

RNA Sensing and Programming Platforms for
Mammalian Synthetic Biology

by

Breanna Elizabeth DiAndreth



B.S. in Chemical Engineering with Biomedical Engineering Double
Major, Carnegie Mellon University, 2013

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

© Massachusetts Institute of Technology 2019. All rights reserved.

Signature redacted

Author
Department of Biological Engineering
April 26, 2019

Signature redacted

Certified by...
.....
Ron Weiss, PhD
Professor of Biological Engineering, MIT
Thesis Supervisor

Signature redacted

Accepted by
.....
Forest M. White, PhD
Professor of Biological Engineering
Chair, Department Committee on Graduate Theses

This doctoral thesis has been examined by the following committee:

Doug Lauffenburger, PhD

Chairman, Thesis Committee

Ford Professor of Biological Engineering, Chemical Engineering and Biology, MIT

Ron Weiss, PhD

Thesis Supervisor

Professor of Biological Engineering, MIT

James J. Collins, PhD

Termeer Professor of Medical Engineering and Science, Department of Biological Engineering, MIT

RNA Sensing and Programming Platforms for Mammalian Synthetic Biology

by

Breanna Elizabeth DiAndreth

Submitted to the Department of Biological Engineering
on April 26, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The field of synthetic biology aims to control cellular behavior using programmable gene circuits. Generally these gene circuits sense molecular biomarkers, process these inputs and execute a desired calculated response. This is especially relevant for gene and cell therapies where integrating multiple disease-related inputs and/or sophisticated control could lead to safer and more effective approaches.

While mammalian synthetic biology has made great progress, few gene circuit-based therapies have entered the clinic. Regulatory issues aside, this lag may be due to several technical impediments. First, the computing part of circuits is often accomplished via transcriptional regulation, which presents challenges as we move toward the clinic. Second, the field relies on a limited set of sensors; the detection of other types of disease biomarkers will help robustly identify cell state. Finally, the design cycle currently used to develop gene circuits is laborious and slow, which is not suitable for clinical development, especially applications in personalized medicine. In this thesis I describe how I address these three limitations. I develop a new post-transcriptional regulation platform based on RNA cleavage that I term "PERSIST" (Programmable Endonucleolytic RNA Scission-Induced Stability Tuning). CRISPR-specific endonucleases are adapted as RNA-level regulators for the platform and we demonstrate several genetic devices including cascades, feedback, logic functions and a bistable switch. I explore sensor designs for relevant biomolecules including mRNAs, miRNAs and proteins via the PERSIST and other platforms. Finally, I present a "poly-transfection" method, associated advanced data analysis pipelines, and computational models that make circuit engineering faster and more predictive.

Taken together, the expanded RNA toolkit that the PERSIST platform offers as well as advancements in sensing and circuit design will enable the more straightforward creation of robust gene circuits for gene and cell therapies.

Thesis Supervisor: Ron Weiss, PhD
Title: Professor of Biological Engineering, MIT

Acknowledgments

It's impossible to mention all of the people that have made this dissertation possible but here is my valiant attempt.

First, a special thank you to mentors at Carnegie Mellon - Ge Yang, Conrad Zapanta, and Kris Dahl - I would have never considered pursuing my PhD if it wasn't for your encouragement, inspiration and guidance.

To my advisor, Ron Weiss, thank you for imparting your passion for "building with biology". The curiosity and excitement with which you approach the work we do is quite contagious and a big part of why I joined your lab. Thank you for your patience and all of the opportunities you gave me to learn and grow scientifically (writing the Science review, writing grants and patents, and building a new platform from scratch) and non-scientifically (learning to snowboard and surf).

I am indebted to my other thesis committee members, Jim Collins and Doug Lauffenburger. Jim, it has been a privilege to receive guidance from such a grounded and brilliant scientist and leader of the field. I can't thank you enough for your support in navigating my PhD and career advice. And Doug, you have always inspired me with the thoughtfulness with which you led my thesis committee and the Department as a whole. You find a way to balance the rigor and scientific contributions with the goals and ambitions of every student, an aspect that is evident in the unique BE culture and for which I will always be grateful.

I've had the honor of collaborating with some remarkable individuals. Thank you for your scientific support, guidance and insight throughout our various projects: Eerik Kaseniit, Liliana Wroblewska, Tasuku Kitada, Velia Siciliano, Blandine Monel, Kiera Clayton, Brian Teague, Jeremy Gam, Ross Jones, Jin Huh, Kevin Lebo, Noreen Wauford, Eileen Hu, and Sebastian Palacios. Thank you in particular to Tasuku Kitada, Brian Teague, Lila Wroblewska, and Velia Siciliano for your mentorship. An additional group, who I didn't directly work with, never hesitated to offer assistance as well as scientific and non-scientific discussion: Allen Tseng, Nika Shakiba, Ely Porter, Lin-Ya Huang, Michelle Chang, Maria Foley, Leonid Gaidukov, Nicholas DeLateur, Jake Becraft, Deepak Mishra, Jesse Tordoff, and Casper Enghuus.

None of this would be possible without the untiring work of those who kept the lab running: Steve Firsing, Cammie Haase-Pettingell, Olga Parkin, Kalpana Jagtap, Kristine Marzilli, as well as lab aides and technicians who have maintained supplies and made my day-to-day lab life a million times easier: John Scarpa, Roger Guevara, Selamawit Mamo, Jonathan Lyles and Aishwarya Jagtap.

My experience would not have been the same without "the burrow" crew- Jeremy Gam, Jake Becraft, Ross Jones, and Noreen Wauford- as well as burrow-frequenter Nicholas DeLateur. Thank you all for making the lab enjoyable. Thank you in particular Jeremy and Ross. I hope we have labs of our own someday and still call each other up to say "*hey look at this cool data*" or discuss how to manage our underlings. I'm so thankful I was able to grow as a scientist alongside you two.

A special thank you to the BE community especially the administrative staff and the incoming class of 2013. I still find it mind-blowing that I had the chance to belong

to a cohort that is so brilliant yet also so thoughtful. I'm so proud of you all and can't wait to see what you do; thank you for your support and friendship. An additional shout-out to the Leaders for Global Operations (LGO) class of 2019. I'm so lucky I got you all as another group to call family in the last two years of my PhD.

To those friends who knew me before graduate school and supported me unwaveringly: Leanne DeLosh, Sarah Robinson, Marianna Sofman, Hannah Gonzalez, Jen Sung, Price Kinney, Abigail Motley, Sarah Musial, Sangita Sharma, Sarah Carlson, and Jason Paul. A special thank you to Nil Gural for your fierce support of me scientifically and emotionally during my PhD. I'm so lucky to have you all in my life.

I would be nowhere without my family. Thank you to my second families: to Jim and Anita for your love and advice, to Lisa and Matt for being the coolest role models, to Amy and Ron for your guidance and making me laugh, and to Bobby and Alex for your encouragement. To my sisters, Sabrina and Tabitha, you are my inspiration. Thank you for always believing me and for being perfect examples of what following your passion looks like and constant reminders of the beautiful things in life outside of lab (like Jayson!). And I especially want to thank my parents, Andy and Susie, who have provided so much support, given me endless opportunities and love, instilled my sense of curiosity, and kindled my love of learning. Dad, thank you for making me believe I can build whatever I set my mind to and, Mom, your positive outlook on life has been my foundation. I'll never forget two mantras that you both brought me up with: making sure I am a "*contributing member of society*" and to "*be the change I want to see in the world*". I think those two ideas- trying to fulfill the most desperate needs of society while following my heart to get there have been a guiding light throughout my life, pursuing science and during my PhD.

Lastly and most importantly, to my husband, Chris, thank you from the bottom of my heart. You were undoubtedly my rock these last 6 years (and before!) and I can't even begin to express how thankful I am for your endless love and support. There were probably a hundred or so times where I talked jokingly about quitting and a handful where I was quite serious, but you always got me back on track. Thank you for always believing in me, for being so understanding of all of my seemingly endless hours in lab, and for making our home with Brandi and Oliver the most joyful and comforting place to retreat to. You are my best friend, I love you, and I'm so thankful for all of the adventures we've had together. I can't wait to see what the next chapter of our life brings.

Contents

Abstract	3
Acknowledgments	5
Preface	15
1 Introduction	17
1.1 Thesis Statement	20
1.2 Summary of Contributions	21
1.3 Thesis Overview	23
2 Background: Programming Gene and Cell Therapies with Synthetic Biology	25
2.1 Biological modules: building blocks for therapeutic circuits	27
2.1.1 Biological modules that control DNA	28
2.1.2 Biological modules that control RNA	30
2.1.3 Biological modules that control proteins	32
2.2 Genetically encoded therapeutic programs	33
2.2.1 Small molecule-based regulation of gene and cell therapies	34
2.2.2 Sensing biomarkers for localized therapeutic activity	38
2.2.3 Feedback control for augmented homeostasis	41
2.3 Perspective	44
3 One-Pot characterization and optimization of genetic circuits	51
3.1 Evaluating fluorescent protein orthogonality	54
3.2 Benchmarking poly- against co-transfection	57
3.3 New characterization and insights	57
3.4 Optimizing a complete genetic system with poly-transfection	60
3.5 Discussion	65

4	Programmable endonucleolytic RNA scission-induced stability tuning (PERSIST) platform	67
4.1	Developing PERSIST "ON" switches	70
4.2	Combining RNA OFF and ON switches for basic logic	81
4.2.1	Evaluating orthogonality	82
4.2.2	Evaluating composability: cascades and positive feedback	83
4.2.3	Evaluating modularity: All 16 two-input logic functions	85
4.3	Dual functionality allows for complex motifs to be easily engineered	86
4.3.1	Feed-forward loop	89
4.3.2	Single-node positive feedback + repression	90
4.3.3	Construction of an endoRNase-based bistable switch	91
4.4	Discussion and Future Directions	92
5	mRNA sensing in mammalian cells using RNA Strand Displacement	95
5.1	FRET-based mRNA sensing	98
5.1.1	Proof-of-concept detection of endogenously-transcribed mRNA	98
5.1.2	Increasing the strand displacement signal	100
5.1.3	Strand displacement characterization	103
5.2	Endogenous reporters for mRNA sensing	107
5.2.1	Endogenous reporter modeling	110
5.2.2	RBP aptamer interference sensor	112
5.2.3	Hairpin releasing sensor	113
5.2.4	miRNA sponge sensor	113
5.2.5	Ribozyme sensor	116
5.3	Discussion and Future Directions	117
A	Appendix: Methods	119
A.1	Updated Golden Gate DNA Assembly Strategies	119
A.1.1	New backbones and standardization for sub-pL0s and pL2	119
A.1.2	Multimerizing Golden Gate (NxGG) protocol for generating repeated elements	121
A.2	Cell culture protocols	124
A.2.1	Lipid transfections: co-/poly-transfection optimization	124
A.2.2	Flow cytometry	124
A.2.3	Advanced flow cytometry analysis	124
B	Appendix: Theoretical modelling	127
B.1	Theoretical model of the RNA-based switch circuit (replicon vs. pDNA)	127

B.1.1	Computational Model	132
B.2	Models for improving protein sensor performance	143
B.3	Strand Displacement FRET model	150
B.4	Endogenous reporter model	151

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

1-1	Synthetic biology and the therapeutic index	18
1-2	Programming gene and cell therapies with synthetic biology	19
1-3	Programming with the central dogma	20
1-4	Navigating my dissertation	24
2-1	Building blocks for therapeutic programs.	29
2-2	Small-molecule regulation enables control for CAR-T cells	35
2-3	Small molecule regulation for beta-like cell lineage control	36
2-4	Genetically encoded therapeutic programs incorporate cell-specific biomarkers for localized activity.	40
2-5	Gene circuits that use feedback regulation to sense systemic biomarkers and secrete systemically acting effector molecules enable homeostasis.	42
3-1	Overview of poly-transfection and comparison to other methods	53
3-2	Spectral Analysis of fluorescent proteins to determine compatible sets.	55
3-3	Simplified assembly strategy for one-pot-ready plasmids	56
3-4	Benchmarking poly-transfection against co-transfection	58
3-5	Novel CRISPRa characterization enabled by poly-transfection	59
3-6	Novel RNAi characterization enabled by poly-transfection	61
3-7	Optimization of a miRNA classifier circuit using poly-transfection	63
3-8	Evaluation of single-plasmid miRNA classifiers <i>in vitro</i>	64
4-1	PERSIST "ON" switch motif design	71
4-2	Degradation motifs robustly degrade RNA	72
4-3	Triplex RNA stabilizer protects transcript after degradation tag removal	73
4-4	Triplex requirement is Cas-specific	74
4-5	PERSIST OFF switch has improved performance in 5' UTR.	75
4-6	Nuclear localized Cas13 endoRNases have improved performance in ON-switch	77
4-7	9 CRISPR endoRNases function as activators and repressors	78

4-8	PERSIST platform functions in a variety of cell types	78
4-9	Ribozymes repress in OFF-switch but activate transcripts in ON-switch	79
4-10	PERSIST switch can function as a single-transcript miRNA sensor . .	80
4-11	PERSIST integrated response	81
4-12	Orthogonality matrix for endoRNases	82
4-13	EndoRNase repression of an endoRNase	83
4-14	EndoRNase activation of an endoRNase	84
4-15	Single-node positive feedback	85
4-16	Three-stage repression cascade	86
4-17	All 16 two-input logic functions	87
4-18	Dual functionality of Cas PERSIST platform	88
4-19	FFL	89
4-20	Positive feedback + repression motif	90
4-21	PERSIST-based bistable toggle switch	92
5-1	Strand Displacement mechanism	97
5-2	FRET reporter sensing of an engineered transcript	99
5-3	Methods for increasing the signal from strand displacement	101
5-4	Kinetic experiments for transfection optimization	103
5-5	Sensor re-evaluation after optimization	104
5-6	Strand displacement characterization of mRNA sensing inside living cells.	105
5-7	ODE model for strand displacement.	108
5-8	Dynamic range of FRET sensor limits its applicability	109
5-9	Endogenous mRNA reporter	111
5-10	Model of various mRNA reporter mechanisms	112
5-11	RBP aptamer interference sensor	113
5-12	Hairpin releasing sensor	114
5-13	miRNA sponge sensor mechanism	115
5-14	miRNA sponge is not able to relieve repression	115
5-15	Evaluation of miRNA sponges with bulged target sequence	116
5-16	Ribozyme inactivation sensor.	118
A-1	An updated "All Golden Gate" assembly framework	120
A-2	NxGG protocol	122
A-3	Morphological Gating	125
A-4	Bleed-through correction	126

B-1	pDNA and replicon toggle switch results	128
B-2	Replicon system analysis of mutual exclusivity.	139
B-3	Theoretical model diagrams	140
B-4	Mutual Exclusivity (MEx) metric	141
B-5	Comparison of long-term behavior	142
B-6	pDNA system analysis of mutual exclusivity.	142
B-7	Protein sensor model species	143
B-8	Protein sensor model reactions	144
B-9	Protein sensor model set-up	145
B-10	Heatmaps for the identification of parameter interactions	146
B-11	Parameter effects summary	147
B-12	Endogenous reporter model reactions	152
B-13	Endogenous reporter model for various mechanisms	152

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

As is common in scientific fields a full research project is rarely carried out by a single person. This thesis contains both writing and data derived from the efforts of colleagues which allows me to present a complete story of my work. I wish to point out the large contributions made by others and have so included them at the beginning of each chapter.

Additionally, much of the writing and results included in this thesis have been either previously published (Chapters 2, B.1, B.2), are undergoing revision (Chapter 3) or are being prepared for submission (Chapter 4) in peer-reviewed journals. Therefore, the bulk of the content and figures presented herein have been based on these manuscripts, in accordance with the copyright policy of each individual publisher, and represent a joint effort between myself and other listed authors.

My work was supported by the National Science Foundation Graduate Research Fellowship Program grant number 1122374, by National Institutes of Health grant number R01CA207029, and the Siebel Fellowship Foundation.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

“Synthetic Biology is characterized by a focus on constructive approaches to understanding and manipulating biological systems. This is a challenging goal, as biology is complex and the principles that determine its operation are not well elucidated.”

- Cheng et al. Annu. Rev. Biomed. Eng. 2012

A goal of mammalian synthetic biology is to enable more sophisticated control over cellular activity which would allow us to reliably "program" cells with therapeutic behaviors, leading to safer and more effective gene and engineered-cell therapies. While gene and engineered-cell therapies promise new treatment modalities often through a "sense-and-respond" approach (Figure 1-1a) for incurable or difficult-to-treat diseases (and first-generation versions are already used in the clinic), safety concerns may hamper the broader adoption of some of these approaches. For example, overexpression of a therapeutic gene product with a narrow therapeutic window may be toxic, and excessive activation of T cells can be fatal (Figure 1-1b). Synthetic biology provides programming platforms and principles for building "genetic circuits" that integrate multiple signals and perform computation before providing therapeutic function at a precise strength, timing, and location (Figure 1-2). The control and specificity could allow us to create gene and cell therapies that sense and self-correct when they start to leave their narrow therapeutic window (Figure 1-1b, top), or also broaden the therapeutic window of the therapy itself by making them safer (Figure 1-1b,

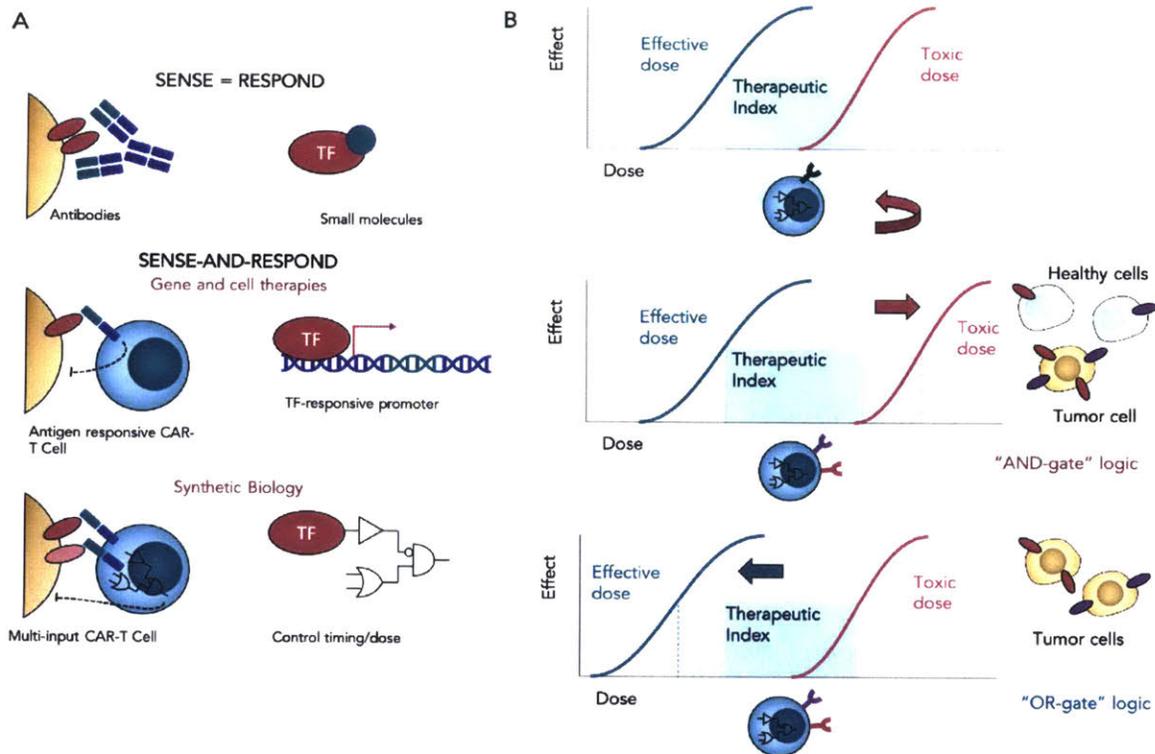


Figure 1-1: Synthetic biology and the therapeutic index: (A) Synthetic biology hopes to develop next generation "sense-and-respond" gene and cell therapies. (B) Sophisticated programming can create therapies that keep themselves from going outside their therapeutic window or expand the range of the therapeutic window itself.

middle) and/or more effective (Figure 1-1b, bottom). Synthetic biology also creates new sensors that convert molecular biomarker input "signals" into signals that can be processed by the genetic circuit, which could provide viable therapeutic strategies in situations where the biomolecular targets have been previously considered "undruggable." And finally, synthetic biology adopts a design-build-test-learn (DBTL) engineering pipeline approach (Figure 1-2b) and has provided marked improvements in DNA synthesis and assembly, gene circuit prototyping, computational modeling for circuit outcome prediction, and high-throughput evaluation and analysis methods.

However some technical hurdles still remain as synthetic biology moves toward therapeutic applications. Many of the gene circuits developed to date are DNA-based and use transcription factors for regulation (Figure 1-3). While transcription factor-based regulation has allowed for some impressive proof-of-concept circuits it presents several challenges for gene and cell therapies. These challenges include epi-

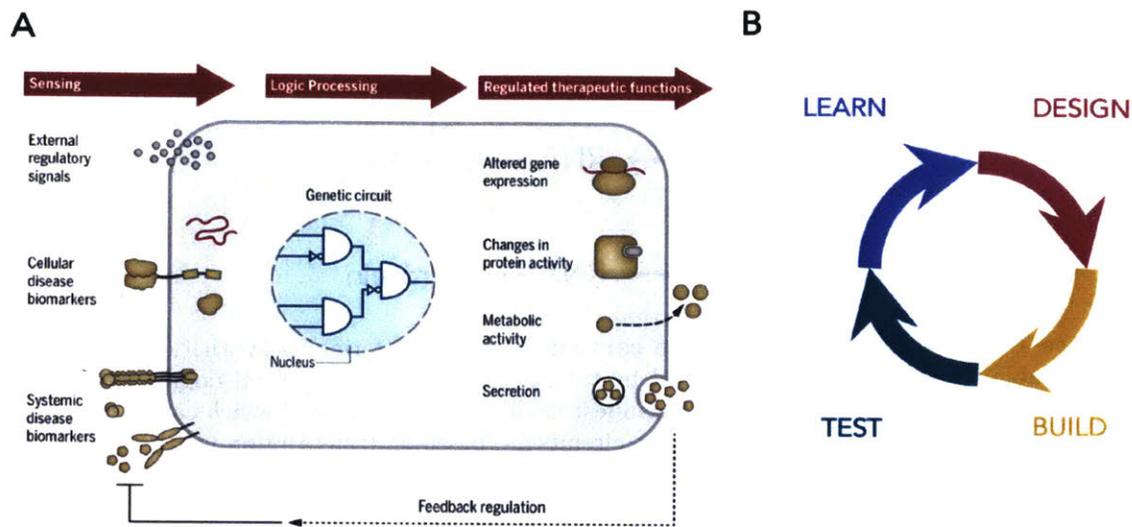


Figure 1-2: Programming gene and cell therapies with synthetic biology: (A) Genetically encoded therapeutic programs can regulate the dosage, localization, or timing of therapeutic function by sensing and processing externally administered signals as well as cell-specific and systemic disease biomarkers. These synthetic gene networks may lead to gene and engineered-cell therapies that are safer and more effective and that can address a broader class of diseases than current approaches. (B) Synthetic biology utilizes an engineering strategy of design-build-test-learn (DBTL) to create gene circuits.

genetic silencing of the circuit in engineered cells, especially due to the transcriptional inactivity caused when using activators and repressors, as well as transcription factor regulation incompatibility with newer RNA-based gene therapies. For this reason, both genetically engineered cells used for cell therapies and biomanufacturing as well as newer RNA-based gene therapies would benefit from post-transcriptional methods of gene regulation, however few such platforms exist.

Also, synthetic biology circuits rely on a limited set of biomarkers, specifically transcription factors, signalling molecules and miRNAs as inputs. These sensors often rely on the activity of the biomolecule itself in order to convert the biomarker into a signal that can be processed by the circuit: for example tissue-specific promoters bind cell-specific transcription factors to regulate expression of a transgene. While all of these molecules have served as suitable markers of cell state, expanding the repertoire would make circuits more robust and some improvements can be made to existing types of sensors themselves. It would be beneficial to detect the concentration

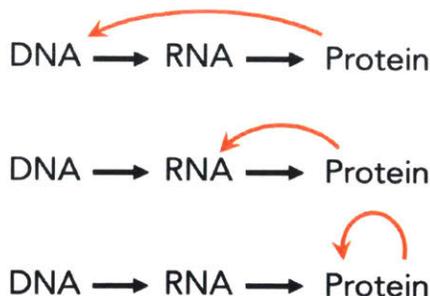


Figure 1-3: Programming with the central dogma: Platforms for synthetic biology must be composable which is often accomplished by taking advantage of the central dogma (e.g. a change at the DNA level due to a protein will manifest itself at the protein level which can in turn regulate again at the DNA level). Many early gene circuits are based on transcription factor regulation, and some recent works have developed protein-protein interaction-based circuits. Here I am developing RNA-level systems that are regulated by proteins or other RNAs.

of "inert" biomolecules such mRNAs and proteins regardless of their activity. Also for miRNAs especially, sensors are suboptimal and often require a double repression cascade to give positive readout of miRNA levels which is difficult to optimize.

Finally, while a longstanding goal of synthetic biology is for the engineering of genetic circuits to be so predictable that a computer program can design circuits based on high-level design specifications, we are still far from that (at least in mammalian systems). In fact most circuits today are still built via trial-and-error through tedious optimization and many rounds of the design-build-test-learn (DBTL) cycle. This is due to several factors including a lack of fully characterized and predictable parts or modules, a lack of easy to use high-information strategies for part characterization and circuit optimization, and, of course, a lack of biological understanding of how our synthetic circuits functions in various contexts.

