spinach molecular beacon triggered by strand displacement. *Rna* **20**, 1183–1194 (2014).
109. Alam, K. K., Tawiah, K. D., Lichte, M. F., Porciani, D. & Burke, D. H. A Fluorescent Split Aptamer for Visualizing RNA-RNA Assembly in Vivo. *ACS Synth. Biol.* **6**, 1710–1721 (2017).
110. Huang, K. *et al.* FASTmiR: an RNA-based sensor for in vitro quantification and live-cell localization of small RNAs. *Nucleic Acids Res.* **45**, e130 (2017).
111. Ying, Z. M., Wu, Z., Tu, B., Tan, W. & Jiang, J. H. Genetically Encoded Fluorescent RNA Sensor for Ratiometric Imaging of MicroRNA in Living Tumor Cells. *J. Am. Chem. Soc.* **139**, 9779–9782 (2017).
112. Karunanayake Mudiyanselage, A. P. K. K. *et al.* Genetically Encoded Catalytic Hairpin Assembly for Sensitive RNA Imaging in Live Cells. *J. Am. Chem. Soc.* **140**, 8739–8745 (2018).
113. Kellenberger, C. a. *et al.* GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5383–8 (2015).
114. Sunbul, M. & Jäschke, A. Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. *Angew. Chemie - Int. Ed.* **52**, 13401–13404 (2013).
115. Dolgosheina, E. V. *et al.* RNA Mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking. *ACS Chem. Biol.* **9**, 2412–2420 (2014).
116. Sato, S. I. *et al.* Live-cell imaging of endogenous mRNAs with a small molecule. *Angew. Chemie - Int. Ed.* **54**, 1855–1858 (2015).
117. Halliday, N. M., Hardie, K. R., Williams, P., Winzer, K. & Barrett, D. A. Quantitative liquid chromatography-tandem mass spectrometry profiling of activated methyl cycle metabolites involved in LuxS-dependent quorum sensing in Escherichia coli. *Anal. Biochem.* **403**, 20–29 (2010).
118. Hallberg, Z. F. *et al.* Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP). *Proc. Natl. Acad. Sci.* **113**, 201515287 (2016).
119. Ponchon, L. & Dardel, F. Recombinant RNA technology: the tRNA scaffold. *Nat. Methods* **4**, 571–576 (2007).
120. Wang, X. GEMM-I riboswitch-based fluorescent biosensors for live cell analysis of cyclic dinucleotide signaling. (University of California, Berkeley, 2016).
121. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
122. Lareau, L. F. & Brenner, S. E. Regulation of Splicing Factors by Alternative Splicing and NMD Is Conserved between Kingdoms Yet Evolutionarily Flexible. *Mol. Biol. Evol.* **32**,

1072–1079 (2015).
123. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
124. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathog.* **13**, 1–19 (2017).
125. Massé, E., Vanderpool, C. K. & Gottesman, S. Effect of RyhB small RNA on global iron use in Escherichia coli. *J. Bacteriol.* **187**, 6962–6971 (2005).
126. Massé, E. & Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 4620–4625 (2002).
127. Massé, E., Escorcia, F. E. & Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. 2374–2383 (2003). doi:10.1101/gad.1127103
128. Wadler, C. S. & Vanderpool, C. K. A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci.* **104**, 20454–20459 (2007).
129. Ren, A. *et al.* Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. *Cell Rep.* **11**, 1–12 (2015).
130. Kellenberger, C. A. & Hammond, M. C. *In vitro analysis of riboswitch-spinach aptamer fusions as metabolite-sensing fluorescent biosensors*. *Methods in Enzymology* **550**, (Elsevier Inc., 2015).
131. Wilson, S. C., Cohen, D. T., Wang, X. I. N. C. & Hammond, M. C. A neutral pH thermal hydrolysis method for quantification of structured RNAs. 1–9 (2014). doi:10.1261/rna.045856.114.3
132. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–34 (2005).
133. Arora, A., Sunbul, M. & Jäschke, A. Dual-colour imaging of RNAs using quencher- and fluorophore-binding aptamers. *Nucleic Acids Res.* **43**, gkv718 (2015).