In my thesis I describe how I address these three mentioned obstacles: a paucity of biomolecular sensors, a lack of robust RNA-level programming platforms, and the inadequacy of the DBTL cycle.

1.1 Thesis Statement

The main goal of this thesis is to improve sensors and processing platforms for mammalian synthetic biology by creating an expanded toolkit of RNA-level systems and

I make the following claim:

Biomolecular sensors and logic processing platforms that function at the RNA level enable the engineering of gene circuits in mammalian cells that are high-performing, predictable, simple to design, and exhibit reliable sustained expression for gene and cell therapies.

This approach involves the development of new experimental and computational methods for interrogating and evaluating circuit performance, the creation and thorough characterization of an entirely new RNA regulation mechanism and associated molecular parts, application of this platform for the detection of biomarkers of interest such as miRNAs and proteins, and finally the evaluation of strand displacement-based reporters for detection of mRNA biomarkers.

1.2 Summary of Contributions

In this dissertation I make the following contributions to the field of mammalian synthetic biology:

- **Poly-transfection method.** In Chapter 3, I contribute to the development of a new transfection method for rapidly characterizing and optimizing genetic constructs. Poly-transfection uses the same reagents as traditional transfection methods but provides substantially more information about the genetic system of interest. Therefore it speeds up the design-build-test cycle of genetic circuit engineering as the process is simplified to a single optimization step. This chapter contains work from the following manuscript currently under revision: Gam, J., DiAndreth, B., Jones, R., Huh, J., Weiss, R. "One-pot transfection method for rapid characterization and optimization of genetic systems" (2019)
- **"PERSIST" platform for RNA regulation.** In Chapter 4, I create a new RNA regulation platform that works via RNA cleavage. In particular I develop an RNA level modular ON-switch that provides a positive readout for RNA

cleavers. I evaluate the ON switch for detecting RNA cleavers such as miRNA, ribozymes and endoRNases. For RNA-level gene programming, I adapt a set of CRISPR-specific endoRNases from other organisms for a non-canonical function in mammalian cells as "activators" and "repressors" in this platform. These regulators allow for the building of complex and sustained logic operations. This chapter contains work from following manuscript currently under preparation: DiAndreth, B., Wauford, N., Palacios, S., Hu, E., Weiss, R. "Programmable RNA circuits using CRISPR endonucleases." (2019)

- **Advancements in mRNA sensing.** In Chapter 5, I make advancements for using the strand displacement mechanism inside living mammalian cells for use in transcript detection. I provide characterization of strand displacement reactions as well as propose and test strand displacement-based mRNA sensing designs in this unpublished work.
- **New cloning schemes.** Throughout the thesis, but described in detail in Appendix A, I develop new circuit assembly frameworks for converting all levels of cloning (i.e. sub-level-0 and level 2) to the efficient golden gate method and provide standardized backbones and overhangs. I also devise a new cloning strategy for generating many repeats of a particular sequence which is still a challenge in DNA assembly. This method is used throughout my thesis, particularly in Chapters 4 and 5.
- **Advanced flow cytometry analysis pipelines.** To facilitate the use of poly-transfection in Chapter 3 and for general analysis of flow cytometry data that has been generated from measured circuit behavior, I provide a MATLAB flow cytometry data analysis pipeline and a repository of custom scripts. Unlike traditional flow cytometry experiments where labeling results in discrete sub-populations, the results from typical synthetic biology experiments are quite different with response values generated for a wide range of input parameters. My analysis pipeline allows for the quantification of genetic systems, taking into account the variation in transfection efficiency through binning or model

fitting to map input parameters to circuit function and inform subsequent design choices.

1.3 Thesis Overview

In the remainder of this thesis, Chapter 2 provides an overview of therapeutic synthetic biology including both descriptions of basic modules as well as circuit design strategies for therapeutic gene circuits.

In Chapter 3 and associated sections in Appendices A and B, I describe efforts at bringing mammalian synthetic biology up to speed as an engineering discipline (Figure 1-4, pink). Chapter 3 provides a new transfection method, while the Appendices contain analysis frameworks, cloning strategies and computational modeling work all of which make circuit engineering more predictable. These methods are used throughout the rest of the Chapters in this thesis for part characterization and circuit evaluation and optimization.

Chapter 4 describes the development of a new platform for RNA-level programming (Figure 1-4, yellow). This "PERSIST" platform is arguably more straightforward to engineer than DNA-based analogs. I demonstrate its modularity and composability and then move forward to benchmark the system against DNA-based computation and construct other high-performing devices.

Some of Chapter 4 and all of Chapter 5 are dedicated to the development of post-transcriptional sensor technologies (Figure 1-4, blue). I describe efforts towards adapting the PERSIST switch platform as a single-transcript sensor for miRNAs and potentially proteins within Chapter 4. In Chapter 5, I focus specifically on RNA strand displacement as a method for mRNA sensing. Here I characterize strand displacement inside cells using a FRET-based reporter and also outline my designs for creating endogenously transcribed strand displacement-based mRNA sensors.

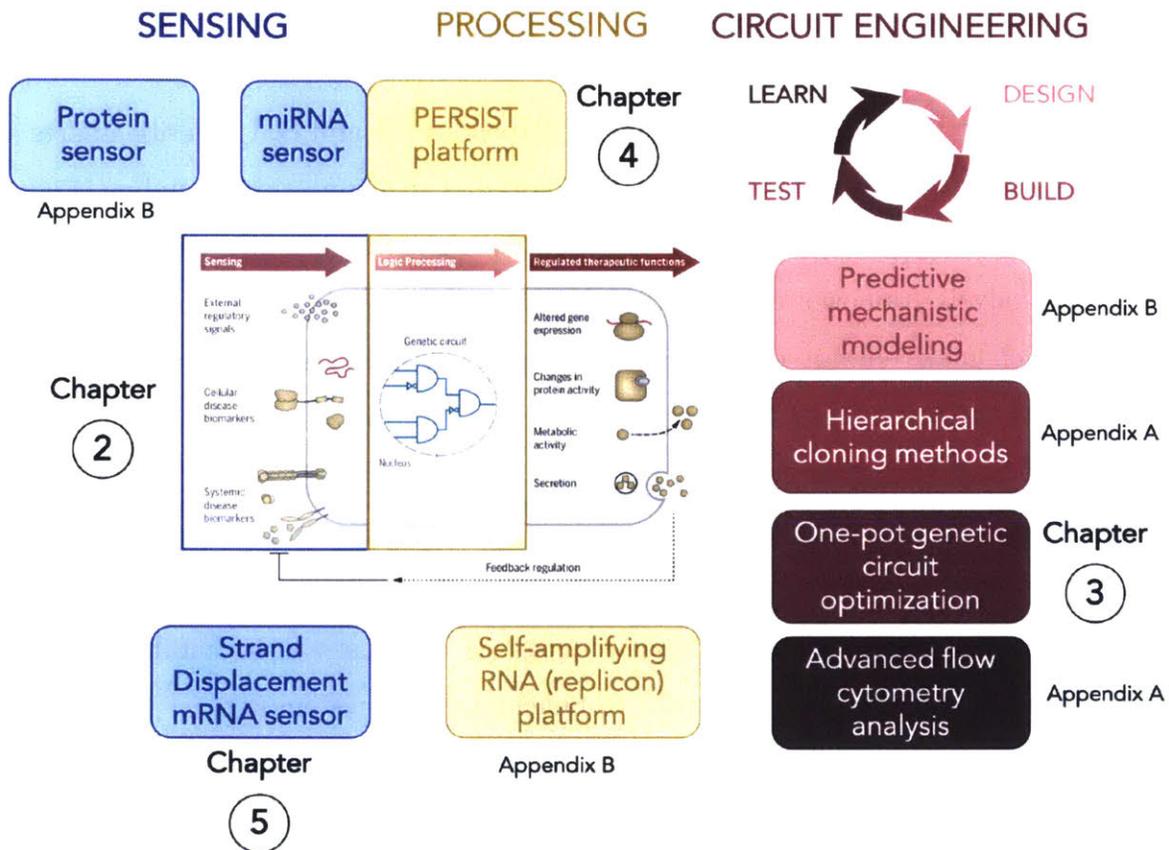


Figure 1-4: Navigating my dissertation: My thesis is arranged into three main developments: sensors (blue, Chapter 5), processing platforms (yellow, Chapter 4), and improvements for the DBTL cycle (pink, Chapter 3). Other projects can also be organized into this schema and are presented in the Appendix. Chapter 2 provides an overview of therapeutic mammalian synthetic biology.

Chapter 2

Background: Programming Gene and Cell Therapies with Synthetic Biology

This chapter contains work from the following published manuscript and therefore represents a collaboration between myself, Tasuku Kitada, Brian Teague, and Ron Weiss.

Kitada, T., DiAndreth, B., Teague, B., Weiss, R. (2018) "Programming Gene and Cell therapies with Synthetic Biology" *Science*, 359 (6376).
science.sciencemag.org/content/359/6376/eaad1067

Gene and engineered cell therapies use nucleic acids to repair or augment a cell's genetic "program" in order to change its behavior in a therapeutically useful manner. For example, transducing the hematopoietic stem cells of β -thalassemia patients with a functional β -globin locus may cure their disease [1], while genetically modifying a cancer patient's T cells *ex vivo* with a chimeric antigen receptor (CAR) may allow them to destroy tumor cells when transplanted back *in vivo* [2]. These approaches represent a new wave of rational therapeutic design and open exciting new avenues to treat or even cure previously intractable diseases [3].

However, while gene and engineered cell therapies are starting to demonstrate promising clinical results, an important limitation of many current approaches is that they provide little control over the strength, timing, or cellular context of the therapeutic effect. This lack of control may hamper broader adoption of these approaches: for example, existing gene therapies usually rely on overexpression of a therapeutic

behavior, leading to new or improved therapies for a host of diseases.

2.1 Biological modules: building blocks for therapeutic circuits

One of the most important principles of engineering, modularity, allows engineers to design complex systems by combining simpler functional units with defined inputs and outputs. These simpler units, or modules, can be designed, tested, and characterized independently before being integrated together. Biological systems can similarly be thought of as a hierarchical connection of many simpler units [10], the simplest of which are molecular interactions. For example, transcription can be thought of as a module with two "inputs" (a DNA molecule containing a promoter and a transcription factor) and one "output" (an RNA molecule). These molecular interfaces allow bioengineers to create synthetic modules that interact with endogenous cellular processes, and support the creation of more complex synthetic systems via the composition of these modules.

Like all abstractions, this definition of biological modularity ignores many important details for the sake of conceptual simplicity. By focusing initially on the modules' inputs and outputs, this abstraction can make designing complex "biological programs" more tractable: the overall desired behavior of a system is expressed as a composite set of logic operations (Figure 2-1A), which in turn are decomposed into modules that encode appropriate logical relationships and whose inputs and outputs can be properly connected. For instance, imagine a "second-generation" monogenic gene therapy that allows a clinician to control the strength and timing of transgene expression via administration of an appropriate small-molecule drug. The overall behavior of this therapy can be expressed by a simple Boolean AND-gate, where the therapy is only "ON" if both the DNA and the small molecule are present. We begin by decomposing the overall desired system behavior, "DNA AND small molecule drug \rightarrow protein," into a set of molecular modules. The required "AND" logic can

be implemented in many ways (Figure 2-1C): for example, the program's transcription module could be one that requires a small-molecule drug activated transcription factor for transgene transcription [11]. Alternatively, we could insert a drug-sensitive ribozyme in the transgene's 5' untranslated region, which regulates the transcript's degradation rate [12]. Choosing an appropriate strategy requires deeper consideration of the clinical requirements, biological dynamics and cellular context of the eventual therapeutic circuit.

The effective design of gene circuits requires a broad understanding of the available modules, and thus we begin this review with a survey of synthetic modules particularly useful for gene and engineered cell therapies. The survey groups modules together based on their output, classifying them as directly affecting the abundance or activity of DNA, RNA or proteins (Figure 2-1C). This structure reflects our focus on therapeutic applications, as it is often useful to map a module's functional effect to the molecular basis of a disease. For example, patients with spinal muscular atrophy (SMA) have decreased levels of correctly-spliced SMN mRNA and therefore modules that affect mRNA could be utilized. Amounts of the functional full length SMN splice isoform could be restored by increasing pre-mRNA production with a synthetic transcription factor, or by targeting a splicing silencer region with a splice-switching oligonucleotide [13, 14]. For each category of modules, we discuss several recent examples, emphasizing both their utility in engineering synthetic gene circuits and their applicability in a clinical context.

2.1.1 Biological modules that control DNA

Biological modules that act on DNA (Figure 2-1C) have the potential to be powerful therapies because editing a cell's genome can cause permanent changes in its phenotype. For example, targeted DNA cleavage and subsequent repair of gene regulatory elements may one day treat β -hemoglobinopathies [15, 16]. While gene editing tools based on zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) have been used in several clinical trials (17, 18), recent advances in clustered regularly interspersed short palindromic repeat (CRISPR) systems are gen-

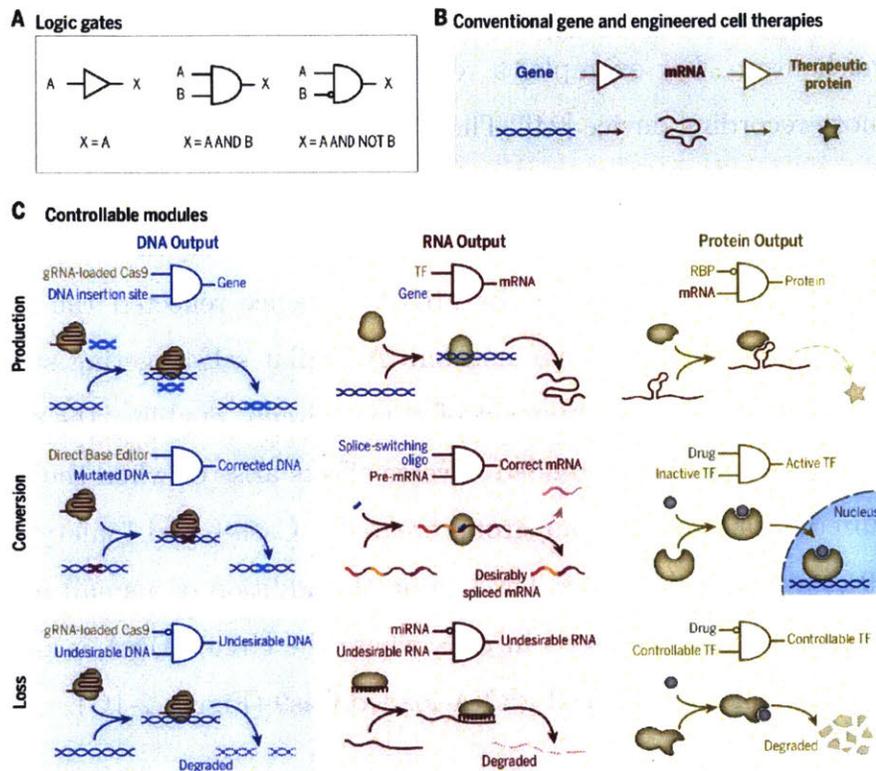


Figure 2-1: Building blocks for therapeutic programs.: (A) Logic gates with inputs (A and/or B) and outputs (X) can be used to represent molecular processes and reactions. (B) Conventional gene and cell therapies require just one exogenous molecular input and lack precise control over the output. Such modules function as buffer gates (i.e., control devices whose output levels correspond to their input levels), because the RNA output will be produced in any cell that the DNA input is delivered to and the therapeutic protein will be translated correspondingly. (C) Engineerable modules can regulate the production, conversion, or loss of specific DNA (blue), RNA (red), or protein (yellow) species by using more than one molecular input. TF, transcription factor; RBP, RNA-binding protein.

erating excitement because their targets are specified by a guide RNA (gRNA) using simple base-pairing rules instead of laborious protein engineering. CRISPR-enabled gene editing is already generating encouraging results in animal models [17–19]; for example, Tabebordbar et al. used a Cas nuclease from *Staphylococcus aureus*, SaCas9, to excise a mutated intron in the mdx mouse model of Duchenne muscular dystrophy. The CRISPR system has also been extended to enable direct base editing [20,21] and

targeted DNA demethylation [22, 23].

Gene editing tools including Cas nucleases can also be targeted to exogenous DNA that encodes the synthetic gene circuit itself as part of the biological program or as a safety mechanism. For example, a recent effort used a Cas9 system as a DNA-based memory recording device [24]. The Cas9 transgene was placed under control of an NF- κ B-responsive promoter and the nuclease was targeted to the DNA sequence encoding the gRNA. When cells containing this memory device were implanted *in vivo*, the number of mutations in the gRNA sequence reflected the intensity and duration of NF- κ B-mediated inflammation. A similar self-targeting strategy might be used as a "self-destruct" component of a gene circuit, enabling the destruction of DNA-encoded therapies if unintended adverse effects arise or when the therapy is no longer required. For example, our group created a Cas9-based "safety switch" [25], where Cas9 cleaves circuit-related DNA upon the addition of a small molecule. The circuit is therefore only expressed in cells where the circuit DNA is present "AND NOT" the small-molecule induced gRNA-loaded Cas9 (Figure 2-1C).

2.1.2 Biological modules that control RNA

Many different cellular processes regulate the production, stability, conformation, splicing, and translation of RNA [26]. The broad range of processes that involve RNA make it a compelling target for engineering efforts, while dysfunctions in these processes make synthetic systems that interact with RNA of significant therapeutic interest. For example, a synthetic splice-switching oligonucleotide (Figure 2-1C) was approved by the U.S. FDA for the treatment of SMA [14], while a synthetic small interfering RNA (siRNA) which targets transthyretin (TTR) mRNAs for degradation demonstrated strong clinical benefits in a trial for patients with TTR amyloidosis [27]. These successes suggest that more sophisticated synthetic biological modules that control the abundance and activity of RNA may have broad therapeutic potential.

One way to control the abundance of RNA is by regulating its transcription from DNA. Synthetic transcription modules (Figure 2-1C) in mammalian systems have typically been built by fusing DNA binding domains with activation and repression

domains [28]. For example, Zhang et al. created synthetic transcriptional activators by attaching the potent synthetic activation domain, VP64, to TALE DNA binding domains designed to target sequences upstream of endogenous genes [29]. Customizing the DNA-binding specificity of synthetic transcription factors has become even easier with the development of catalytically "dead" dCas9, which still binds DNA but does not cleave it. For instance, dCas9-based transcriptional activation systems designed to target the 5' LTR of dormant HIV proviral DNA have been used to induce HIV in cell culture models of latency [30].

Additionally, the abundance, availability, and translational or catalytic activity of RNA can be modified directly via interactions with other biomolecules. For example, Chen et al. created a synthetic RNA device that deactivates an adjacent ribozyme upon binding theophylline [12]. When inserted into the 3' UTR of a transcript encoding IL-15 in a mouse cytotoxic T cell line (CTLL-2), the theophylline switch implements a logic AND module: only in the presence of both the RNA transcript and theophylline is IL-15 expressed. This enabled control over IL-15-mediated proliferation of these T cells in mice to improve clonal expansion following adoptive transfer. More recently, orthologs of the RNA-guided RNA-targeting CRISPR-Cas effector Cas13 have been used in mammalian cells to both knock down and directly edit endogenous mRNA transcripts with high specificity [31,32]. Such programmable RNA-binding proteins enable precise post-transcriptional regulation of endogenous RNAs, which has exciting therapeutic implications.

Finally, just as siRNAs can be used to target endogenous mRNAs for degradation (Figure 2-1C), synthetic biological modules can regulate RNA abundance by interfacing with endogenous host RNA degradation machinery. For example, engineered mRNAs containing miRNA target sites in their 3' UTRs can be used as miRNA sensors because their abundance (and the abundance of the translated protein) is inversely related to the amount of cognate miRNA in the cell. This approach was used to control the replication of an engineered HSV-1 oncolytic virus by inserting miR-124 target sites into the 3' UTR of an essential viral early gene [33]. Because miR-124 is expressed in neurons but not glioblastoma cells, viral replication occurred

in cancerous cells but not in healthy neuronal tissue.

2.1.3 Biological modules that control proteins

Proteins are a biochemically diverse molecular species that transduce information, catalyze synthesis and conversion of other biomolecules, and serve as structural components inside and outside the cell. Due to this functional diversity, clinicians have been able to use protein "biologics" (i.e. complex drugs manufactured in and isolated from living cells) to treat diseases including autoimmune, metabolic, and cardiovascular disorders as well as cancer [34]. Synthetic biology promises to improve upon these therapies by providing precise control over the abundance, localization and activity of therapeutic proteins, making them safer and more effective.

Because most gene circuits express proteins from RNA, any of the strategies discussed above for controlling RNA abundance can also be used to modulate protein levels. Additionally, an mRNA's translation can be influenced by its interactions with other proteins and small molecules. For example, the archaeal ribosomal protein L7Ae binds the box C/D kink-turn (K-turn) and related K-loop RNA motifs and this binding has been shown to strongly inhibit translation of the RNA [35,36]. This module therefore produces a protein output when the logic function "mRNA AND NOT L7Ae" is true (Figure 2-1C). Endogenous RNAs can be modulated in this way as well, via "programmable" RNA-binding proteins (RBPs) such as the Pumilio/fem-3 mRNA binding factors (PUF) [37], pentatricopeptide repeat (PPR) proteins [38] and, more recently, RNA-targeting Cas effector proteins [31,39–44]. Together, these new tools are rapidly expanding the list of RNAs whose translation can be controlled by engineered systems.

Finally, a protein's abundance and activity can be modulated by fusing it to domains that respond to small molecules (Figure 2-1C). For example, a number of modular degradation domains are stabilized by small molecule ligands, allowing for direct external control of protein levels [45–47]. Banaszynski et al. fused one of these domains to the cytokine TNF- α , then used the small molecule ligand Shield-1 to control the strength and timing of the cytokine's expression in mice. The fusion

protein was expressed from a strain of vaccinia virus that preferentially replicates in tumor cells. When Shield-1 was administered three days after viral delivery, TNF- α expression was localized primarily to the tumor cells, demonstrating control over both the localization and timing of an otherwise-toxic gene product [48]. Another recent example is the synthetic thyroid hormone homeostasis system developed by Saxena et al. [49], which used hormone-binding domain of human thyroid receptor alpha fused to a DNA binding domain to activate a reporter transgene in response to thyroid hormone. When they replaced the reporter transgene with a thyroid-stimulating hormone receptor antagonist, the construct was able to restore thyroid hormone homeostasis in a mouse model of Graves' disease. This last example of a prototype therapeutic gene circuit uses feedback to modulate the level of its therapeutic output: the output of the circuit decreases thyroid activity, which reduces thyroid hormone levels that are in turn the input to the circuit. This negative feedback is one strategy that therapeutic gene circuits can use to control the strength, timing and context of their therapeutic output. Such control strategies, and gene circuits that implement them, are the subject of the next section of this review.

2.2 Genetically encoded therapeutic programs

While recent years have witnessed an increase in the number of successful gene and engineered cell therapy trials for various diseases, more precisely regulating the dosage, localization, or timing of a treatment's therapeutic activity may lead to improved safety and efficacy profiles for existing treatments as well as enable new modes of therapeutic intervention. For instance, implanted cells engineered with "prosthetic" gene circuits may one day treat autoimmune disorders by sensing systemic disease-associated biomarkers and secreting immune-modulating proteins in response (52). In this section, we describe three synthetic biology strategies to more precisely control gene and engineered cell therapies. In the first strategy, a synthetic gene circuit's activity is externally modulated via small molecules, affording a clinician precise control over the intensity and timing of therapeutic functions. In the second strategy, gene

circuits sense intracellular and extracellular biomarkers to spatially restrict therapeutic activities to diseased cells and tissues. Finally, in the third strategy, gene circuits use feedback control loops to adaptively modulate the activity levels of therapies to treat diseases caused by disrupted homeostasis. For each strategy, we examine several recently reported gene circuits, exploring how therapeutic requirements drive the design of synthetic biology solutions. Together, these examples highlight how synthetic biology could pave the way to safer and more effective gene and engineered cell therapies.

2.2.1 Small molecule-based regulation of gene and cell therapies

Successful clinical translation of gene and engineered cell therapies would be facilitated by the ability of a clinician to control the activity of a therapy once it has been administered to a patient. In some settings, external control can decrease the likelihood that excessive activity of an engineered cell or gene therapy results in harm to the patient. For example, serious adverse events have been recorded in a number of recent CAR T cell cancer immunotherapy trials [2]. CAR T cell therapy is a potent new cancer treatment where a patient's T cells are harvested, genetically engineered with a CAR against a tumor antigen, expanded and re-infused into the patient. Unfortunately, in some patients, through a not yet fully characterized process linked to excessive activation of the infused cells, CAR T cell therapy can cause severe neurotoxicity or cytokine release syndromes that have proved fatal [2].

One possible way to address such problems is to engineer gene circuits whose activities are regulated by the administration of small molecules. While gene and engineered cell therapies may persist in the body for a relatively long period of time, small molecules typically have short half-lives *in vivo* and thus can be used to precisely control the activities of gene and engineered cell therapies. For example, to enable external control over cytotoxic T cell function, Wu et al. created a CAR whose activation was dependent on a rapamycin analog (rapalog) [50]. CAR proteins are typically

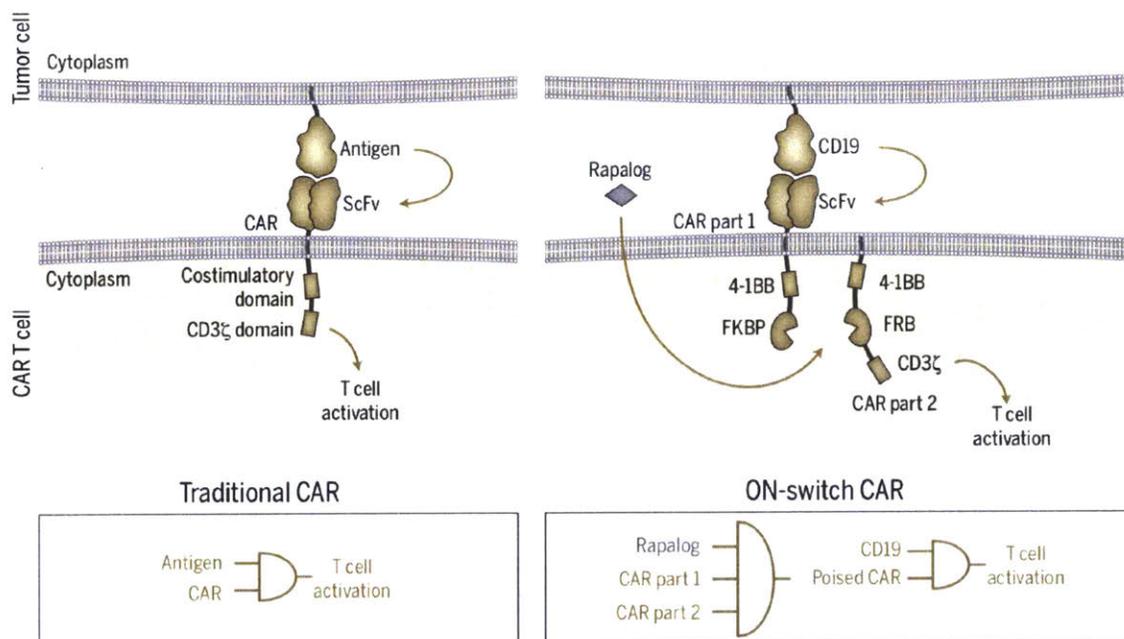


Figure 2-2: Small-molecule regulation enables control for CAR-T cells: Traditional CARs are activated when the T cell encounters a target antigen. ON-switch CARs respond to antigens only when a small molecule, such as rapalog, is administered.

composed of a single-chain variable fragment (scFv) fused to activating domains of the T cell receptor CD3 ζ intracellular domain and co-stimulatory domains such as CD28 or 4-1BB (Figure 2-2). Wu et al. modified the original CAR to develop an "ON-switch" CAR system consisting of two modular transmembrane proteins: one containing the extracellular scFv domain, a 4-1BB co-stimulation domain, and an FKBP domain; and the other composed of the CD3 ζ T cell activator, a second 4-1BB domain, and an FRB domain. The FKBP and FRB domains interact to create a complete receptor only when rapalog is present (Figure 2-2). Small molecule-dependent activation of the ON-switch CAR T cell was demonstrated *in vitro* using a cell killing assay, as well as by tumor clearance in a xenograft mouse model of CD19-positive lymphoma, establishing the proof of concept of a titratable CAR T cell system.

The key to this circuit's function is the careful design of the split CAR. In particular, the CAR's activity is controlled by rapalog post-translationally. This allows for a timely response to the small molecule and might allow the CAR T cells to be shut down rapidly in patients experiencing cytokine release syndrome or neurotoxicity.

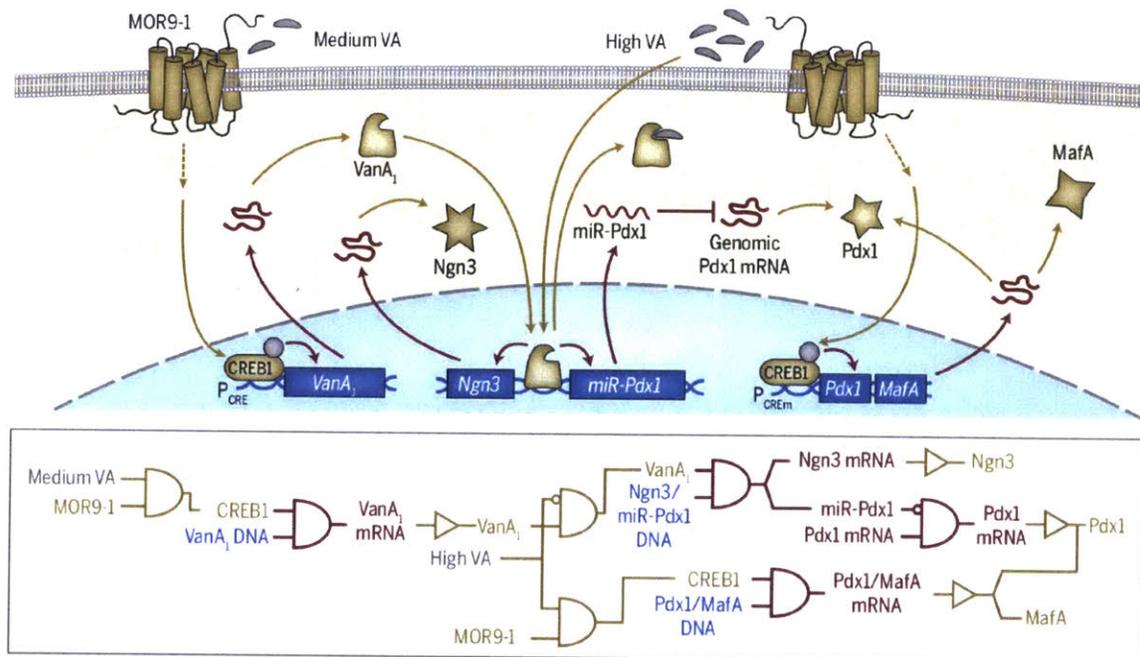


Figure 2-3: Small molecule regulation for beta-like cell lineage control: The pancreatic progenitor-to- β -like cell differentiation circuit is controlled by VA. Increasing levels of VA establish three different gene-expression profiles for the transcription factors PDX1, NGN3, and MAFA to drive differentiation. The final concentration of PDX1 is a summation of translation from two mRNA sources akin to a wired-OR operation in electronic logic circuits. Dashed arrows indicate multiple steps. The same drawing conventions are used as in Figure 2-1

This strategy is advantageous over a design in which a small molecule controls the transcription of a CAR: in such a scenario, full activation of the CAR T cell would be delayed by the transcription and translation processes necessary to produce the receptor, and shutoff of T cell activity would be slow because of the time required for the CAR protein to be degraded.

While small molecule control over gene circuit activity has obvious safety benefits, the same strategy can also be used more broadly to control therapeutic gene circuit behavior to facilitate new modes of therapeutic interventions. For example, Saxena et al. developed a synthetic gene network which uses increasing concentrations of vanillic acid (VA) to differentiate pancreatic progenitor cells into insulin-producing β -like cells [51] (Figure 2-3). Saxena et al. use discrete concentrations (zero, moderate, and high) of VA to sequentially establish three distinct patterns of gene expression, resulting in an OFF-ON-OFF pattern for Ngn1; an ON-OFF-ON pattern for Pdx1;

and an OFF-OFF-ON pattern for MafA. Two different VA sensors enable the circuit's concentration-dependent response: MOR9-1, which is activated by VA, and Van A1, which is inhibited by high levels of VA. Plasmids encoding the circuit were transfected into pancreatic progenitor cells, which were differentiated from iPS cells using growth factors and small molecule inducers. In the absence of VA, circuit-containing pancreatic progenitor cells express endogenous Pdx1, but no genes from the circuit (Pdx1:ON, Ngn3: OFF, MafA: OFF). At moderate levels of VA, the odorant receptor MOR9-1 is activated and this in turn generates moderate levels of activated endogenous CREB1. Activated CREB1 binds the highly sensitive PCRE promoter, inducing expression of the synthetic transcription factor VanA1, which in turn activates transcription of both Ngn3 and a synthetic miRNA against endogenous Pdx1 mRNA (Pdx1:OFF, Ngn3: ON, MafA: OFF). This drives differentiation of the pancreatic progenitor cells into endocrine progenitor cells. Finally, at high levels of VA, and therefore high levels of activated CREB1, the less sensitive PCREm promoter is activated and induces Pdx1 and Maf1, while VanA1 is inhibited by the high concentration of VA. Inhibition of VanA1 stops induction of Ngn3 and Pdx1 miRNA (Pdx1:ON, Ngn3: OFF, MafA: ON), producing β -like cells. By changing the concentration of one exogenous molecule, three different gene expression signatures are generated, guiding the stepwise differentiation of pancreatic progenitor cells into β -like cells.

An important feature of this lineage-control network is the tight integration between the synthetic gene network and endogenous machinery for transmitting information and effecting changes to cell state. The VA-activated G-protein coupled receptor MOR9-1 is expressed from a transgene but it transmits information to the rest of the synthetic gene circuit via an endogenous G protein, adenylyl cyclase, kinase, and transcription factor. This takes advantage of the signal amplification properties of the endogenous signaling network [52]. Additionally, the three circuit-encoded effector genes that drive cell differentiation, Ngn3, Pdx1 and MafA, also drive transcription from their endogenous loci. These feedback loops, initially activated by the synthetic gene network, serve as signal amplifiers and help the combined gene network achieve each intermediate state necessary to drive the pancreatic progenitor

cell's differentiation.

While the synthetic gene circuit that Saxena et al. describe operates *in vitro*, circuits that guide multi-stage (trans)differentiation based on the concentration of a single small molecule may be useful for *in vivo* therapeutic applications as well. For instance, there is continued interest in using cell-based therapies to repair damaged or diseased tissues *in situ*, but engraftment of immature cells into existing structures remains a challenge [53]. Such therapies generally involve creation of patient-specific pluripotent stem cells, then implanting them into damaged tissue. Engineering these cells with a gene circuit that directs an engineered stem cell's differentiation *in vivo* could improve the likelihood that it engrafts successfully. This strategy would be particularly powerful when combined with *ex vivo* gene editing for monogenic disorders: for example, Duchenne muscular dystrophy could potentially be treated by 1) generating iPSCs from a patient, 2) correcting the mutated DMD gene, then 3) implanting them into the patient's muscles and directing their differentiation into mature myofibers. For transdifferentiation, an analogous approach might involve delivering a similar multi-stage gene circuit to damaged tissues *in vivo*, which could guide the direct and efficient conversion of one cell type to another. Activating such a circuit in specific cell types is the subject of the next section.

2.2.2 Sensing biomarkers for localized therapeutic activity

Mechanisms to regulate gene circuit activity based on endogenous cellular biomarkers can enable gene and engineered cellular therapies to be more context-specific, activating only in the proper cellular environment. This additional specificity can increase potency and reduce off-target activation, making engineered gene and engineered cell therapies safer and more effective. CAR T cell-based cancer immunotherapy, discussed above, is a particularly ripe target for biomarker-based gene and engineered cell therapies. Since an antigen targeted by a CAR is rarely expressed exclusively on tumor cells, T cell activation is sometimes observed in tissues other than the tumor, leading to adverse events in clinical trials. For example, leukemia patients who were administered autologous T cells engineered with CARs against CD19 experienced

cancer remission but also long-term depletion of normal CD19+ B cells [54]. The engineered effector cells in these trials are potently activated by the tumor antigen, but if their activation could be made more specific, they may serve as the basis for safer therapeutics.

One approach to improving the specificity of engineered CAR T cells is to make them recognize multiple antigens instead of just one. Roybal et al. did so by engineering a CAR T system that could recognize two independent antigens [55], based on a synthetic Notch juxtacrine sensor developed by Morsut et al. [56]. The native Notch receptor recognizes a cognate ligand and releases an intracellular domain that activates endogenous gene expression. Morsut et al. replaced both the extracellular ligand-binding domain and the intracellular transactivation domain, creating a synthetic Notch receptor (a "synNotch") whose signaling is orthogonal to native cellular machinery. Roybal et al. then used the synNotch receptor platform to engineer CAR T cells with improved specificity [55]. Their system is composed of several modules: (1) a synNotch receptor, which binds to a tumor antigen and releases a synthetic transactivator; (2) the DNA encoding a CAR, which is activated by the synNotch transactivator; and (3) the CAR itself, which binds to a second tumor antigen and activates the T cell. This results in a very specific AND-gate: only T cells presented with both the synNotch ligand and the CAR ligand are activated (Figure 2-4A). As a proof-of-concept, Roybal et al. built a synNotch sensor for GFP and a CAR specific to CD19. When they implanted a CD19/GFP-expressing xenograft into mice and injected their engineered T cells, the tumor was cleared. Importantly, the AND-gate T cells did not elicit a response against tumor cells that only express GFP or CD19, demonstrating the specificity of the engineered T cells.

The dual-antigen CAR T cell system is a promising platform for personalized medicine because both the synNotch receptor and the CAR can be engineered with virtually any ligand binding domain. In the future, it may be possible to engineer T cells that precisely target a patient's tumor by using synNotch receptors and CARs or TCRs that recognize unique cell surface markers or MHC epitopes expressed on that patient's tumor cells [57]. There are also opportunities to add other effectors to the

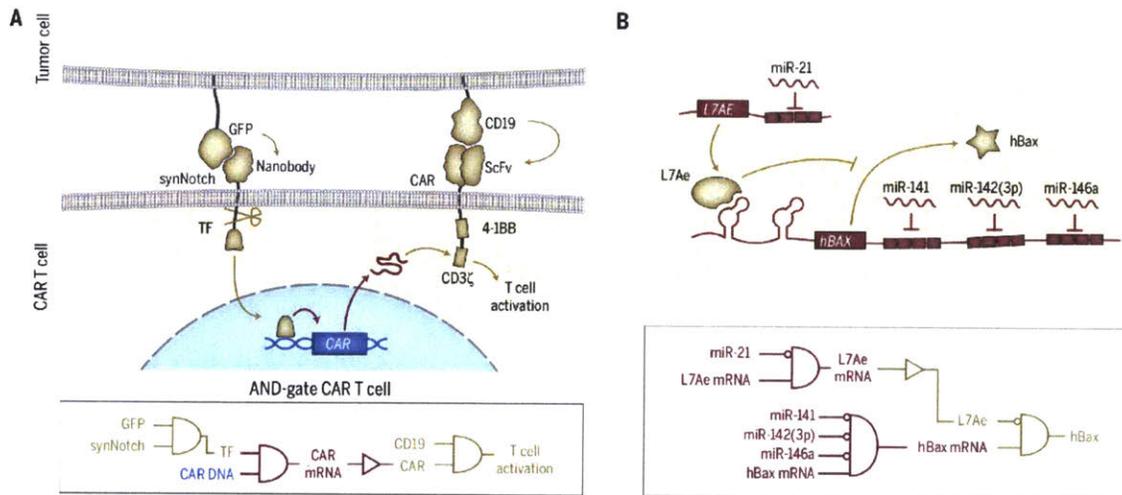


Figure 2-4: Genetically encoded therapeutic programs incorporate cell-specific biomarkers for localized activity. : (A) In an AND-gate CAR T cell, the activation of a synNotch receptor by a first antigen induces the expression of a CAR, which in turn is activated by a second antigen to ultimately activate the T cell. (B) RNA-encoded miRNA-classifier circuit selectively kills cancer cells characterized by high levels of miR-21 and low levels of miR-141, miR-142(3p), and miR-146a. The same drawing conventions are used as in Figure 2-1

output of the circuit: because synNotch can activate any exogenous gene, it would be straightforward to expand the circuit with other modules that (for example) modulate the properties of T cells. Indeed, in a recent publication, Roybal et al. engineered synNotch T cells that express the T helper type 1 (Th1) specific transcription factor Tbet upon recognition of the CD19 antigen [58]. These synNotch T cells differentiated into interferon gamma ($\text{IFN}\gamma$) secreting Th1 cells when co-cultured with leukemia cells ectopically expressing CD19 but not when co-cultured with control leukemia cells.

There are biomarkers besides cell surface antigens that can be used to distinguish cancer cells from healthy tissues. Intracellular biomarkers such as miRNAs can also be detected by gene circuits, providing specificity to broadly cytotoxic anti-cancer mechanisms that may otherwise have significant off-tumor activity. In this vein, in collaborations with the laboratories of Benenson, Saito, and Xie, we have created several "cancer classifier" circuits that use endogenous miRNA expression signatures to distinguish between HeLa cells and healthy cells and activate apoptosis selectively in "malignant" HeLa cells while sparing surrounding cells. The original

circuit was built using plasmid DNA by Benenson and our group [8] and was subsequently modified by Lapique and Benenson to reduce leaky output expression by introducing a recombinase-mediated delay in expression of the toxic load [59]. In a related project, with Xie’s group, we demonstrated HeLa/HEK cell classification using a cross-repressed TALE repressor circuit [60]. This circuit architecture decreased leaky expression and improved the signal-to-noise ratio of cell type classification. More recently, we implemented another version of the classifier based on the original circuit design [8] except using only post-transcriptional regulation (Figure 2-4B) [61]. This allowed us to encode the circuit entirely in synthetic mRNA, a safer therapeutic modality compared to DNA (65). The RNA-encoded circuit consists of mRNA for the RNA-binding protein L7Ae with target sites for miR-21 in its 3’ untranslated region, and a second mRNA encoding the pro-apoptotic hBax protein regulated by an upstream K-turn motif and followed by target sites for miR-141, miR-142(3p) and miR-146a. L7Ae binds the K-turn motif and represses expression of hBax. Thus, hBax is expressed only in cells with high levels of miR-21 and low levels of miR-141, miR-142(3p) and miR-146a, a signature specific for HeLa cells. When this mRNA-encoded circuit was transfected into a co-culture of HeLa and HEK cells *in vitro*, the circuit induced apoptosis specifically in HeLa cells.

Several features make miRNA-sensing a particularly attractive strategy for designing cell type classifier circuits. First, current sequencing methods make holistic comparisons of miRNA signatures from different tissues rapid, inexpensive and reliable. Second, miRNA sensing modules are easy to design, function potently, and can be combined in tandem to create more complex logic. One caveat is that miRNA abundance does not always correlate well with miRNA sensor activity [62]. Thus, experimental verification of sensor libraries may be critical for predictable and accurate design of cell type classifier circuits.

2.2.3 Feedback control for augmented homeostasis

Gene circuits that sense and respond to disease biomarkers via feedback loops can regulate therapeutic functions so that they are activated only at the right intensity and

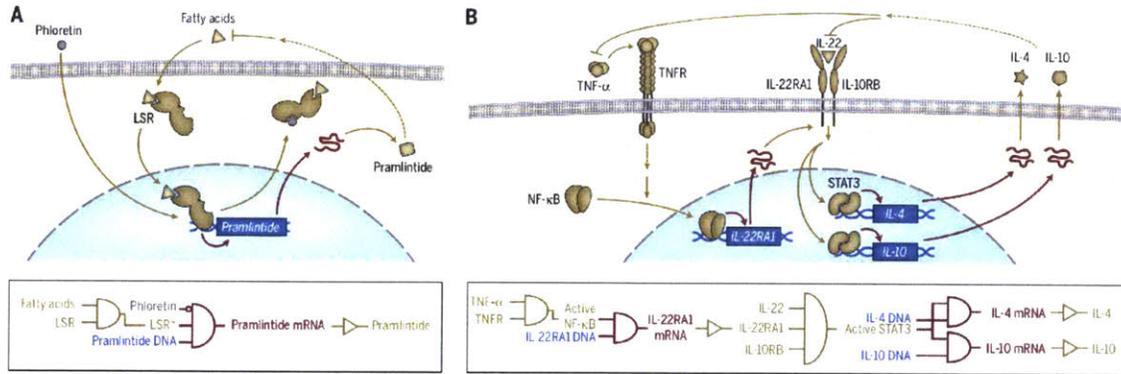


Figure 2-5: Gene circuits that use feedback regulation to sense systemic biomarkers and secrete systemically acting effector molecules enable homeostasis.: (A) A closed-loop circuit to treat obesity responds to fatty acids and produces pramlintide to slow gastric emptying, reduce glucagon, and modulate satiety. (B) A cytokine converter circuit to treat psoriasis responds to inflammatory signals TNF- α and IL-22 and produces anti-psoriatic and anti-inflammatory cytokines, IL-4 and IL-10, respectively. Dashed arrows indicate multiple steps. The same drawing conventions are used as in Figure 2-1

time. Such a feature could be particularly beneficial when systemic modes of interventions are used for treatment of disorders related to disrupted homeostasis [63]. For example, diet-induced obesity may be treated with this approach. While the primary treatment is a change in the patient’s lifestyle and dietary habits, pharmacological and surgical interventions can assist weight loss by suppressing appetite and reducing fat absorption. One such intervention is treatment with pramlintide, an analog of amylin, which aids in blood glucose regulation and promotes satiety. Pramlintide is approved by the U.S. FDA for the treatment of type 1 and type 2 diabetes in patients who use meal time insulin, but it has also been investigated as an adjunct to lifestyle intervention in obesity treatment [64]. Because pramlintide is a peptide, it can be synthesized via a transgene, which makes it an attractive effector module for a gene circuit to treat diet-induced obesity.

In order to enable autonomous dosing of pramlintide, Rössger et al. created a feedback loop which coupled its expression to a sensor for an appropriate metabolite. They built such a sensor, dubbed the lipid-sensing receptor (LSR), by fusing the ligand-binding domain of peroxisome proliferator-activated receptor alpha (PPAR α) to the phloretin-responsive repressor TtgR (Figure 2-5A) [65], which allowed it to bind to synthetic promoter containing the TtgR operator. The PPAR α subunit re-

cruits transcriptional co-activators in the presence of fatty acids, and co-repressors in their absence, expressing the transgene strongly in the ON state but abolishing its expression in the OFF state. The DNA-binding domain TtgR provides another level of control since its binding to the TtgR operator sequence is repressed by phloretin, an apple-derived small molecule found in many cosmetics. The small molecule control could serve as an external way to tune system response or abrogate circuit activity. Thus the LSR forms a logical two-input AND-gate with one inverted input, where the transgene under LSR control is active only when lipids are present but phloretin is absent. In cell culture, this sensor was highly sensitive to exogenous fatty acids, with transgene expression increasing over 100-fold in response to some treatments, such as with rapeseed oil.

Delivery is one of the major obstacles in translating such advances into clinically useful therapeutics. To deliver their prosthetic gene circuit, Rössger et al. engineered the human fibrosarcoma cell line HT-1080 to express LSR and LSR-controlled pramlintide transgene, then encapsulated the cells in alginate-poly-L-lysine beads and injected them intraperitoneally into mice. The alginate encapsulation protects the implanted cells from the host immune response but provides them with access to host metabolism. When obese mice fed a high-fat diet were implanted with the cells, they showed high levels of circulating pramlintide and corresponding decreases in blood fat, food intake and body weight. Prosthetic gene circuits have also been developed to regulate urate [66], blood pressure [67], blood pH [68], thyroid hormone levels [49], enterohepatic signaling [69], blood glucose [9], and insulin levels [70].

In addition to metabolic disorders, feedback control of therapeutic gene circuits is also appropriate for chronic diseases that flare up occasionally but for which prophylactic treatment has safety concerns. One example is psoriasis, a common autoimmune disorder that causes inflamed skin lesions, and whose comorbidities include psoriatic arthritis, Crohn's disease, metabolic syndrome, and cardiovascular disease [71]. The inflammation characteristic of psoriasis is caused by overexpression of cytokines such as tumor necrosis factor alpha (TNF- α) and interleukin 22 (IL-22). Existing therapies include antibodies against TNF- α or Th1 and Th17-related cytokines as well

as various oral or topical treatments, but long-term suppression of the immune system is associated with side-effects such as infection [72]. Recent phase 2 trials of anti-psoriatic and anti-inflammatory cytokines IL-4 and IL-10 have shown efficacy in treating psoriasis [73, 74], but the short half-lives of these compounds means they require almost continuous administration to be efficacious.

To address these challenges, Schukur et al. used a feedback regulation strategy to engineer a gene circuit that senses TNF- α and IL-22 and drives the expression of IL-4 and IL-10 (Figure 2-5B) [75]. The sensing half of the circuit shares a similar "serial sensor" design with the synNotch CAR: the endogenous TNF- α receptor (TNFR) activates expression of an IL-22 receptor IL-22RA via an endogenous NF- κ B signaling cascade; the IL-22RA transgene then senses IL-22 and communicates the signal to the nucleus via an endogenous JAK/STAT cascade. Finally, synthetic STAT3-responsive promoters activate expression of IL-4 and IL-10. The circuit therefore encodes a logical two-input AND-gate: the therapeutic outputs (IL-4 and IL-10) are expressed only in the presence of TNF- α and IL-22. In a cultured human cell assay, the activity was reversible, with production of IL-4 and IL-10 falling after the withdrawal of TNF- α and IL-22, which is a precondition for reacting to changing levels of pathological cytokines during a psoriatic flare-up. HEK-293T cells engineered with the homeostatic circuit were encapsulated in alginate-poly-L-lysine and injected intraperitoneally into mice, where the prosthetic gene circuit successfully controlled inflammation caused by topical application of imiquimod, a common model for psoriatic lesions. The sensors also responded to cytokines in the blood of psoriasis patients, suggesting that it is sensitive enough to detect circulating TNF- α and IL-22 in humans.

2.3 Perspective

Synthetic biology is poised to improve gene and engineered cell-based treatments for many diseases by providing precise control over the intensity, timing, and context of therapeutic intervention. Synthetic biology-inspired modules such as safety-switches and gene editing technologies are being introduced to clinical trials, and more sophis-

ticated gene circuits may one day enable advanced therapies like direct *in vivo* trans-differentiation. However, while complex synthetic systems have been demonstrated by a growing number of proofs-of-concept in the lab, challenges remain in developing synthetic biology-enabled therapies. This section explores several of these challenges, including designing synthetic gene networks that meet specified performance goals, translating them from *in vitro* testing environments to an *in vivo* therapeutic context, and delivering them efficiently into the patient. We also discuss recent advances in all three areas, which together are moving synthetic biology closer to the clinic.

Despite synthetic biology’s explosive development in the last decade, it is still challenging to build gene circuits that behave as anticipated [76], often requiring many design-build-test iterations before a synthetic gene network meets its performance goals. One common reason is unaccounted-for context effects: genetic circuits operate using host cell resources, which vary from cell to cell and form a finite pool from which all cellular processes (both native and engineered) must draw. Resource competition and other contextual effects can lead to modules that have different behavior depending on other modules in the circuit, and circuits that have different behavior depending on the cell type they operate in [77]. Characterizing genetic modules in multiple cell types or specific contexts of interest as well as in combination with different modules might help better account for this context sensitivity when designing larger circuits and systems. In parallel, by developing a deeper understanding of how cellular environments influence the behavior of synthetic modules, bioengineers may design modules and circuits that are better insulated from the cellular context.

One manifestation of context sensitivity is the poor agreement observed anecdotally between circuit performance *in vitro* and *in vivo*. Mammalian gene circuits are typically developed in cell lines cultured in artificial environments, which facilitates rapid testing of many circuit variants. While these culture systems are experimentally tractable, they do a poor job recapitulating the heterogeneous, dynamic *in vivo* environment in which a therapeutic circuit must ultimately operate. For instance, differences in intracellular biomolecule levels between cell lines and primary cells [78] could be debilitating for synthetic gene circuits whose proper function depends sen-

sitively on the concentrations of their modules' inputs.

Bridging this gap in circuit behavior may be possible using *in vitro* test systems that more closely resemble the environment inside the body. Promising technologies include organoids, which are three-dimensional "organ buds" grown *in vitro*, and "organ-on-a-chip" systems where cells are grown in microfluidic systems that mimic tissue properties. Both of these platforms have shown utility in emulating disease pathologies [79], [80]. For example, Ogawa et al. created cholangiocyte organoids from iPSCs of cystic fibrosis (CF) patients that recapitulated important aspects of the CF disease phenotype [81]. While organoids derived from normal iPSCs could regulate cyst swelling via cystic fibrosis transmembrane conductance regulator (CFTR)-mediated fluid transfer, this capability was lost in organoids derived from CF patients' cholangiocytes. The researchers then demonstrated that cyst swelling could be rescued in the diseased organoids by modulators of CFTR. Such organoids may one day facilitate testing of therapeutic synthetic gene circuits in a setting that closely mimics relevant disease pathologies. Another approach might be to characterize genetic circuits embedded in engineered "designer cells" using human whole-blood culture systems [75,82]. These systems simulate the environment the engineered cells will be exposed to in a patient, helping more accurately predict their performance *in vivo*. For example, engineered HEK-293T cells encapsulated in alginate and co-cultured with whole blood were able to respond to the TNF- α produced by primary immune cells stimulated with bacterial lipopolysaccharides [82]. The continuing development of such platforms will enable high-throughput circuit characterization and optimization in more physiologically relevant settings.

A complementary approach to address gene circuits' unwanted context sensitivity is to develop biological modules and circuit designs that are better insulated from cellular context, for example by minimizing spurious interactions (crosstalk) with other molecular species or by reducing reliance on host factors and processes. Such modules and designs should be easier to model computationally and may better maintain their behavior as they are moved from model systems to *in vivo* use. One design technique to achieve this is to import modules and molecules from other organisms

that are expected to have minimal crosstalk with native mammalian molecular networks. For example, a modified form of the *E. coli* transcription factor TetR and its cognate DNA sequence has found broad use in mammalian systems [11]. While this module relies on endogenous transcription machinery, there is minimal non-specific interaction between TetR and other DNA sequences in mammalian cells, which allows for some degree of logical insulation of the module from the cellular context. Another design strategy is to use modules whose interaction partners can be controlled more rationally, such as CRISPR-Cas and RNAi modules whose binding is based on Watson-Crick base pairing; although in both cases off-target binding effects are still being studied.

Both of these approaches to building more predictable gene circuits can be supported by advances in computer-aided design tools. Better software for designing and simulating biological circuits could reduce the number of circuit variants that need to be built and tested, leading to faster and cheaper development of synthetic biology therapies. Early efforts to simulate the behavior of synthetic gene networks relied on mechanistic models that captured each species' production, transport, binding, etc. [83], but these models' predictive power decreased as the size of the gene networks grew. Some recent efforts have taken a less mechanistic, more phenomenological approach: for example, Nielsen et al. developed a software package, Cello, that automates the design of biological circuits whose desired behavior is expressed in the digital logic design language Verilog [84]. Of the 60 circuits that they designed in *E. coli*, 45 performed correctly for every predicted output. Cello relies on a library of genetic modules with well-characterized input-output relationships, composing modules together by mapping the output range of one module to the input range of another. Importantly, circuit predictions became more accurate after incorporation of constraints to exclude combinations of modules that behaved unpredictably as well as mechanisms to insulate individual modules.

This phenomenological approach parallels recent progress in modeling and predicting the behavior of mammalian gene circuits from Davidsohn et al [85], who adopted a hybrid phenomenological/mechanistic strategy for modeling transcriptional cascades

in transiently transfected mammalian cells. Because such experimental systems never reach a steady state, Davidsohn et al used a set of rate functions to model the production and loss of each transcriptional product. By characterizing several input/output relationships of regulatory parts and then keeping track of their evolution over time, the investigators achieved predictions with a 1.6-fold mean error over a 261-fold range in output. Such a hybrid approach may enable more advanced and predictive biodesign and modeling tools for therapeutic gene networks.

Finally, a key hurdle to therapeutic deployment of synthetic gene circuits is safe and efficient delivery into a patient's body. One promising approach for delivery directly into patient cells is to use adeno-associated virus (AAV) since it efficiently delivers genetic material to cells in the human body with minimal immune response [3]. While AAV has been used in several recent gene therapy clinical trials [3], the nucleic acid packaging capacity of AAV is only approximately 5kb [86], which is too small for many of the synthetic gene circuits described above (although AAV co-delivery is being investigated). Other viral vectors such as herpes simplex virus type 1 (HSV-1) have a much larger packaging capacity, well over 100 kilobases, but their immunogenicity limits their applications [87].

A number of alternative nucleic acid delivery methods are also under development. One approach is to introduce purified DNA or mRNA into cells using physical forces such as electroporation or synthetic carriers such as zwitterionic lipids or cationic polymers [88]. These methods are not subject to the same packaging capacity limits as viral vectors, and furthermore can be produced in a completely cell-free manner, simplifying the manufacturing process and reducing the risk of unexpected contaminants in the final product. Unfortunately, non-viral delivery systems have their own challenges: mechanical methods such as electroporation work well *in vitro* but are difficult to use in human subjects, and synthetic carrier-based delivery of large nucleic acids often triggers an undesirable immune response [89]. Chemically modified nucleic acid vectors and biodegradable synthetic carriers that have reduced toxicity due to their rapid elimination represent a promising step forward in advancing these systems into the clinic [89, 90].

In contrast, engineered cell therapies deliver genetic material to cells *ex vivo*, which are then used as "living therapeutics." In methods based on adoptive cell transfer, a patient's own cells are extracted, engineered and expanded in a laboratory, then transplanted back into the patient. This approach allows for efficient *ex vivo* delivery of genetic material and has seen recent successes in clinical trials, including CAR T cell-based cancer therapies [2] and engineered hematopoietic stem cells used to treat β -thalassemia [1] and adenosine deaminase severe combined immunodeficiency (ADA-SCID) [91]. Alternately, genetically engineered cells can be microencapsulated in a biocompatible polymer matrix such as alginate and transplanted directly into the body [63], as described in the diet-induced obesity and psoriasis circuit examples above. Because the encapsulated cells do not provoke an immune response, these "prosthetic" genetic circuits can be tested and optimized *in vitro* in the same cell line in which they will operate *in vivo*, increasing the likelihood that the therapeutic circuit will function as desired.

More broadly, synthetic biology-based therapies carry the risk that either the delivery vector or the proteins expressed by the gene may cause T cells to become reactive or induce so-called anti-drug antibodies (ADAs) [92]. The consequences of an immune response can vary from a reduction in therapeutic efficacy all the way to a life-threatening reaction. For example, a T cell response to non-human proteins in *ex vivo* engineered cell therapies may cause those cells to be destroyed, blunting the efficacy of the therapeutic. On the other hand, ADAs against human proteins or non-human proteins expressed by gene therapies *in situ* could lead to a severe autoimmune response. To address these issues, strategies have been developed to suppress or mitigate the immune responses against therapies. For example, administration of corticosteroids was used to dampen the T cell response to AAV capsids in a clinical trial to express factor IX in hemophilia B patients [93]. Importantly, the expression levels of factor IX were sustained for several years in those patients who were promptly treated with steroids following a T cell response [94]. More recently, Kishimoto et al. co-administered therapeutic biologics along with poly(lactic-co-glycolic acid) (PLGA)-encapsulated rapamycin nanoparticles to mice and non-human

primates in order to induce immunological tolerance towards the biologics and prevented anti-drug immune responses [95,96]. A recent clinical trial used this strategy to induce tolerance against a yeast uricase enzyme for the treatment of gout [97]. In the future, it may be possible to encode mechanisms to prevent anti-drug immune responses in therapeutic gene circuits themselves.

The rapid development of our ability to manipulate biological systems using synthetic genes has direct implications for medicine. More than two decades after the first gene therapy trial was initiated [98], we have witnessed several regulatory approvals of gene and engineered cell-based therapies [99, 100]. The advancement of gene and engineered cell therapies into the clinic brings with it opportunities for synthetic biologists to create new treatments using synthetic gene circuits. These circuits promise to make gene and engineered cell therapies both safer and more effective as well as enable treatment options for diseases, genetic and otherwise, that are currently intractable. Taken together, these prospects will continue to propel synthetic biology into the clinic where it can have significant impact on human health.

Chapter 3

One-Pot characterization and optimization of genetic circuits

This chapter contains work from the following manuscript currently under revision and therefore represents a collaboration between myself, Jeremy Gam, Ross Jones, Jin Huh and Ron Weiss.

Gam, J., DiAndreth, B., Jones, R., Huh, J., Weiss, R. (2019) "One-pot transfection method for rapid characterization and optimization of genetic systems"

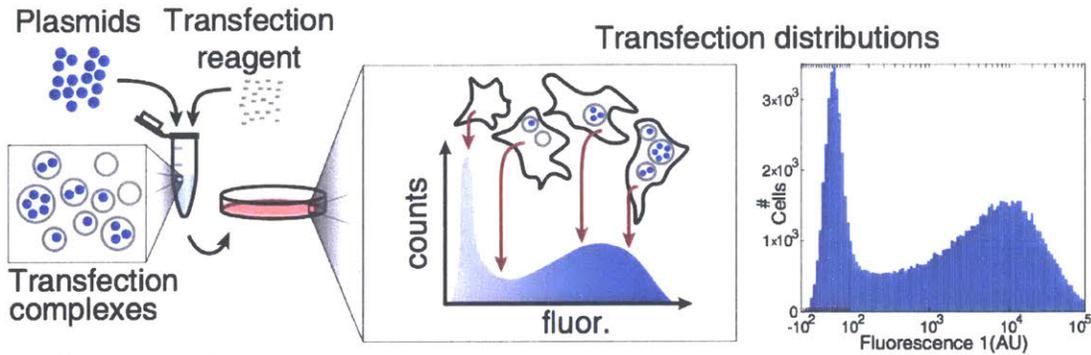
Biologists frequently use transient co-transfections to introduce exogenous DNA, RNA, or proteins into cells to modify cellular signaling or behavior. In many cases, multiple genetic elements (e.g. plasmids) must be delivered simultaneously. For instance, simultaneous expression of several transcription factors has been used to reprogram cell fate [101], and both guide RNA (gRNA) and Cas9 are needed for targeted nuclease activity [102]. Often, the relative ratios of these elements are important for function; for example, specific stoichiometries of reprogramming transcription factors improve reprogramming efficiency by several fold [103, 104]. However, ratios are often chosen based on intuition, trial and error, or coarse optimization, which can hinder realization of the desired phenotype. For instance, popular plasmids for CRISPR/Cas9 (e.g. pX330 [105]) encode constitutive high expression of both gRNA and Cas9, but recent results show significant gains in editing efficiency after optimizing the ratio of gRNA to Cas9 [106, 107]. Moreover, when delivering many plasmids, the exponential

increase in possible stoichiometries can drastically reduce the likelihood that subjectively chosen expression levels will yield optimal or even functional behavior [108]. Several methods have been prototyped to address this problem, but these involve complex and expensive pooled experiments [109–111], microfluidics [108,112], or time-consuming manual approaches [113]. Pooled experiments in mammalian cells include many steps that each require significant expertise, including: design of pooled DNA, assembly into a uniform library, delivery of the library into cells using virus, and preparation and analysis of high throughput sequencing experiments [114,115]. On the other hand, microfluidic experiments require dedicated clean rooms, equipment, and expertise [116]. More commonly, coarse optimization is achieved by manually varying stoichiometry in many different independent co-transfection samples. Thus co-transfection experiments are often tedious for smaller systems and infeasible for larger systems such as those being developed in synthetic biology.

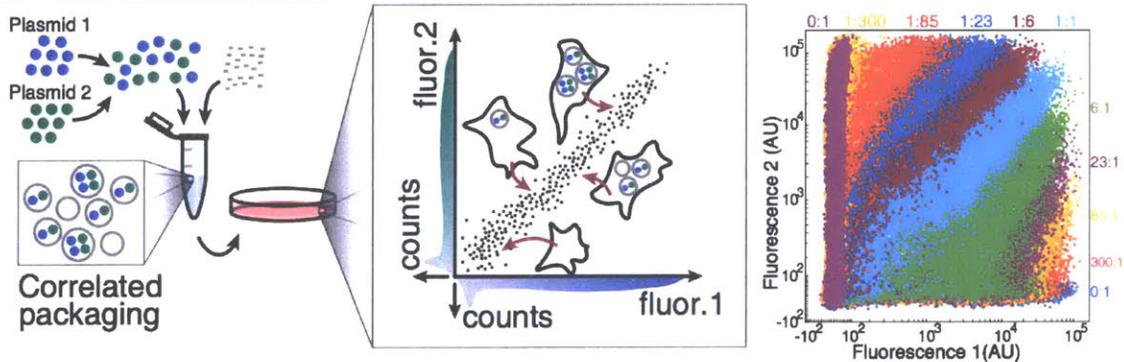
Here we introduce a one-pot ‘poly-transfection’ method, which enables the evaluation of a wide range of plasmid ratios for a given genetic system, all in a single sample. Notably, poly-transfection requires only a very simple change to a typical co-transfection protocol. In a typical co-transfection, plasmids are mixed before the transfection reagent is added, which results in highly correlated delivery of the plasmids. Therefore, even though transfection efficiency results in expression that varies over several orders of magnitude (Fig. 3-1a), only a single stoichiometry can be evaluated for the genetic system across that broad range (Fig. 3-1b) [117,118]. For poly-transfections, we instead mix each plasmid separately with transfection reagent before adding them to cells, resulting in decorrelated delivery of each plasmid (Fig. 3-1c) [117,118]. Rather than including a single transfection marker for an entire co-transfection sample, we include a distinct transfection marker with each plasmid, such that each fluorescence color intensity serves as a proxy for the concentration of a specific plasmid. Therefore, poly-transfections truly take advantage of broad transfection distribution in order to broadly sample various plasmid stoichiometries.

When combined with single-cell analysis methods such as flow cytometry, each poly-transfected cell provides an independent measurement of how the system behaves

a Single transfection



b Co-transfection



c Poly-transfection

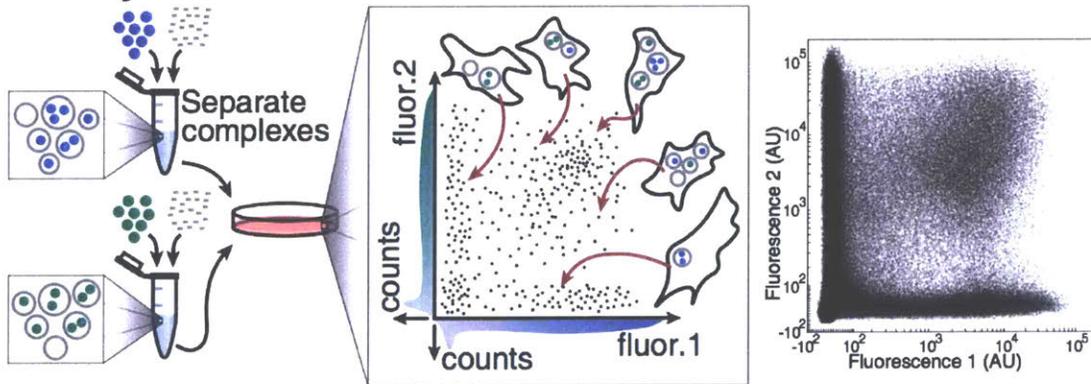


Figure 3-1: Overview of poly-transfection and comparison to other methods: For each transfection method, the leftmost diagram shows formation of transfection complexes between negatively charged DNA and positively charged lipid. In these examples, each color plasmid encodes expression of a different fluorescent protein. The center diagram shows examples of plasmid delivery to cells and also a schematic for the expected distributions in a histogram or scatter plot. Coloring intensity on the histogram corresponds to fluorescence from the corresponding plasmid color. The rightmost diagram shows real data from cells transfected using the given method.

at a specific combination of plasmid concentrations. In addition to providing more informative data, poly-transfections simplify experimental planning and execution to a single sample, resulting in significant savings in active experiment time. We have

demonstrated poly-transfections in systems containing up to four different plasmids, and depending on the experiment up to seven would be possible, enabling efficient study and optimization of many current and future genetic systems [119]. While in this study we focus on effects of stoichiometry of DNA plasmids, poly-transfections can likely also be applied to other types of molecules like RNA or proteins and possibly nucleic acids that encode multiple genes or cistrons.

3.1 Evaluating fluorescent protein orthogonality

For the purposes of this study, we examined whether poly-transfection behavior would be suitable for up to five-dimensional circuit characterization, which would allow for up to 4 circuit parts in addition to one output. We chose to examine up to five dimensions since the laser/filter settings for most available flow cytometers limits the use of fluorescent proteins to around five, further fluorescent proteins would result in significant spectral overlap [120]. A first step was to determine a compatible set of fluorescent proteins that could be spectrally distinguished. We tested a panel of 22 different fluorescent proteins to determine how many could be measured without significant bleedthrough in signal between channels in our flow cytometer (Figure 3-2a). Fluorescence was measured for each fluorescent protein in each channel and scores were calculated for all combinations using the following equation: $score = \log_{10}(99th\text{percentile}\text{channel}\text{fluorescence}\text{from}\text{sample}) - \log_{10}(\text{channel}\text{background}\text{from}\text{untransfected})$. Fluorescent proteins were chosen for emission across a range of spectra and for excitation with the 5-laser cytometer we used for the study. We also tested several high stokes shift proteins (LSS-mOrange, mBeRFP, mT-Sapphire) to determine whether two fluorescent proteins could be measured from the same laser, though those still resulted in significant bleedthrough to other channels. We observed several incompatibilities of note: mKate2 and mRuby2 have high bleedthrough into many channels, yellow fluorescent proteins (EYFP and mVenus) show high signal in the BUV 396 channel used to measure Sirius so green FPs are more desirable when both are to be measured, and all red fluorescent proteins

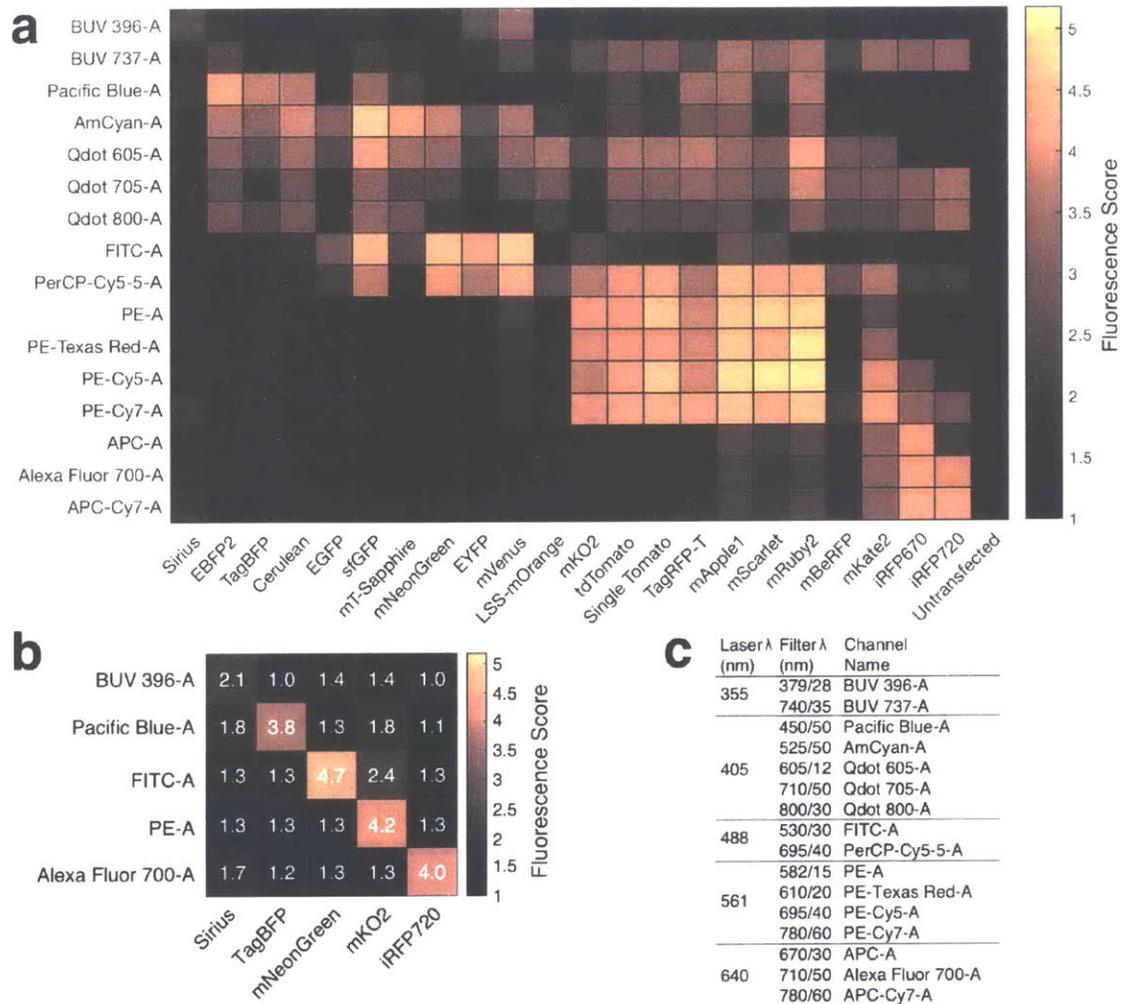


Figure 3-2: Spectral Analysis of fluorescent proteins to determine compatible sets.: (a) Fluorescent scores for each of the 22 fluorescent proteins measured in each channel. (b) A compatible set of fluorescent proteins where one fluorescent protein can be measured with each laser. (c) Laser and filter setting used to analyze the fluorescent proteins.

we measured (notably tdTomato, mScarlet, TagRFP-T, mApple1) showed stochastic signal in the Pacific Blue channel which meant compensation could not be used to correct for bleedthrough - instead orange fluorescent proteins (mKO2 measured in the PE channel) appear to be a good alternative. We determined that the set comprised of Sirius, TagBFP, mNeonGreen, mKO2, and iRFP720 (measured in BUV396, Pacific Blue, FITC, PE, and Alexa Fluor 700 channels respectively) show the most orthogonal signals and bleedthrough that could be compensated for. Note that when designing experiments with fewer than five fluorescent proteins, the lower brightness of Sirius and moderate bleedthrough of mKO2 into mNeonGreen may not be ideal.

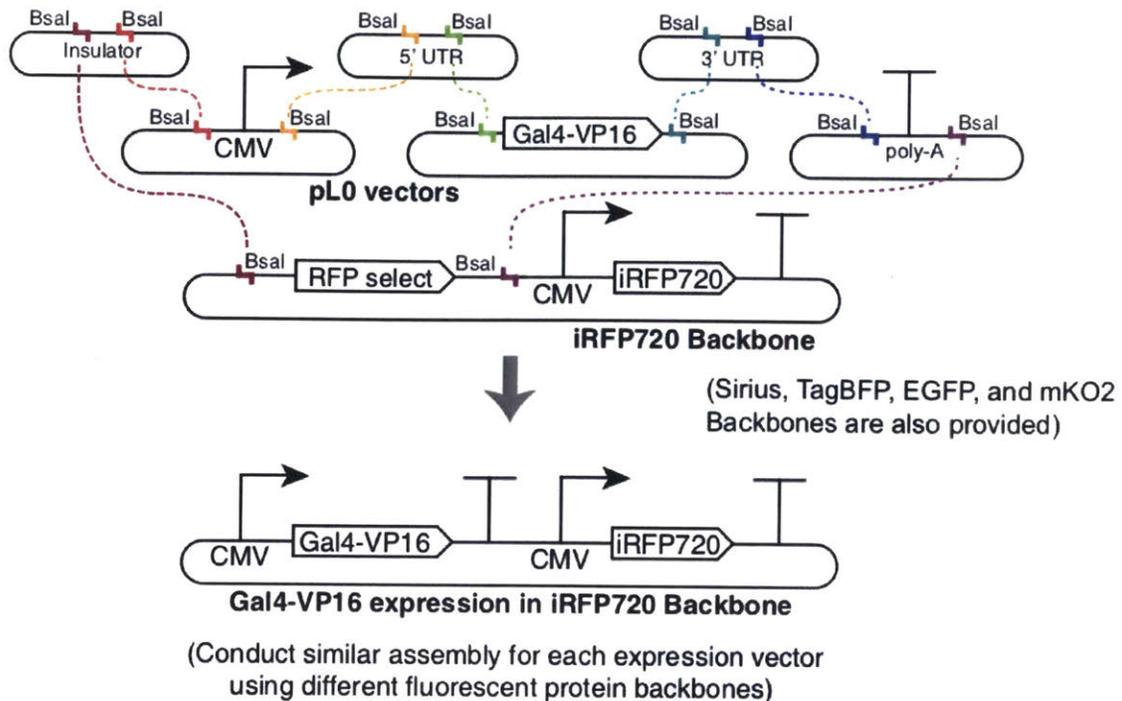


Figure 3-3: Simplified assembly strategy for one-pot-ready plasmids: Backbone vectors were cloned using the 5 compatible fluorescent proteins identified in Figure 3-2 with EGFP replacing mNeonGreen to facilitate cloning components and transfection markers on single-plasmid.

Listed are the laser and filter configuration of the BD LSRFortessa we used to collect data.

We next constructed a set of backbone plasmids containing constitutive expression of one of five fluorescent proteins (Sirius, TagBFP, EGFP, mKO2, iRFP720) along with restriction sites and colorimetric selection for BsaI-mediated Golden Gate assembly (Figure 3-3). In a single Golden Gate step, level 0 plasmids (pL0) encoding insulator, promoter, 5' UTR, gene, 3' UTR, and poly-A sequences can be integrated into the backbone and white colonies can be selected against red-expressing background colonies. Plasmids assembled in this way are immediately ready for use in poly-transfections since they encode simultaneous expression of a circuit part (Gal4-VP16 in the illustrated example) and a fluorescent transfection marker (iRFP720 in the example). Note that EGFP is substituted for mNeonGreen since the latter is not available for distribution, though both show similar spectral properties.

3.2 Benchmarking poly- against co-transfection

We first determined whether a single poly-transfection can accurately recapitulate results from co-transfections. For this purpose we tested two regulatory modules: translational repression by RNA binding protein L7Ae and transcriptional activation by Tet3G (a variant of rtTA). Each of these systems has two constituent plasmids: one encoding the regulatory protein (L7Ae or Tet3G), and a second plasmid encoding a fluorescent reporter that is controlled by the regulatory protein (Fig. 3-4d,h). We also encoded expression of a different constitutive fluorescent protein on each of the two plasmids in order to track their concentrations. For a given two-plasmid poly-transfection, we found that 11 co-transfections were required to cover the two-dimensional concentration space, achieved by varying the ratio of these two plasmids in each co-transfection (Fig. 3-4b-c). Next, data from these sets of co-transfections were collated to form a single dataset that could be compared to poly-transfection data. We then binned both the poly-transfection data and collated co-transfection data according to the two transfection markers and evaluated output reporter fluorescence in each bin (Fig. 3-4e,i). Output fluorescence proved to be quite similar (Fig. 3-4f,g,j,k) with 66% of all L7Ae values and 87% of all Tet3G values achieved by poly-transfection falling within 2-fold change of their associated co-transfection-derived values. Despite this bias, we obtained sufficient numbers of cells spanning the concentration space. Overall this similarity provides confidence that cells with a given transfection marker fluorescence represents the evaluation of a specific plasmid concentration regardless of the transfection method used to achieve it.

3.3 New characterization and insights

In addition to benchmarking poly-transfections and comparing to traditional co-transfections, we used poly-transfection to generate new insights and characterization for biological parts. We first determined the activation strength for a dCas9-VPR/gRNA transcriptional activation (CRISPRa) system. CRISPRa systems have

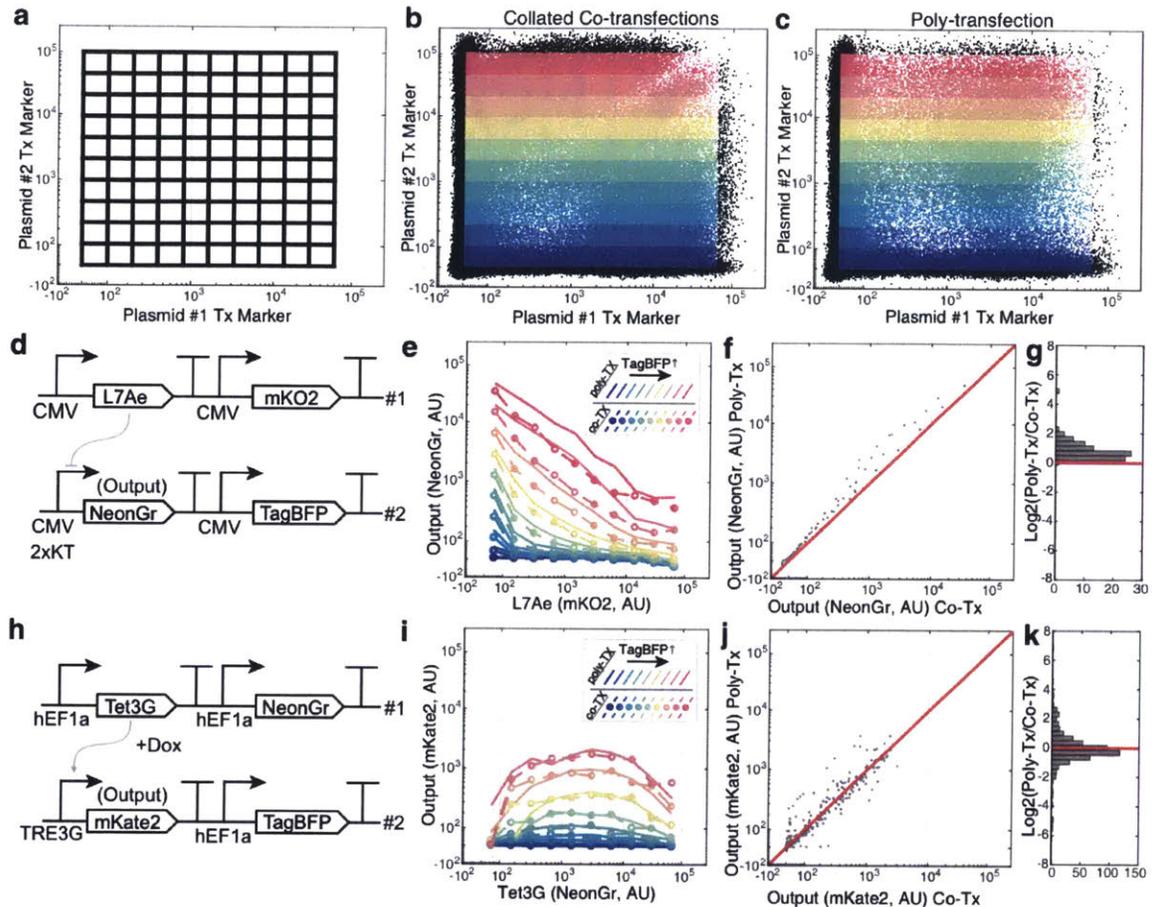


Figure 3-4: Benchmarking poly-transfection against co-transfection: Benchmarking poly-transfection against co-transfection. (a-c) Binning workflow for bench-marking analysis. Ten biexponentially-spaced bins were assigned for each transfection marker dimension which approximate the levels of each of the two plasmids (a). Binning was performed on both a collated set of 11 co-TX samples spanning various plasmid ratios (b) and data from a single poly-TX (c). Colors correspond to sets of bins defined by gene 2 (TagBFP) levels. (d-k) Median output fluorescence was evaluated for the cells in each bin and compared between methods for each system: L7Ae translational repression (d-g) and TetG transcriptional activation (h-k). (d) Constructs for measuring L7Ae activity. (e) Multi-dimensional titration curves for L7Ae. Each line represents the set of bins at one level of TagBFP as denoted in (b) and (c). (f) Visual comparison between co-TX and poly-TX for L7Ae system. Each point represents the measured output in corresponding bins for poly-TX and co-TX measurements, where more equivalent values are closer to the red 1:1 line. (g) Histogram for discrepancy between poly-TX and co-TX for the L7Ae system. The values are the log₂ fold-changes from the co-TX to poly-TX for each point in (f). Mean = 0.87, standard deviation = 0.65. (h) Constructs for measuring Tet3G activity. (i) Multi-dimensional titration curves for Tet3G at the highest level of doxycycline (Dox = 2.3 μ M). (j) Visual comparison between poly-TX and co-TX for Tet3G system (inclusive of all doxycycline levels)(k) Histogram for discrepancy between poly-TX and co-TX for Tet3g system, calculated as in (g). Mean = -0.09, standard deviation = 1.31.

been used for various applications, though quantitative information about how activation strength depends on the stoichiometry of dCas9-activator and gRNA has not been

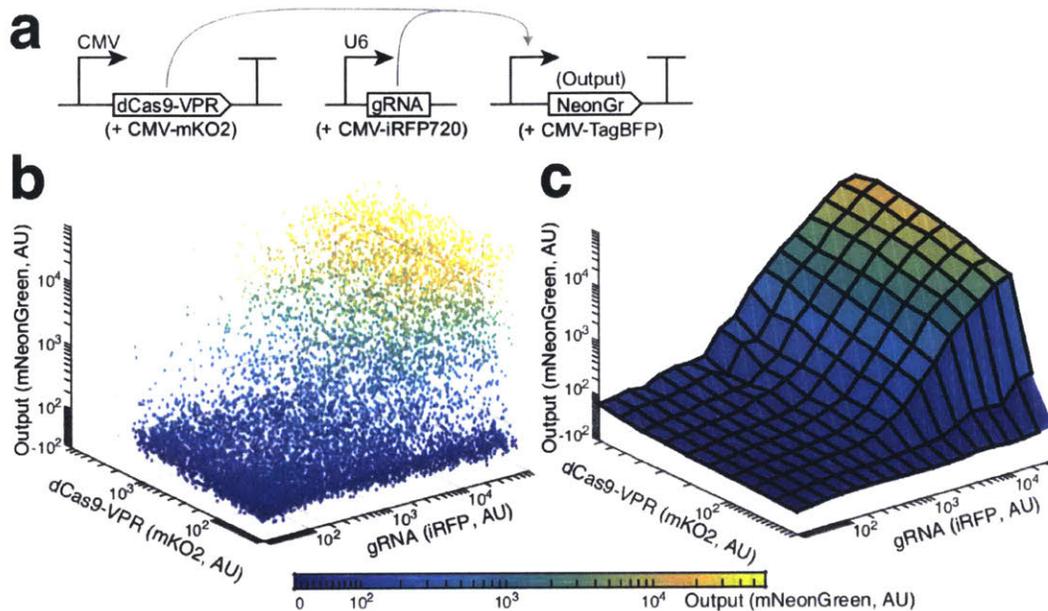


Figure 3-5: Novel CRISPRa characterization enabled by poly-transfection: Characterization of transcriptional activation of a fluorescent reporter by dCas9-VPR. Contributions to activation by dCas9-VPR and gRNA can be measured simultaneously with our one-pot method. Shown are the circuit diagram (a), subsampled scatter (b), and a surface plot (c) of output as a function of gRNA and dCas9-VPR at intermediate TagBFP transfection marker levels ($1e3$ to $1e4$ AU). Surface plot indicates the medians of mNeonGreen within each bin.

explored in depth [121, 122]. Using a single poly-transfection, we were able to quantify how concentrations of a gRNA-expressing plasmid and a dCas9-VPR-expressing plasmid contribute to target gene activation. Interestingly, we found that increasing gRNA concentration contributes to significant increases in activation, whereas dCas9-VPR is more uniformly effective across many concentrations as long as it is above a relatively low threshold (Fig. 3-5). These results are consistent with CRISPR editing studies where efficiency of Cas9 editing depended more on gRNA concentration than Cas9 mRNA concentration [106, 107]. Therefore activation efforts with CRISPRa and editing should benefit more from improved delivery of DNA coding for gRNA rather than Cas9.

We also used poly-transfection to examine the potency of synthetic miRNA sequences [123] in repressing a gene at the post-transcriptional level [8, 124]. Synthetic miRNA sequences have been recently leveraged to reduce leaky gene expression or generate more complex systems like feedforward loops [125, 126]. We sought to enable

these applications by quantifying both the potency of synthetic miRNAs as well as the orthogonality between miRNAs and the targeted genes, which would allow for more informed design choices. miRNA-expressing plasmids and target fluorescent reporters were each associated with a transfection marker (Fig. 3-6a). Similarly to other poly-transfections, a broad combinatorial space of combinations between the miRNA and miRNA sensors was explored (Fig. 3-6b). As miRNA concentration increased, reporter fluorescence decreased more significantly compared to the transfection marker in a threshold-like manner as previously observed from similar sensor constructs (Fig. 3-6c,d) [127, 128]. We then demonstrated that detailed information about miRNA orthogonality could be obtained from poly-transfection data, at a reduced number of transfections compared to co-transfections. We conducted poly-transfections using all possible combinations of eleven different miRNA and sensor sequences, resulting in $11 \times 11 = 121$ separate poly-transfections which were completed in less than two hours. To obtain somewhat similar data using co-transfections would have required approximately $10 \times 11 \times 11 = 1210$ transfections, since around 10 co-transfections are required to adequately cover a 2 dimensional concentration space (Fig. 3-1b). Analysis showed that synthetic miRNAs FF5 and FF6 have the strongest potency and all tested synthetic miRNAs have minimal cross-talk except for the miR-FF6 / miR-SHC007 pair, since both are derived from the same luciferase sequence. We were also able to easily extract key miRNA activity metrics of M and K_M for each synthetic miRNA [128]. These examples show how CRISPRa and synthetic miRNA systems - and by extension, future yet uncharacterized systems - can be easily characterized using poly-transfection.

3.4 Optimizing a complete genetic system with poly-transfection

Next, we used poly-transfections to rapidly engineer a difficult-to-optimize genetic system for discriminating cancerous from noncancerous cells [8]. Cancer cells often

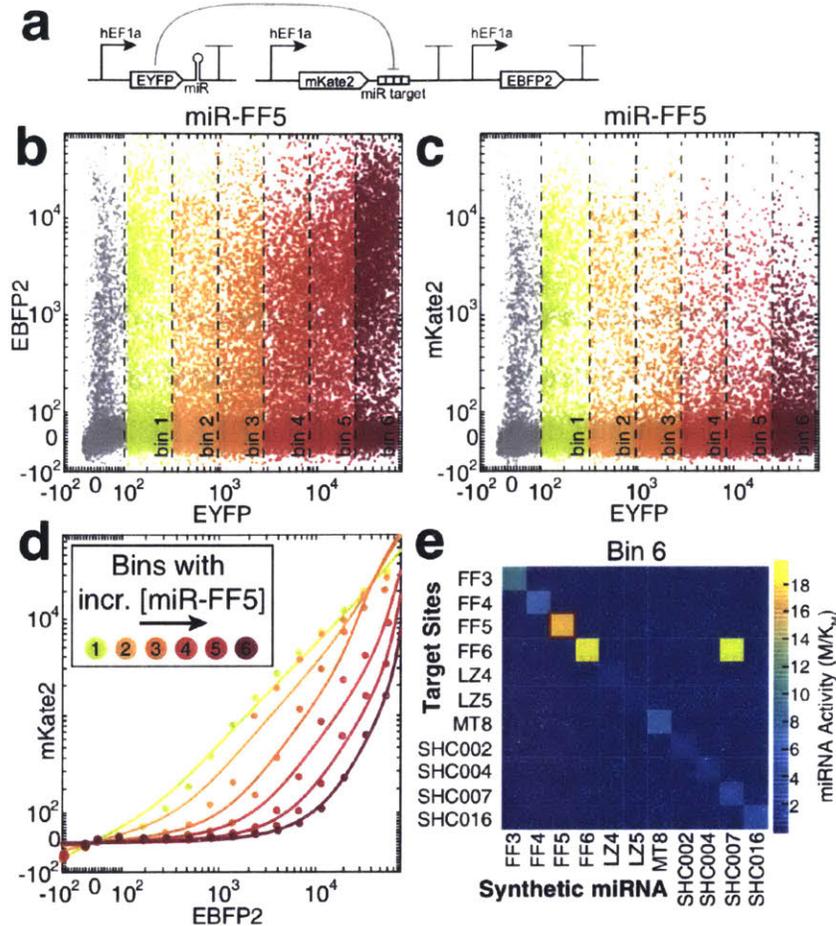


Figure 3-6: Novel RNAi characterization enabled by poly-transfection: (a) Genetic circuit for testing miRNA behavior. The circuit consists of two plasmids where the first expresses a synthetic miRNA based on a miR-155 expression platform and also EYFP marker serving as an indicator of miRNA expression. The second plasmid encodes a miRNA low sensor where mKate2 fluorescent protein contains miRNA target sites in the 3' UTR which mediate repression of mKate2 in the presence of the complementary miRNA. EBFP2 transfection marker indicates how much of this sensor plasmid is delivered to each cell. (b) Scatter plot of poly-transfection input parameter space for miR-FF5. (c) Scatter plot of output as a function of miR-FF5 concentration. (d) Sensor performance at different miR-FF5 concentrations. We binned data in each of the EYFP bins further using EBFP2 and calculated the mKate2 medians (circles). Lines indicate fits to a miRNA repression model. (e) Orthogonality of synthetic miRNAs and sensors. We generated orthogonality matrices for the six EYFP bins, where rows indicate which target sites were present on the sensor and columns indicate the synthetic miRNA introduced. Shown is the sixth bin, which corresponds to the highest synthetic miRNA expression.

express the biomarker miRNA-21-5p highly; [129] thus, we built and optimized a cell classifier which produces a genetic output only in the presence of miR-21-5p (i.e. a single high miRNA classifier). Our classifier is composed of three DNA components and responds to miR-21-5p in the following way: (i) at high miR-21-5p activity, the miRNA degrades the transcript encoding the BM3R1 transcriptional repressor,

(ii) degradation of BM3R1 allows transcription from the promoter driving mKO2 reporter, and (iii) a Gal4-VP16 transcriptional activator is needed for output expression (Fig. 3-7a). Constraining this system to accurately produce output only in miR-21-5p expressing cells requires balancing the ratios of plasmid concentrations for BM3R1, Gal4-VP16, and reporter.

We optimized transfection-based classification of cells via the miR-21-5p classifier by using poly-transfection to identify an optimal DNA ratio of the three classifier plasmids that would specifically distinguish HeLa from HEK293FT cells, which have high and low miR-21-5p activity respectively. We evaluated classification accuracy at different plasmid ratios by computationally subsampling the 3-dimensional poly-transfection data, designating stoichiometric ‘trajectories’ through the concentration space (Fig. 3-7b), and including only cells that were in close proximity to the trajectory. Subsampling in this manner provided data similar to traditional co-transfections from which classification metrics including specificity, sensitivity, and accuracy could be calculated. Using an optimization method, we identified a well-performing ratio of 10.9 : 1.5 : 1 for Gal4-VP16 : output : BM3R1 plasmids at which HeLa and HEK293FT cells were distinguished with 91% specificity, 62% sensitivity, and 77% accuracy (Fig. 3-7c). We verified this prediction by performing a co-transfection at the corresponding ratio of plasmids and found that the specificity, sensitivity, and accuracy metrics agreed with the values computed from the subsampled poly-transfection (99%, 68%, 84% respectively; Fig. 3-7d). Further comparisons between poly- and co-transfections at 27 different DNA ratios representing titrations of each unique plasmid showed good correlation for specificity, sensitivity, and accuracy (Figure 3-7e). Analysis showed that high classification accuracy was possible with Gal4-VP16 expressed at a wide range of high levels, and with BM3R1 and output within a narrow band of lower concentrations (Fig. 3-7e). Overall, these data indicate that subsampled data from single poly-transfection samples can be used to quickly evaluate system performance, reducing or eliminating the need to iterate through different physical designs.

We then applied our findings to construct optimized single-plasmid versions of the

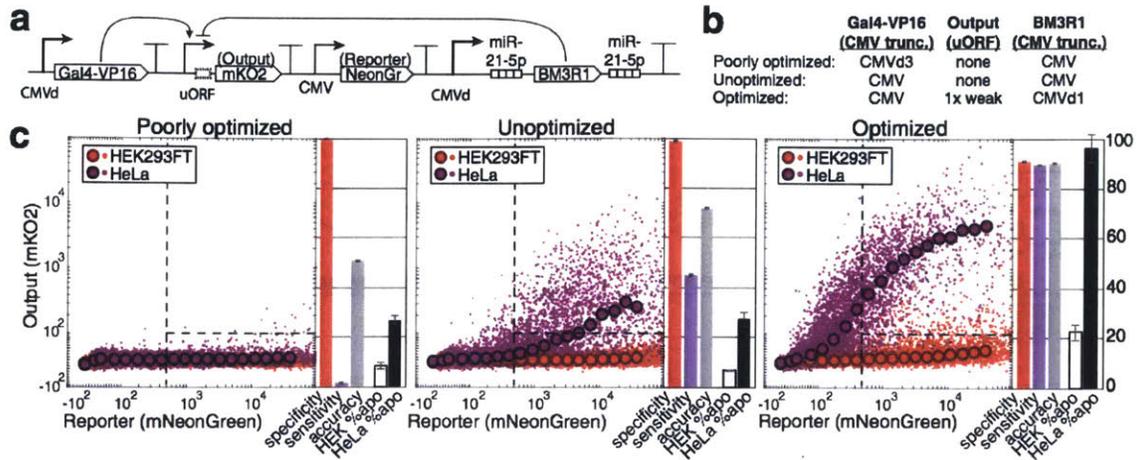


Figure 3-8: Evaluation of single-plasmid miRNA classifiers *in vitro*: (a) Circuit diagram for single-plasmid miRNA high sensors. Tunable parts include either upstream open reading frames (uORFs) or CMV truncations (CMVd), both of which tune down expression to a defined degree [130,131]. (b) Table of CMV truncations and uORFs used in single-plasmid constructs. The optimized circuit encodes high expression of Gal4-VP16 and reduced levels of output and BM3R1, the unoptimized circuit encodes high expression of all three, while the poorly optimized circuit encodes low expression of Gal4-VP16 and high expression of output and BM3R1. (c) Performance of an optimized single-plasmid high sensor compared to unoptimized and poorly optimized variants. Scatter plots of output (mKO2) as a function of transfection marker (mNeonGreen) in the poorly optimized high sensor (left) show low output in both HEK293FT and HeLa cells, resulting in high specificity (100%) but inferior sensitivity (2%) and classification accuracy (51%). The unoptimized high sensor (middle) shows low output in HEK293FT (specificity = 99%), but only high output in a fraction of HeLa cells (sensitivity = 45%), and an overall lowered ability to classify cells (accuracy = 72%). For the optimized high sensor (right), output remained low in most transfected HEK293FT cells (specificity = 91%) and high in most transfected HeLa cells (sensitivity = 90%), resulting in a generally high classification accuracy (accuracy = 90%). Results from the optimized classifier with Bax as output show much higher degree of killing in on-target HeLa cells (black) with low killing in off-target HEK293FT cells (white). Apoptosis percentages are relative to positive controls with constitutively expressed Bax. Error bars for bar charts indicate standard deviations for technical triplicates.

ratios of each gene to the optimized levels identified above by inserting a repressive upstream open reading frame (uORF) into the 5'UTR of the mKO2 reporter and truncating the CMV promoter driving BM3R1 (Fig. 3-8c) [130,131]. The miR-21-5p classifier optimized in this way showed higher classification accuracy in single-plasmid transfections compared to both an unoptimized classifier (high expression of all genes) and a poorly optimized classifier (low Gal4-VP16 expression via truncated CMV) (Fig. 3-8c). Additionally, when we replaced the fluorescent output with the apoptosis regulator Bax, the optimized design showed greater ability to selectively induce apoptosis in HeLa cells compared to the other classifier variants (Fig. 3-

8c), and demonstrated similar selectivity compared to other more complex miRNA classifiers [8, 61].

3.5 Discussion

Poly-transfections bridge the gap between simple but low information co-transfections and complex high-information pooled approaches. By taking advantage of innate variations in transfection efficiency, poly-transfections yield high-information data similar to modern single-cell analysis techniques, while also being substantially simpler. Specifically, the increased cost and complexity of FlowSeq methods [109] and droplet microfluidics [108] makes them unsuitable for small- to medium-scale systems (2-4 genes) commonly used in biology. Additionally, poly-transfections are simpler to perform and scale drastically better than co-transfections, which require exponentially more samples for each unique plasmid within the system. Furthermore, our poly-transfection method can be readily extended to newer technologies such as spectral analyzers and mass cytometers (e.g. CyTOF) that enable even higher dimensional datasets than are possible with the 5-color flow cytometers used in this study.

Overall our poly-transfection method represents a convenient and powerful one-pot approach for evaluating effects of DNA stoichiometries for diverse experimental systems. In addition to Cas9 and classifier systems shown here, we anticipate that future insights gained from poly-transfections will help biologists extract deeper information from genomic perturbations, design reporters with higher signal to noise ratio, and build tools that more robustly modulate gene expression, among other applications. Therefore, in addition to poly-transfections being immediately applicable to fields such as synthetic biology where optimization of complex genetic systems is already essential, we envision that the ease of our method will make such one-pot strategies more attractive to a wide range of other fields within biology.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Programmable endonucleolytic RNA scission-induced stability tuning (PERSIST) platform

The ability to regulate the level of synthetic gene expression within cells is a vital requirement for synthetic biology. Gene circuits are created by regulating the level of each synthetic gene, which in turn can regulate the level of other synthetic genes. Most mammalian genetic circuits remain based on transcription factor-based regulation due to its robust behavior, modularity, composability and large design space: promoters and genes can be easily swapped to exhibit programmable behavior. However, as the field moves towards applications such as gene and cell therapies, and biomanufacturing, DNA-based regulation may cause complications.

An important challenge is epigenetic silencing of the synthetic payload that has been engineered into the genome of cells used for cell therapies or biomanufacturing. Transgene silencing has been observed in a large number of cell types including stem cells [132], neurons [133], CHO cells used for antibody production [134], and Hek293 cells. While the causes for this epigenetic silencing are still being uncovered some factors contributing to avoiding silencing are becoming evident. First, the genomic location of transgene insertion must be strongly considered. Several "safe harbor" locations have been identified to help promote expression primarily in re-

gions of high transcription (open chromatin). Second, it appears epigenetic silencing is largely dependent on specific promoter sequence. Several viral promoters such as cytomegalovirus (CMV) promoter [134–137] and other sequences including the UAS binding site [138, 139] have been shown to induce silencing. And finally, irrespective of the promoter sequence epigenetic silencing has also been shown to be dependent on the transcriptional state of the gene, where strong constitutive expression may abrogate silencing [140–142]. In fact it was demonstrated that even a transient decrease in transcription (via TetOn or TetOff systems) could cause permanent silencing [132, 140, 143]. In fact, the gene therapy field has been developing a library of strong constitutive promoters that avoid silencing [144, 145]. Since current transcription factor-based circuits often require the replacement of these tried-and-true promoters with (often CMV-based) regulatable ones and depend entirely on changing transcriptional activity, perhaps it is no surprise that these suffer from epigenetic silencing.

Newer gene therapies represent a second challenge as many are encoded in mRNA. RNA-based therapies have several benefits including improved delivery since nuclear delivery is unnecessary, decreased chance of accidental genomic integration, and potentially more transient expression profiles. However because they are encoded in and delivered as RNA, they cannot utilize on transcriptional logic.

A post-transcriptional regulation platform would overcome these obstacles. Such a platform would allow integrated transgenes to be continuously expressed from verified constitutive promoters to combat epigenetic silencing (Fig 1A), while regulation could be programmed post-transcriptionally; such a platform would also find use in RNA-based therapeutics. However there are few examples of such platforms that enable activator-like activation, repressor-like repression, and subsequent layering into circuits. Post-transcriptional platforms based on proteases have recently been developed [146–148]. The Chomp platform developed by Elowitz et al. [147] comprises proteases that cleave off or expose a degron on a fluorescent reporter. Unfortunately layering these proteases was not as easy as placing appropriate target sites and actually involved engineering split protease versions with leucine zippers, where proteases

could repress another protease with a serious cost in performance (all less than 10 fold change). While protease-mediated activation and repression of a fluorescent protein was shown for three proteases, protease-mediated activation of another protease was never demonstrated.

Introducing an RNA intermediate, or in other words, having proteins that act on another protein's RNA instead of directly on that other protein has some favorable characteristics: one can avoid challenging protein engineering efforts, separation of a component's function and its regulation can lead to more composable and modular behavior, it allows interaction with endogenous biomarkers that act on RNA, it could potentially rely on WC base-pairing, and it might introduce a small but sometimes advantageous delay due to translation. There are several commonly-used devices for regulating RNA, such as the translational repressor L7Ae [35, 36], mRNA-degrading miRNAs [8], and RNA-cleaving Csy4 [149] however due to their limited numbers or mechanisms of action they are not easily composable. Additionally all of these examples typically reduce expression and therefore act only as repressor-like "OFF" switches. In fact, very few RNA-based activator-like "ON" switches have been developed for RNA [35, 149–152] and none of these are composable.

Here we engineer an RNA-based "ON"-switch motif that can be engineered into the 3' untranslated region (UTR) of any transcript and whose cleavage results in stabilization and expression of a transgene's transcript rather than canonical degradation. We demonstrate that the ON switch can be modified to respond to a variety of RNA cleavers including exogenous and some endogenous endonucleases, miRNAs and ribozymes. Importantly to our knowledge, although modest in performance, this represents the first single-transcript miRNA "high" sensor which could enable miRNA-based sense-and-respond therapeutics without the expression of potentially immunogenic intermediate proteins. We identify a set of 9 CRISPR-specific endoRNases that are orthogonal in their cleavage and can "activate" expression by cleaving their cognate ON-switch motif. When the recognition sequences of these 9 endoRNases are instead engineered into the 5' UTR of transcripts they canonically cleave and reduce expression by often over two orders of magnitude. Therefore these endoR-

Nases indeed function as RNA-level "activators" and "repressors" of gene expression. We term this platform for RNA regulation a "programmable endonucleolytic RNA scission-induced stabilization tuning" (PERSIST) platform.

First, to benchmark the platform against DNA-based analogs we show that this platform is quite modular and composable. It is modular because we can combine endoRNase-responsive 3' UTR ON-switches with 5' UTR OFF switches on the same transcript. Due to the multi-modal (RNA-Protein) nature, the platform can be easily composed to create layered circuits by swapping the regulated reporter gene for the genes of other CRISPR endoRNases. To demonstrate these features we create multi-level cascades, engineer all 16 two-input logic functions, and demonstrate a positive feedback motif. All motifs and functions behave impressively and often exhibit at least an order of magnitude difference between off and on states.

An interesting feature of these endoRNase regulators is that they can simultaneously function as both repressors and activators based on where their recognition site is located in the transcript, unlike transcription factors whose activator or repressor functions are most often mutually exclusive. As a result of this dual functionality, there are several useful motifs that can be engineered quite readily. We therefore demonstrate a FFL as well as a combined positive-feedback/repression motif. Importantly both motifs results in ultrasensitive responses which is a requirement for building more complicated circuits such as bistable switches and oscillators and is a behavior obtained in transcription factor activity through multimerization. We therefore expand the positive-feedback/repression motif to create an RNA-level toggle switch.

4.1 Developing PERSIST "ON" switches

In order to create an RNA-level "ON" switch we designed a fluorescent reporter expressing EYFP that contains the PERSIST-ON motif in its 3'UTR. This motif has three domains (see Figure 4-1): (1) degradation signals that cause the rapid degradation of the transcript, (2) a cleavage domain which allows the removal of the

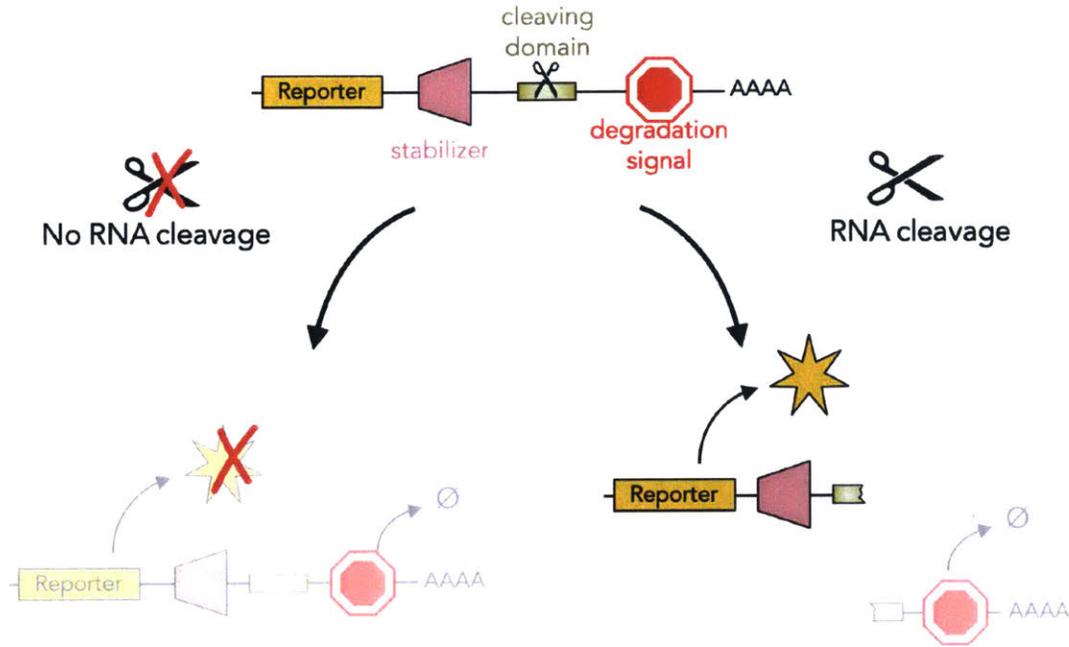


Figure 4-1: PERSIST "ON" switch motif design: For the PERSIST "ON" switch motif, the transcript will get quickly degraded by the cell's endogenous machinery due to included degradation domains. When an RNA cleaver removes these tags the transcript is stabilized and expression is restored.

degradation tag, and (3) a stabilizer that allows efficient protection and translation of the mRNA after the removal of the degradation signal.

First we sought to identify a mechanism that would lead to rapid mRNA degradation. A short RNA motif has been recently identified that binds hnRNPs which in turn recruit deadenylase complexes and results in mRNA degradation in a variety of cell types [153]; in this study 10 repeats led to an approximately 7 fold decrease in expression from an engineered reporter. Here we created a multimerizing golden gate reaction protocol (see Appendix A.1.2) to add many repeats of this domain, "wt1". We transfected plasmids encoding a reporter gene appended with various numbers of repeats included within its 3' UTR into HEK293FT cells and evaluated its fluorescence. The reporter containing 30 repeats was efficiently degraded with over 300 fold reduction in fluorescence, while fewer numbers of repeats allow for varied degrees of repression (Figure 4-2). Notably, while there are methods to program rates for RNA production, protein production and protein degradation (e.g. with promoter design [130, 154], uORFS [131], and destabilization domains [45, 46]) respectively),

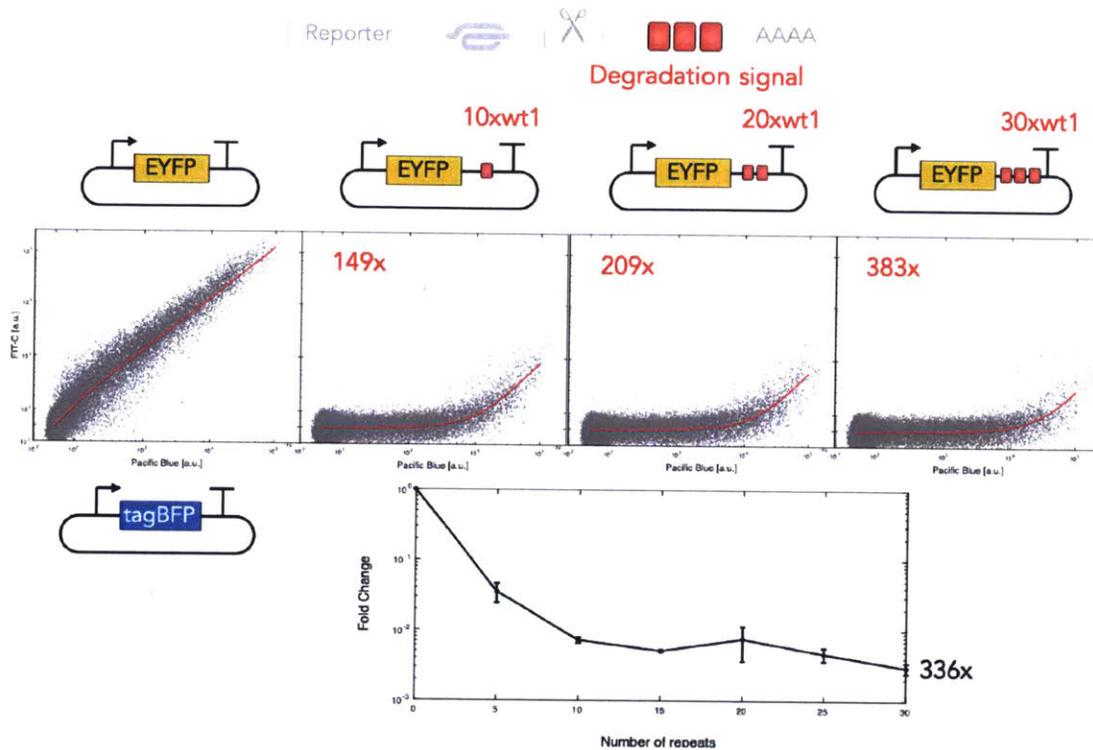


Figure 4-2: Degradation motifs robustly degrade RNA: Multiple degradation motifs, "wt1", were appended within the 3' UTR of an EYFP reporter plasmid which is tracked by the transfection marker, tagBFP. tagBFP vs EYFP fluorescence data was fit to a line in linear space to evaluate degradation (approximated by the slope of the line). Increasing numbers of repeats of the degradation signal causes rapid degradation of the reporter and reduced signal.

these degradation motifs represent a new way to program mRNA half-life.

Next, we looked for a sequence that would stabilize the transcript and allow for translation only after the removal of these degradation tags. Importantly, this stabilizer should not stabilize the transcript when the degradation tags are still present. An RNA triple helix (triplex) structure was identified to stabilize a highly-abundant long noncoding RNA, MALAT1, which lacks a poly-A tail [155–157] and has been used in several synthetic biology studies to stabilize engineered transcripts [149, 158, 159]. We tested the MALAT1 triplex followed by 30 repeats of the wt1 degradation motif (Figure 4-3). The construct exhibited low expression, similar to constructs containing the degradation motif alone. When a mascRNA sequence, which is cleaved by RNase P, was included between the triplex and degradation motif, expression was restored to constitutive levels (Figure 4-3). Overall this 3' UTR ON switch exhibits a dynamic range of 166, which is promising for its adaptation as an RNA ON-switch

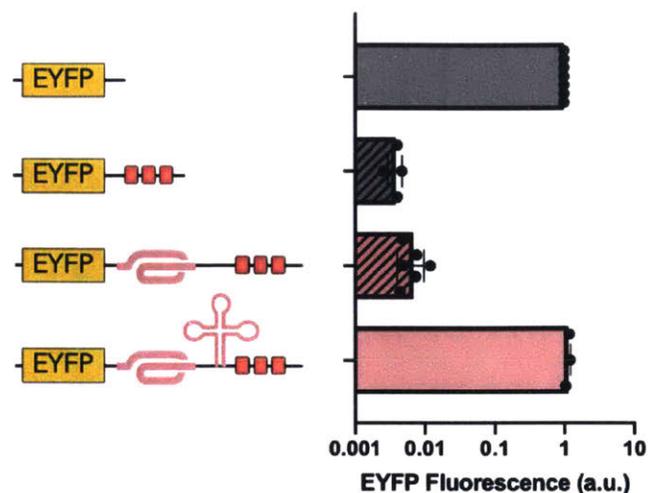


Figure 4-3: Triplex RNA stabilizer protects transcript after degradation tag removal: Evaluating the effect of adding the triplex stabilizer to the ON-switch motif. Red squares indicate 10x degradation signals. When triplex (pink curl, pink striped bar) is added upstream of the degradation signals, fluorescence still remains quite low and the transcript is not protected from degradation via the degradation motifs. When the RNaseP target, *mascRNA* (pink clover-leaf, pink bar) is added directly downstream of the triplex, the degradation signals are removed and the transcript is able to be stabilized by the triplex, resulting in full restoration of fluorescent signal. All values normalized to constitutive control (gray bar).

for synthetic regulation.

In order to generate a library of RNA-regulators we next sought to evaluate whether endonucleases could cleave and activate the ON-switch motif and canonically cleave and repress the transcript otherwise. The Cas6 family and Cas13 family of CRISPR-specific endonucleases process pre-cRNAs to produce the shorter gRNA sequences used for DNA-targeting and RNA-targeting respectively [160]. Importantly, these endonucleases cleave specific (often hairpin-structured) sequences and orthologs are thought to be quite orthogonal [161]. While Csy4 (PaeCas6f) has been used in mammalian synthetic biology applications before [149,158], due to the abundance of these endoRNases we hypothesized that other orthologs would function as well. We evaluated 9 total endonucleases; from the Cas6 family we used Csy4, Cse3 (TthCas6e) [162], CasE (EcoCas6e) [163] and Cas6 (PfuCas6-1) [164] and used 5 from the Cas13 family: LwaCas13a [31], PspCas13b [32, 43], PguCas13b [32, 43], RanCas13b [32, 43], and RfxCas13d [165]. As Csy4 and all of the Cas13 family en-

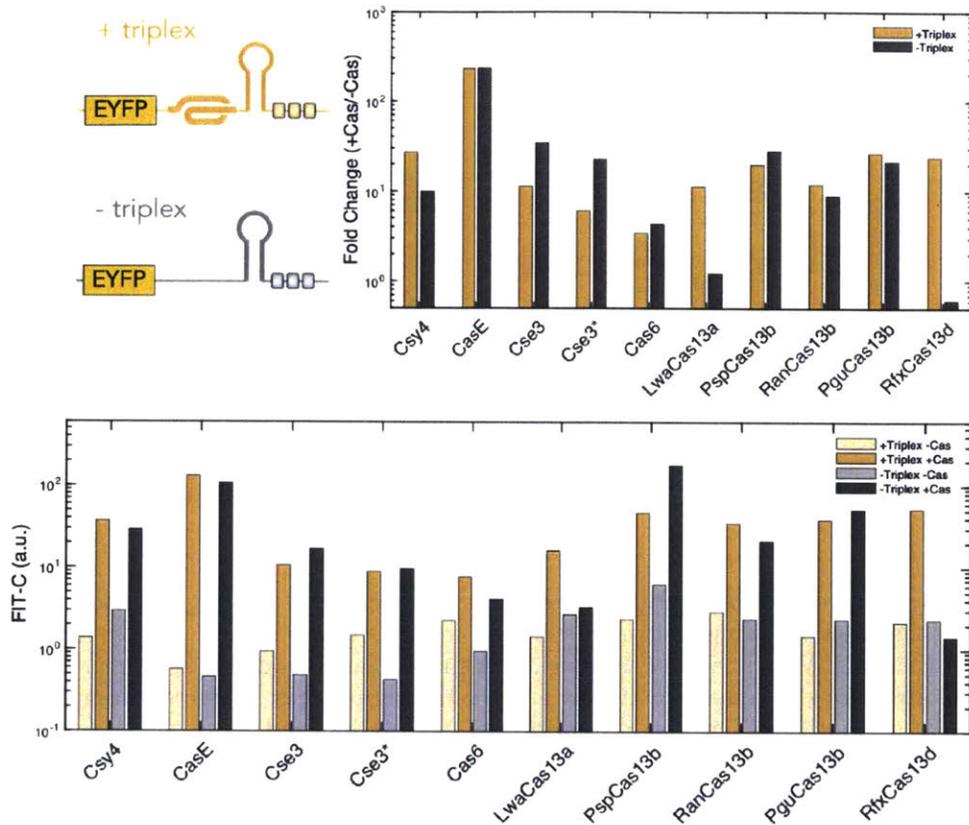


Figure 4-4: Triplex requirement is Cas-specific: ON-switch EYFP reporters with (yellow) and without (gray) the triplex were constructed for each endoRNase. Values for -Cas and +Cas (bottom) are fold change compared to a control with no target site.

endonucleases have previously been used in mammalian cells, only Cse3, CasE, and Cas6 required mammalian codon optimization. We observed robust "ON" response to most endoRNases when their cognate recognition sequences were placed between the triplex and degradation signal in the ON-switch motif (Figure 4-4). Interestingly, for some Cas endoRNases, the triplex was not required for ON-switch functionality. Of note are CasE and the Cas13b family for which the triplex does not make a significant difference, and for Cse3 where no triplex provides a larger dynamic range. This is likely due to the fact that the endoRNases stay bound to their cleaved product and protect the 3' end of the transcript as previously shown [166,167]. However, unlike the other Cas proteins, LwaCas13a and RfxCas13d cleave on the 5' side of their recognition site and would therefore stay bound to the degradation domain side of the transcript. In this case the protein would not be able to protect the 3' end from

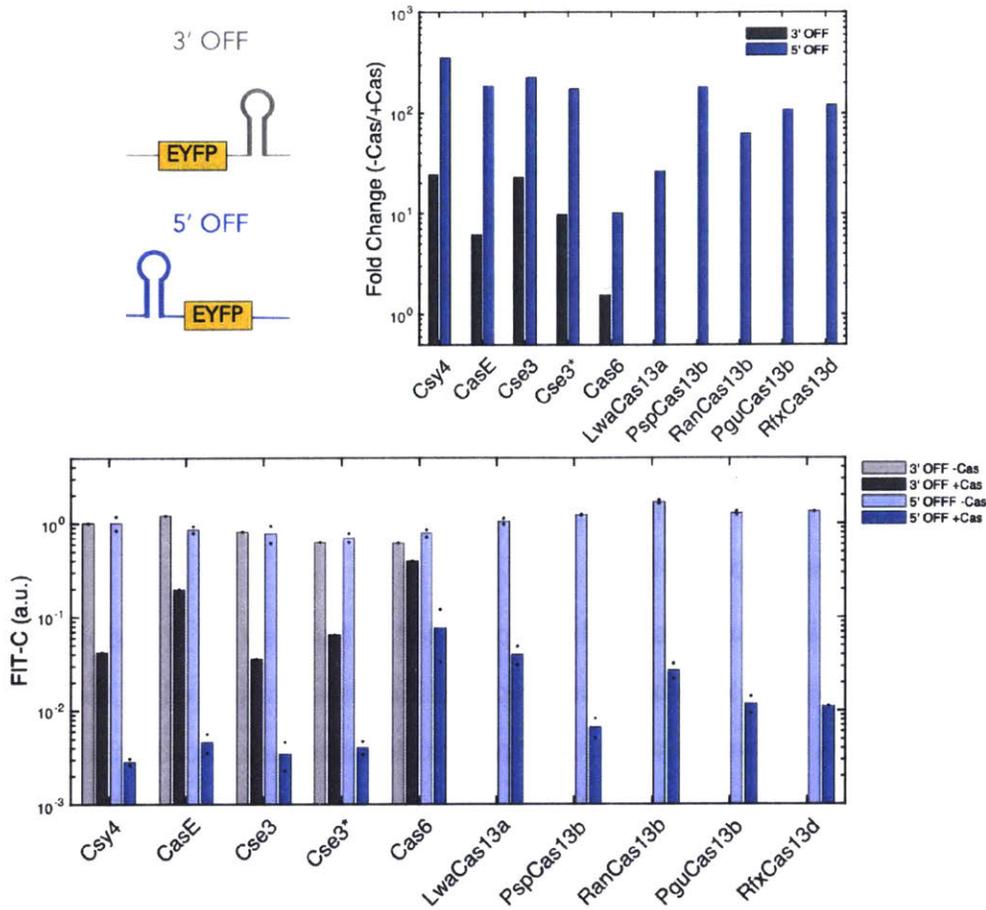


Figure 4-5: PERSIST OFF switch has improved performance in 5' UTR.: OFF-switch EYFP reporters with endoRNase target in the 5' UTR (blue) and 3' UTR (gray) were constructed for each endoRNase. Values for -Cas and +Cas (bottom) are fold change compared to a control with no target site.

degradation which is confirmed by our results where we see a strong requirement for the triplex for these proteins.

These same endoRNases led to robust repression when their target sequences are placed in the transcript outside of the PERSIST-switch, with larger repression conveniently occurring when placed in the 5' UTR of the transcript for those tested (Figure 4-5). As mentioned above, this is likely due to the fact that the majority of the Cas proteins remain bound to the 5' cleavage product which would protect the 3' end of transcripts when placed in the 3' UTR but would leave a free end at the 5' end of transcripts when placed in the 5' UTR allowing for rapid degradation (and likely preventing translation initiation). Other than Pfcas6 the Cas6 family of endonucleases

perform quite well compared to the Cas13 family. The Cas6 family of endonucleases are quite small, less than 30 kDa, which means that they can likely diffuse into the nucleus. We hypothesized that some of the success of these proteins in particular could be attributed to this ability since they are able to process transcripts before they have the chance to get translated in the cytoplasm. The Cas13 endoRNases are quite large, over 100 kDa and all but Cas13d had been published (and tested here) with nuclear export signals. To evaluate whether nuclear localization could improve the performance of these endoRNases we removed the NES tags on Cas13 proteins and placed NLS tags on either end. While nuclear localization did not seem to have a large effect on the OFF switches, it did seem to have a slight beneficial effect for the 3' ON switches (Figure 4-6)

The optimization mentioned above provided a final set of 9 endoRNases that can act as both repressors when their recognition sites are placed in the 5' UTR of transcripts and as activators when their recognition sites are placed in the 3' UTR ON switch position (Figure 4-7). While both the degradation domains and triplex structure are known to function in various cell types we verified that the PERSIST ON-switch responded to the endoRNase CasE in several other relevant cell lines: CHO-K1, HeLa, and Jurkat (Figure 4-8).

In addition to endonucleases we also evaluated whether this PERSIST ON-switch design could be adapted to other types of RNA cleavers. When RNaseP targeted mRNA, HDV ribozyme, or sTRSV ribozyme are placed in the 5' UTR of reporter genes expression is diminished, while when these same ribozymes are inserted into the PERSIST-switch motif we found that expression is restored (Figure 4-9). This could be useful for detecting endogenous ribozymes or to regulate expression based on interactions with endogenous RNAs (discussed in Chapter 5). Note that more studies need to be performed with inactive versions of these ribozymes to confirm that the effect seen is indeed caused by their cleavage.

Of particular interest, next we sought to identify whether the RNA ON-switch motif could be adapted to respond directly to miRNAs. Typically reporters that detect high levels of miRNA (i.e. "miRNA high sensors") have been difficult to design

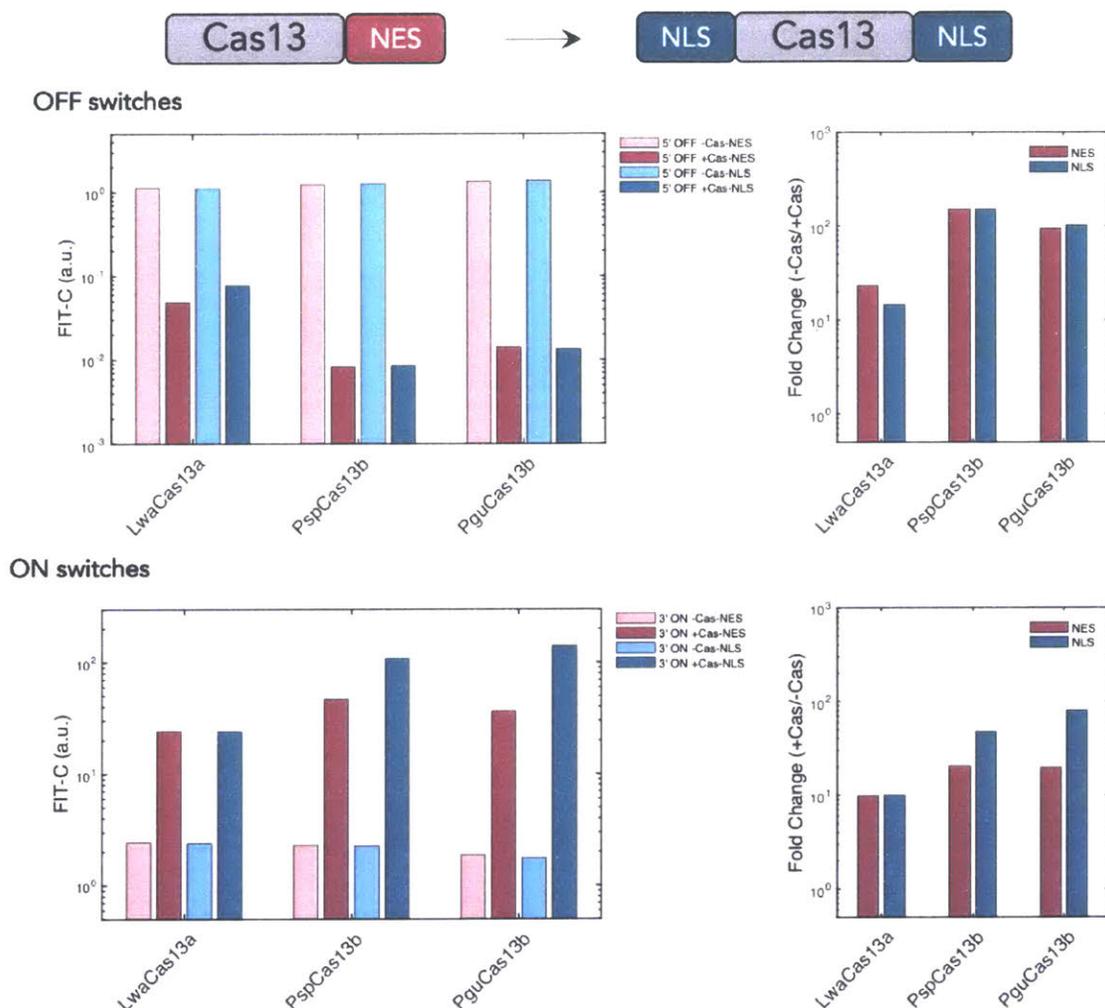


Figure 4-6: Nuclear localized Cas13 endoRNases have improved performance in ON-switch: SV40 nuclear localization tags were cloned onto each end of the Cas13 family of CRISPR endoRNases and evaluated for their 5' OFF-switches and +Triplex ON-switches. Values for -Cas and +Cas (left) are fold change compared to a control with no target site.

and optimize in part because they often require a double-inversion step [8] or multiple components. Recently a positive readout for miRNA was designed based on activation of gRNA but requires additional components: the Cas9 protein and a reporter that can be activated by the gRNA-Cas9 complex [168]. A single-transcript reporter that increases output in direct response to miRNA activity would be very beneficial but has not yet been demonstrated. We modified the PERSIST-switch motif to contain a miRNA target site between the triplex and degradation motifs and evaluated response to several synthetic miRNAs. While modest in performance, reporter expression increased in response to miRNA FF5, while placing miRNA target sites in the 5'

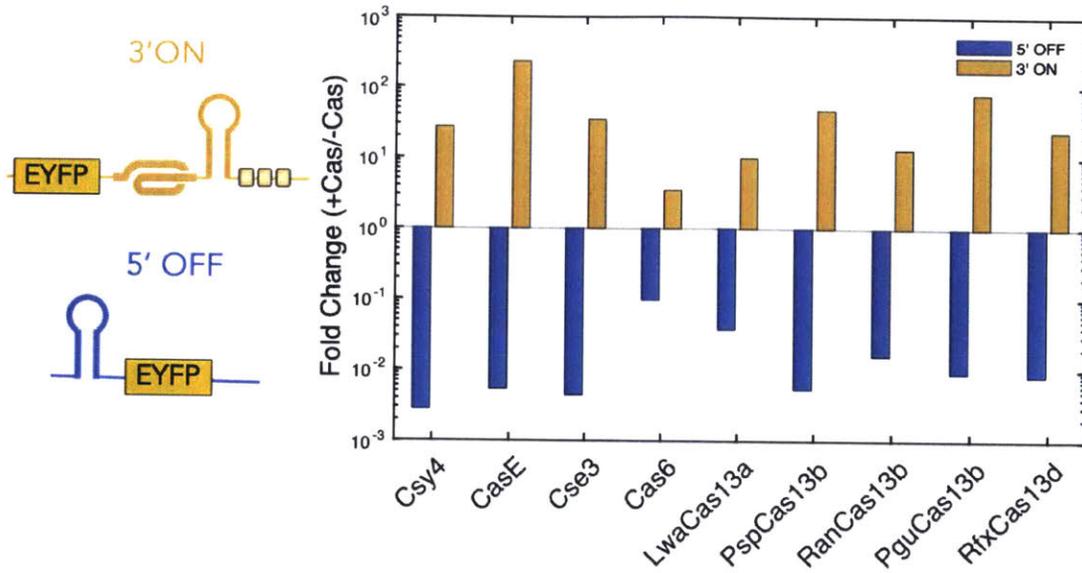


Figure 4-7: 9 CRISPR endoRNases function as activators and repressors: Summary of best-performing variants from switch motif and protein optimizations. Yellow bars show fold activation for each protein added to its respective PERSIST ON-switch while blue bars show fold repression for each protein added to its respective PERSIST OFF-switch

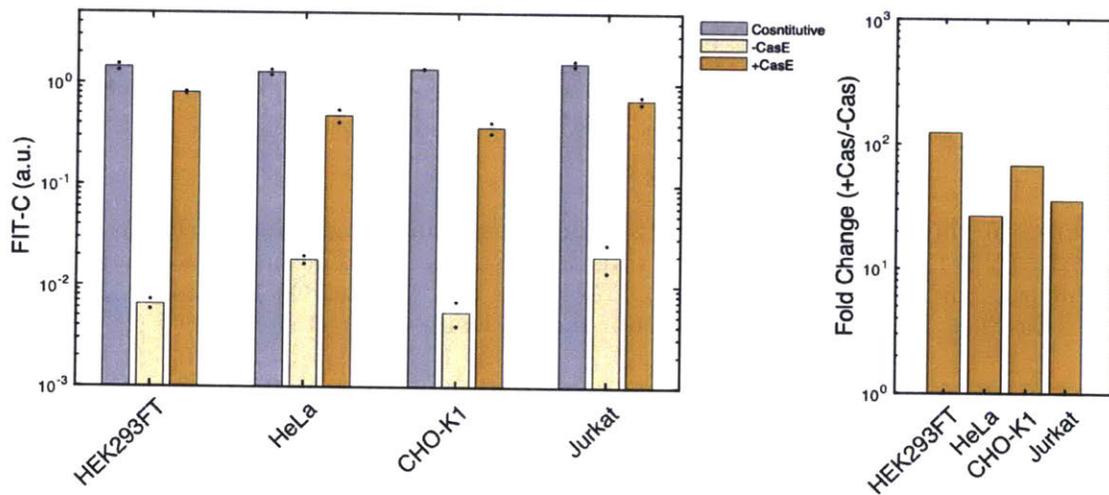


Figure 4-8: PERSIST platform functions in a variety of cell types: A CasE-responsive PERSIST ON-switch EYFP reporter was transfected into various cell lines with and without a plasmid encoding CasE (light yellow and dark yellow respectively). All cell lines show robust degradation of the ON-switch plasmid in the absence of endoRNase and activation of the plasmid close to constitutive values (gray) in the presence of the endoRNase.

UTR canonically decreased expression as has been shown previously [128] (Figure 4-10). While adding multiple target sites in the 5' improves repression, 1 target site works best for persist switch (data not shown).

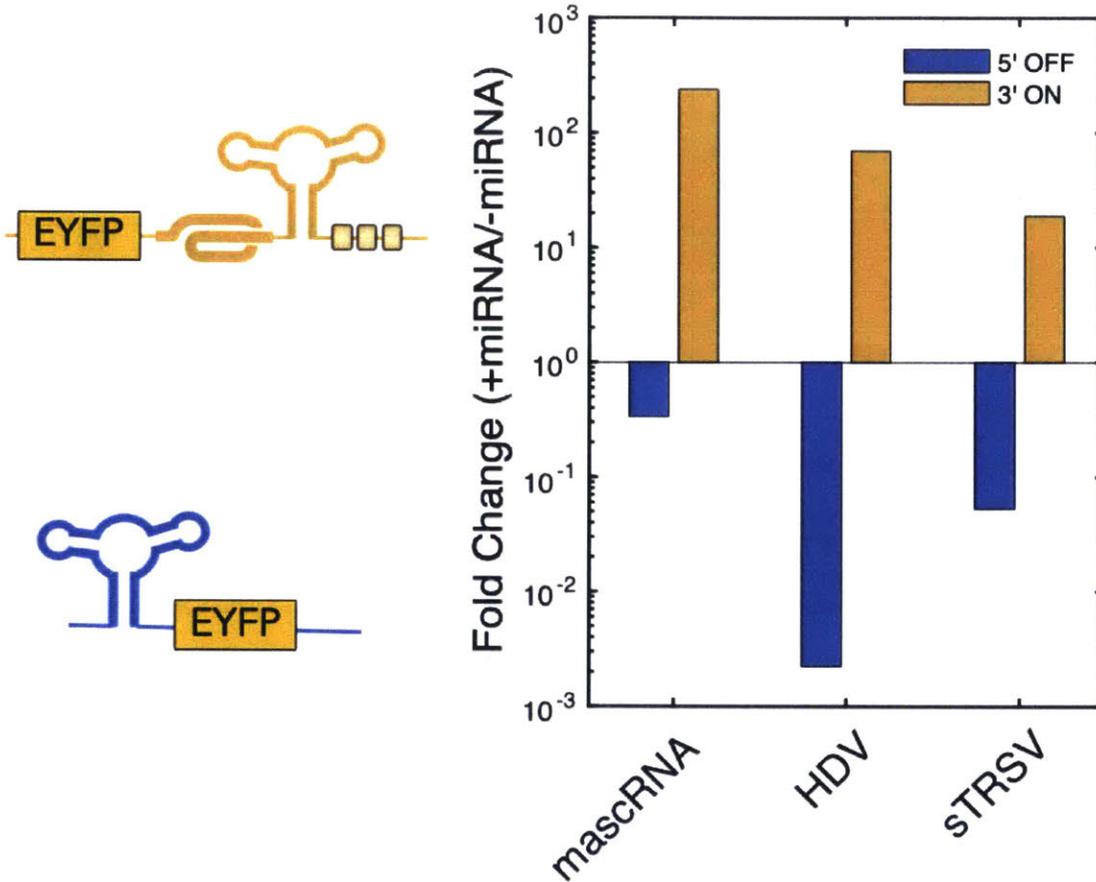


Figure 4-9: Ribozymes repress in OFF-switch but activate transcripts in ON-switch: Ribozymes were inserted into either the 5' PERSIST OFF-switch position or within the 3' PERSIST on switch. Yellow values show fold activation compared to negative ON-switch control with no ribozyme while blue bars show fold repression compared to constitutive control with no ribozyme.

Finally, an important claim of this PERSIST platform is that regulation can be programmed at the RNA level in order to facilitate constitutive transcription to combat epigenetic silencing. In order to evaluate this statement, we integrated the PERSIST RNA ON-switch and two transcriptional ON-switches into CHO-K1 cells to compare their response over time. We chose to evaluate both an activator-based ON-switch (the Tet-On system) and a repressor based ON-switch (the PhIF system). In the Tet-On system, rtTA3 is induced by doxycycline to activate expression at the TRE promoter. In the PhIF system, the transcriptional repression of PhIF is released through addition of small molecule DAPG and the gene can then be expressed through Gal4-VP16-based gene activation. To isolate any differences in response to the expres-

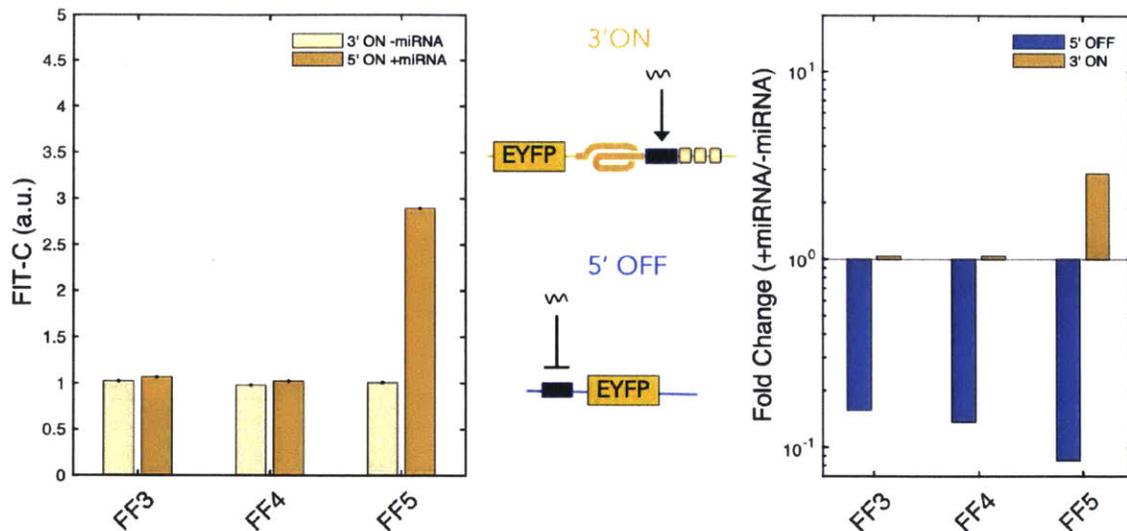


Figure 4-10: PERSIST switch can function as a single-transcript miRNA sensor: PERSIST ON and OFF switches were constructed to contain 1x miRNA recognition sites for miRFF3, miRFF4, and miRFF5. Values for -miRNA and +miRNA (left) are fold change compared to a control with no target site.

sion of the reporter itself, we created plasmids that contained all required parts for all three switches and only varied the last transcription unit and included a constitutive positive control plasmid (Figure 4-11a). After antibiotic selection, cells were induced (either with small molecule or transfected with Csy4) and sorted on +EBFP and +mKO2 fluorescence. Figure 4-11 shows that following sorting, the PERSIST ON-switch responded quite well compared to the transcriptional ON-switches giving us confidence that the system functions at a single-copy level. Tracking these integrated cells over time will help reveal whether constitutive transcription as exhibited in the constitutive case, the PERSIST ON-switch, and the continuously induced transcriptional ON-switches exhibit prolonged response rates compared to the transcriptional ON-switches that remain in the OFF-state but are transiently induced.

Overall, the PERSIST switch platform encodes the ability to respond both positively and negatively to RNA cleavers based on recognition site position. The dynamic range, robustness and adaptability for different types of inputs makes it promising for use in synthetic biology applications.

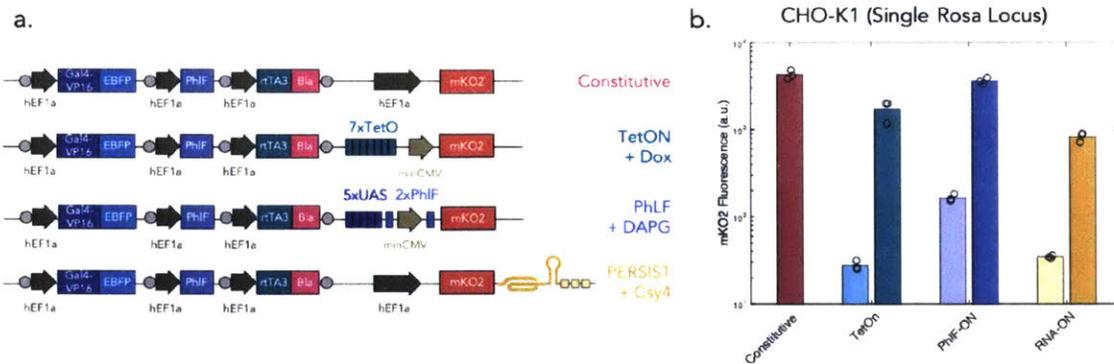


Figure 4-11: PERSIST integrated response: (a) Plasmids each containing four transcription units were constructed using MoClo and custom Golgen Gate assembly strategies. For all four types of plasmids the first three transcription units were kept constant: (1) Gal4-VP16 2A-linked to EBFP, (2) PhIF, and (3) rtTA3 2A-linked to Blasticidin. The fourth transcription unit encoded mKO2 under control of each regulatory mechanism: Constitutive, TetON, PhIF, and PERSIST ON switch. (b) Integrated fluorescence for the 'OFF' and 'ON' state of each system. N=3

4.2 Combining RNA OFF and ON switches for basic logic

The impressive responses to CRISPR-endoRNases in both the 5' UTR "OFF"-switch and 3' UTR "ON"-switch positions gave us confidence that these proteins could be used as RNA-level activators and repressors for programmable gene circuits. In fact, the recognition sites function similarly to transcription factor operons, except the repressor/activator functionality of the endoRNases is determined by the location of their cognate recognition sites in the transcript rather than encoded in the protein itself.

To demonstrate the utility of such a platform, we first sought to evaluate the orthogonality, modularity and composability of the Cas-based PERSIST-switches. In particular, in order to be a platform that could replace transcription-factor-based logic processing the endoRNase platform should have several characteristics: (1) the Cas proteins should be orthogonal to each other, or in other words have little cross-talk with each other's recognition sites, (2) the Cas protein-recognition site pairs should be modular where recognition sites can be placed in either the 5' or 3' PERSIST switch locations (in any order or combination) and have predictable behavior and (3) the regulation enabled by the endoRNases should be composable, which means that it

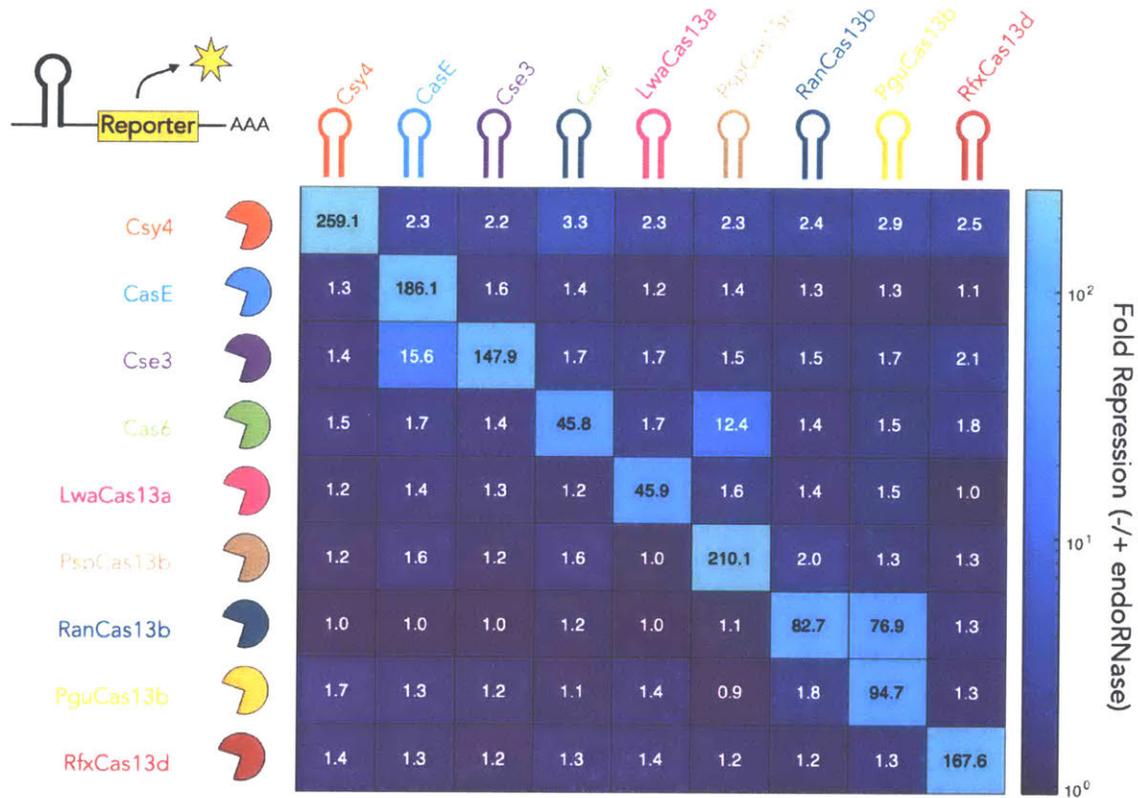


Figure 4-12: Orthogonality matrix for endoRNases: Fold repression was calculated as done previously (Figure 4-5) for each combination of endoRNase and 5' PERSIST OFF-switch.

should be possible to replace the reporter gene used previously with any endoRNase to create layered circuits.

4.2.1 Evaluating orthogonality

We first evaluated the orthogonality of the Cas proteins using poly-transfection to test every combination of Cas protein and recognition site. Reporters were constructed with each recognition site in the 5' UTR of an eYFP-expressing plasmid. All nine recognition sites were tested along with a mutated version of the Cse3 recognition site [162]. As seen in Figure 4-12, the endoRNases cleave quite orthogonally, and fortunately the mutated Cse3(U5A) recognition site is cleavable only by Cse3 and not CasE. The orthogonality shown here, while not surprising, makes these set of proteins usable within the same circuit. Of note some pairs should be avoided (unless

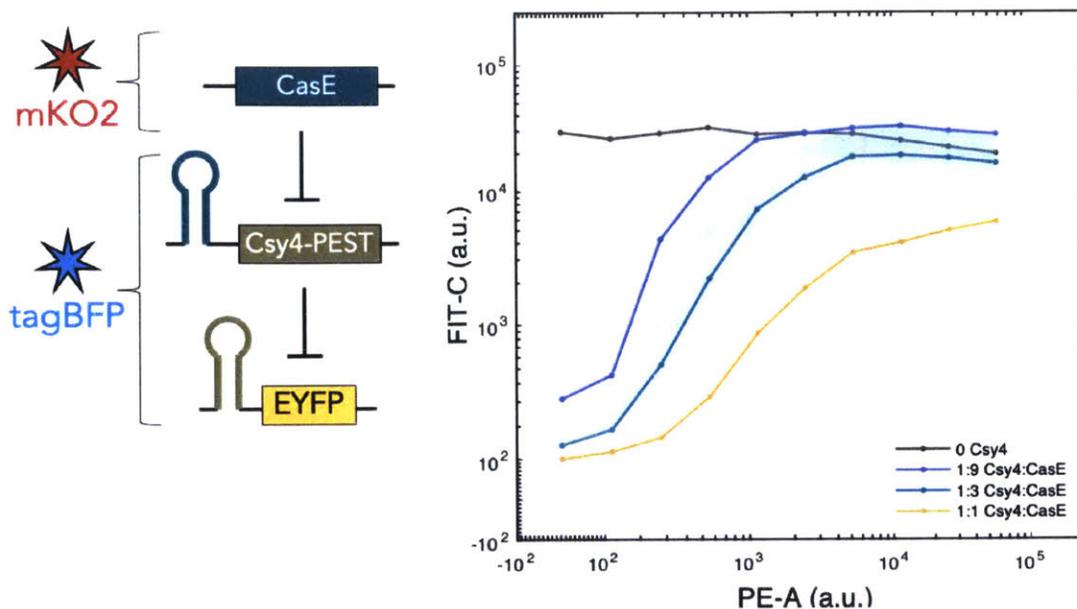


Figure 4-13: EndoRNase repression of an endoRNase: A simple repression cascade was constructed where CasE \dashv Csy4 \dashv EYFP. The EYFP reporter was co-transfected with varying levels of Csy4 with the amount tracked with a constitutive tagBFP plasmid. The amount of CasE was poly-transfected in a separate complex with its level tracked by mKO2 fluorescence. Shaded areas show the interquartile range of each bin and solid lines show medians for each bin.

beneficial for circuit design) which include RanCas13b:PguCas13b and CasE:Cse3 (with the wt Cse3 recognition site).

4.2.2 Evaluating composability: cascades and positive feedback

Next we evaluated the composability of the PERSIST platform by layering the endonucleases, or in other words, engineering constructs that allowed for endoRNases to regulate other endoRNases. This was accomplished by swapping the reporter gene (EYFP) for genes encoding one of the endoRNases, thus creating cascades. In Figure 4-13 we demonstrate endoRNase repression of another endoRNase (that in turn represses a reporter). In the absence of all endoRNases, EYFP is strongly expressed (gray line). When Csy4 is added in the absence of CasE, the reporter is repressed strongly. However when CasE is added, it represses Csy4 and expression of EYFP is restored. Note that as more Csy4 is added, CasE is oversaturated and cannot fully

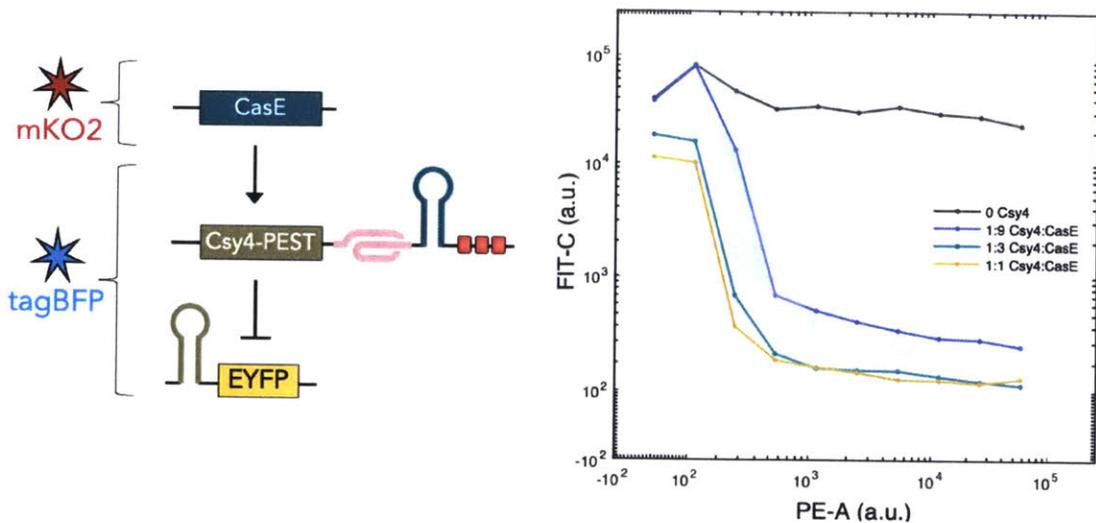


Figure 4-14: EndoRNase activation of an endoRNase: A simple activation cascade was constructed where CasE \rightarrow Csy4 \dashv EYFP. The EYFP reporter was co-transfected with varying levels of Csy4 with the amount tracked with a constitutive tagBFP plasmid. The amount of CasE was poly-transfected in a separate complex with its level tracked by mKO2 fluorescence. Shaded areas show the interquartile range of each bin and solid lines show medians for each bin.

repress Csy4 and therefore full fluorescence levels cannot be restored. While other post-transcriptional platforms, namely protease-mediated regulation [146–148] have show repression of another repressor as we have above, none have demonstrated direct activation of a repressor. In Figure 4-14 we demonstrate endoRNase activation of another endoRNase. In this cascade, Csy4 contains a PERSIST ON-switch tag which quickly degrades its transcript and prevents it from targeting the reporter. When CasE is added, it activates the Csy4 transcript which can then repress fluorescent output expression. In the complete absence of Csy4 plasmid the reporter is expressed strongly (Figure 4-14, gray line). At low CasE and low Csy4 levels, Csy4 transcript is degraded quickly and is unable to repress the output. As the amount of CasE increases, so does the amount of Csy4 and thus we see a decrease in reporter fluorescence. We show the robustness and further composability of the platform by creating slightly more complex systems. Figure 4-15 shows a positive feedback motif where the endoRNase is able to self activate itself when its recognition site is placed in the 3' ON-switch. Note that the level of output fluorescence corresponds closely with the constitutively expressed version of the construct. We also build a three stage

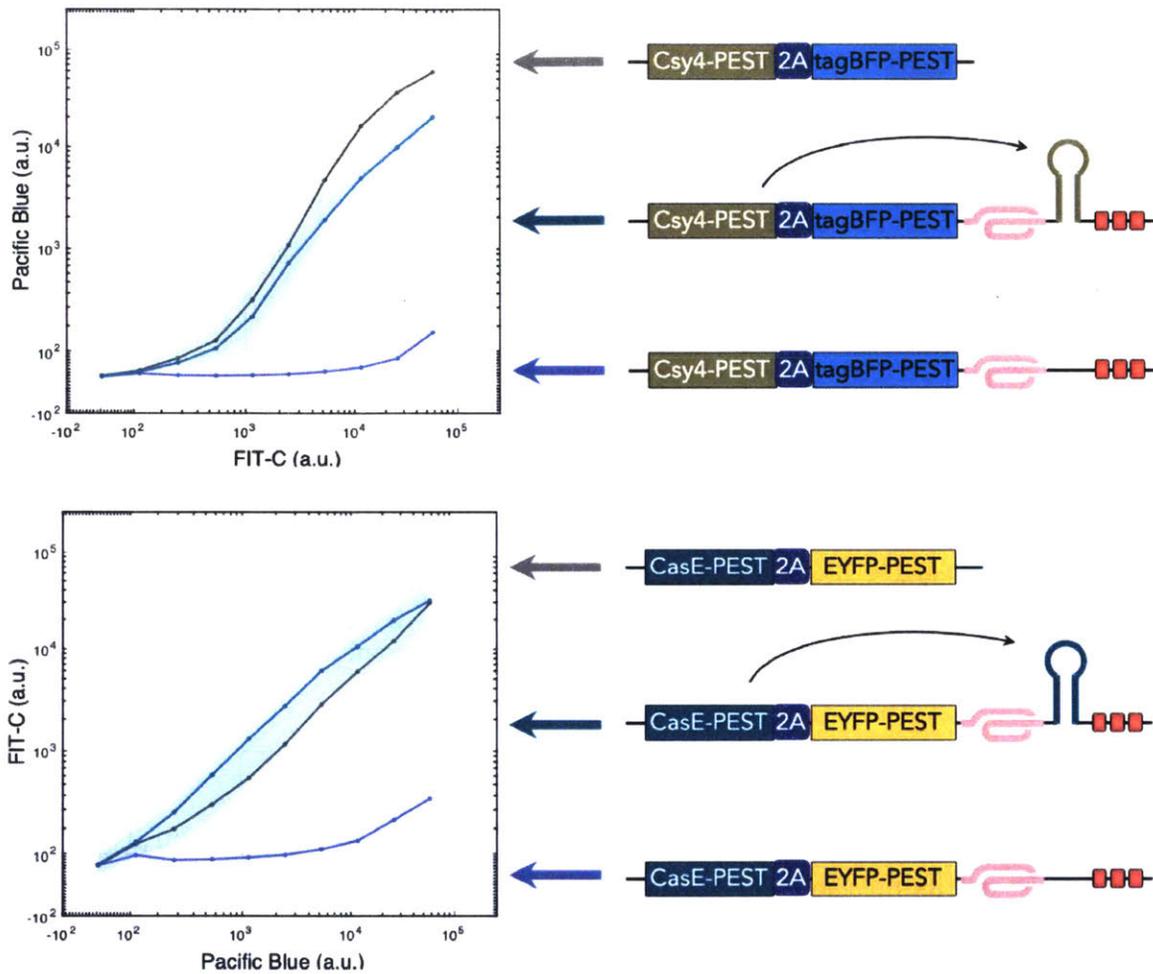


Figure 4-15: Single-node positive feedback: Positive feedback loops can be easily created with the PERSIST platform when endoRNase proteins can activate their own transcripts via the PERSIST ON-switch. To track the positive feedback, CasE and Csy4 were tagged with EYFP and tagBFP respectively and co-transfected with transfection markers encoding tagBFP and EYFP respectively. Shaded areas show the interquartile range of each bin and solid lines show medians for each bin.

repression cascade (Figure 4-16). While there is indeed a drop in signal restoration at stage 2, this cascade outperforms previous RNA-based repression cascades and further optimization as in Figure 4-13 above can be employed to improve performance.

4.2.3 Evaluating modularity: All 16 two-input logic functions

Finally, we demonstrate the modularity of the platform by combining 5' OFF and 3' ON switches on the same transcripts in order to create all 16 two-input logic

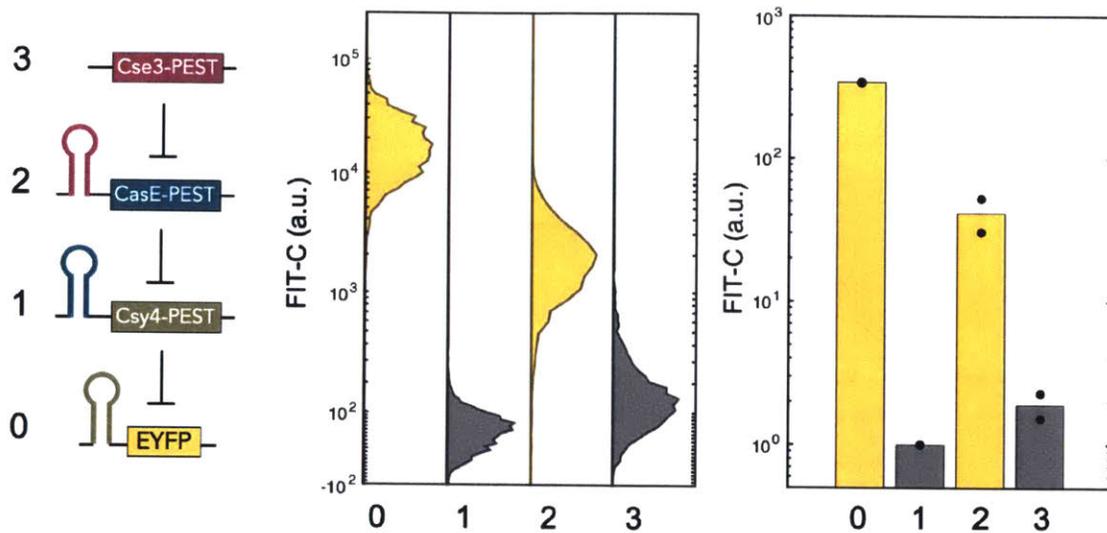


Figure 4-16: Three-stage repression cascade: A 3-stage repression cascade was created where Cse3 \dashv CasE \dashv Csy4 \dashv EYFP (left). Histograms of highly-transfected cells show good separation between ON and OFF states (middle). Bar chart shows slopes of linear fits normalized to Stage 1 value.

functions (Figure 4-17). Importantly, multiple recognition sites can be combined within the same 5' OFF or 3' ON PERSIST motif (see "NOR" and "OR") or on separate motifs on the same transcript (see "NIMPLY"). In fact 10 of the 16 logic gates can be constructed with a single transcript and they all behave quite robustly with dynamic ranges approaching two orders of magnitude. Overall, these logic gates compare well to previous post-transcriptional platforms as well as DNA-based logic gates [169–171].

4.3 Dual functionality allows for complex motifs to be easily engineered

Importantly, unlike transcription factors in DNA logic-based systems, which function to either repress or activate transcription with mutual exclusivity, these endoRNases can act as both "activators" and "repressors" just by changing the location of their recognition sequence (Figure 4-18). There are several transcriptional systems that have devised clever ways of switching between activator and repressor, for example,

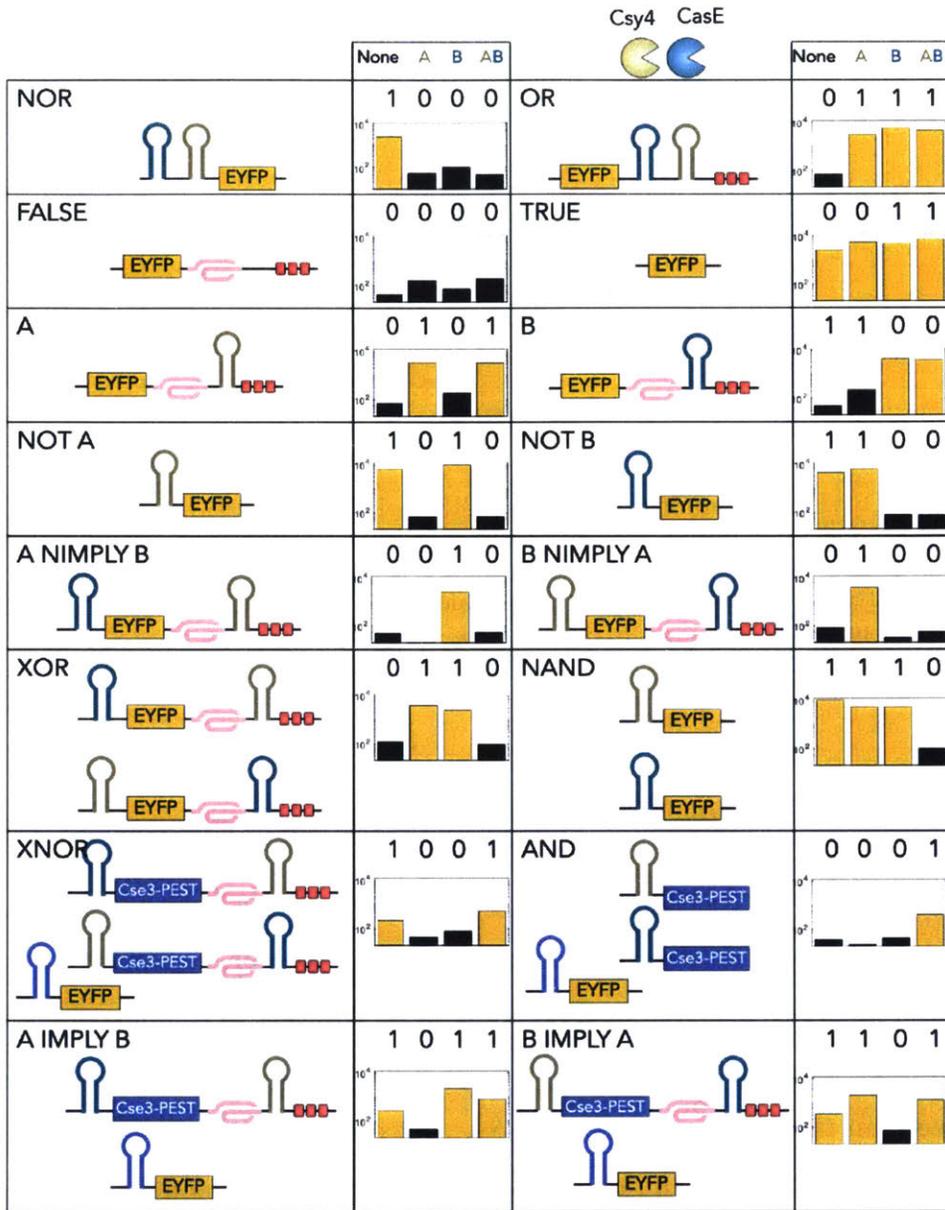


Figure 4-17: All 16 two-input logic functions: All 16 two-input logic functions were constructed using Csy4 and CasE as inputs and Cse3 as an intermediate where necessary. Logic gate constructs were co-transfected with the transfection marker tagBFP, and then separate complexes containing Csy4 with mKO2 and CasE with iRFP720 were poly-transfected for each logic gate sample. Data was gated on +tagBFP and then separated into four subpopulations: -mKO2/-iRFP720 = 'None', +mKO2/-iRFP720 = 'A', -mKO2/+iRFP720 = 'B', and +mKO2/+iRFP720 = 'AB'. Negative fluorescence cutoffs (-mKO2, -iRFP720) were calculated as 99th percentile of fluorescence in each channel for the non-transfected control and positive fluorescence cutoffs (+mKO2,+iRFP720) are 970 AU and 26400 AU respectively. Median EYFP fluorescence was measured for each subpopulation.

Kiana et al [25] created a Cas9-based system where the protein has either repressive or activation functions based on the length of the gRNA. However even still in these

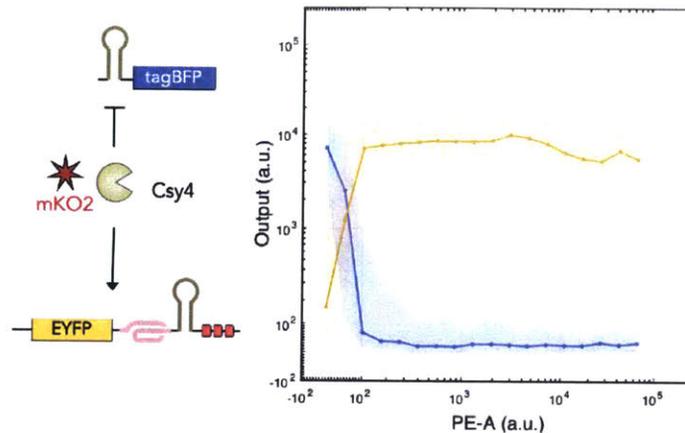


Figure 4-18: Dual functionality of Cas PERSIST platform: A PERSIST ON-switch reporter encoding EYFP and PERSIST OFF-switch reporter encoding tagBFP were co-transfected with the transfection marker iRFP720 and a separate complex was poly-transfected containing Csy4 with mKO2 plasmid. Shaded areas show the interquartile range of each bin for tagBFP and EYFP and solid lines show medians.

systems the function must be encoded in the regulator. In the PERSIST platform, the exact same protein can act as both an activator and repressor. In other words, a given endoRNase will act as a repressor if its recognition site is placed in the 5' OFF switch position while the same endoRNase will function as an activator if its recognition site is placed in the 3' ON switch position. Not only are these endoRNases capable of both functions but they can exhibit this dual functionality simultaneously, which makes some interesting and useful motifs quite easy to build. For example, as shown in Figure 4-18 both a reporter encoding a Csy4-repressible tagBFP transcript and a reporter encoding a Csy4-activatable EYFP transcript can be acted on at the same time in the same cells by Csy4. In fact several motifs that would take several nodes or protein engineering of multi-functional transcription factor modules to accomplish with transcription factors can be accomplished with a single PERSIST node. We took advantage of this feature of the PERSIST platform to build several motifs including a feed-forward loop (FFL), a single-node positive feedback/repression motif, and a two-node bistable switch.

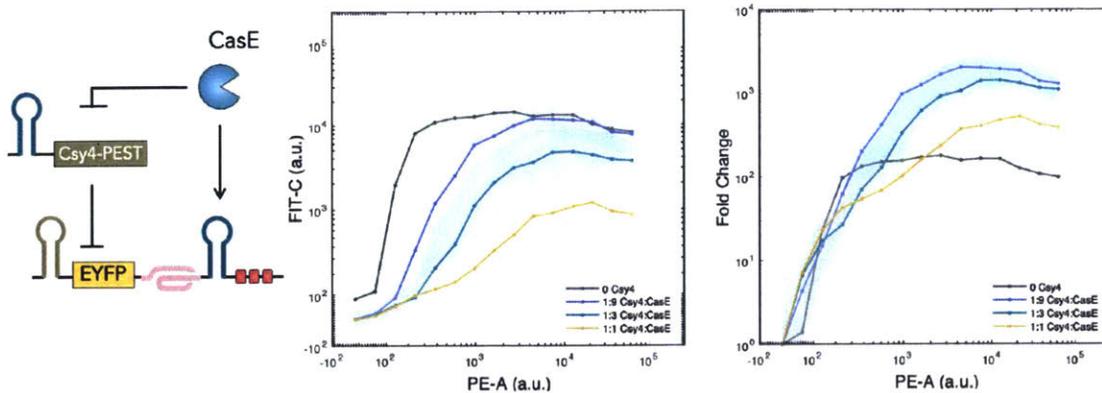


Figure 4-19: FFL: A coherent feed-forward loop was constructed where $\text{CasE} \dashv \text{Csy4} \dashv \text{EYFP}$ and $\text{CasE} \rightarrow \text{EYFP}$. The EYFP reporter was co-transfected with varying levels of Csy4 with the amount of transfection of both plasmids tracked with a constitutive tagBFP plasmid. The amount of CasE was poly-transfected in a separate complex with its level tracked by mKO2 fluorescence. Shaded areas show the interquartile range of each bin and solid lines show medians for each bin. Right plot shows values normalized to lowest mKO2 (-CasE) bin to evaluate dynamic range of FFL ON-switch.

4.3.1 Feed-forward loop

As observed in Figure 4-2 while the "OFF" state of the PERSIST ON switch is indeed low it still has some leaky, well above-zero expression at high transfection efficiency which limits the dynamic range of the switch. We hypothesized that one way to decrease the expression of the leaky OFF state while maintaining responsiveness to a given endoRNase is through an FFL: a second endoRNase could further repress the ON switch to decrease background fluorescence which could itself be repressed by the input endoRNase. Importantly, this FFL is possible because the input endoRNase is able to act as an activator for the reporter and as well as a repressor for second endoRNase. When the input endoRNase is absent, the reporter is both degraded by the ON switch motif and repressed by the second endoRNase resulting in very low expression. When the input endoRNase is present the reporter is activated by the endoRNase-responsive ON switch and the second endoRNase is repressed resulting in output derepression. The FFL was constructed with CasE as the input endoRNase and Csy4 as the intermediate second endoRNase (Figure 4-19a). Additionally a PEST tag was added to Csy4 to increase its degradation rate which would enable efficient derepression by CasE. As seen in Figure 4-19b, compared to no Csy4, there is a

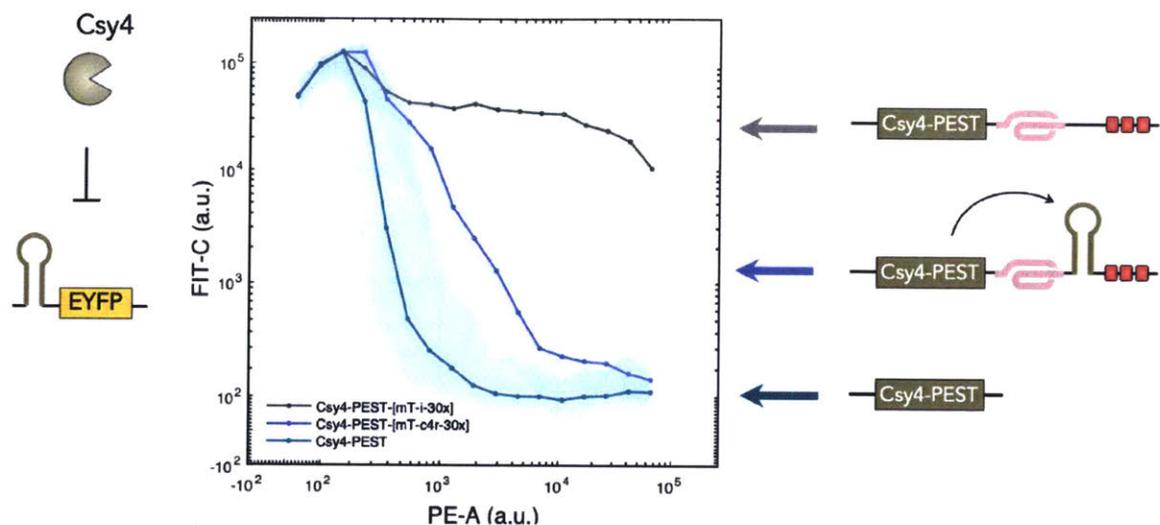


Figure 4-20: Positive feedback + repression motif: A single-node positive feedback + repression motif was constructed where Csy4 \dashv EYFP but where the PEST-tagged Csy4 is either constitutively expressed (green), constitutively degraded by its own RNA degradation tags (gray), or can activate itself through cleaving off the degradation tags (blue). The EYFP reporter was co-transfected with a constitutive tagBFP plasmid transfected marker. The CasE plasmid was poly-transfected in a separate complex with its level tracked by mKO2 fluorescence. Shaded areas show the interquartile range of EYFP in each mKO2 bin and solid lines show medians for each bin.

decrease in the OFF state. When low amounts of Csy4 are added, the motif's OFF state is close to zero while high CasE amounts enable signal restoration close to the ON state of the original ON switch resulting in dynamic ranges well over 1000 fold change. As expected, when increasing amounts of Csy4 plasmid is added to the system the ON switch is unable to recover full fluorescence. Overall the FLL indeed represents a way to increase the dynamic range of the ON switch, especially in applications where extremely low background expression is necessary.

4.3.2 Single-node positive feedback + repression

Another interesting motif that is enabled by the dual functionality of this platform is a single-node positive feedback + repression motif (Figure 4-20). In this motif, an endoRNase is able to both activate itself and repress another transcript, in this case, a reporter. We constructed such a motif using Csy4 and an EYFP OFF-switch reporter containing the Csy4 recognition site in its 5' UTR. When Csy4 functions as just a repressor (without an ON switch motif), it is expressed stably and greatly

represses the reporter even at low values. When Csy4 repressor contains the ON switch degradation tags but is unable to self-activate itself (the ON switch does not contain a recognition site), the Csy4 transcript is degraded resulting in poor repression of the reporter. But finally, when the Csy4 recognition site is placed in the ON switch motif, Csy4 is able to activate itself through a positive feedback loop and recover its ability to target and repress the reporter.

4.3.3 Construction of an endoRNase-based bistable switch

The response shown in Figure 4-20 is actually ultrasensitive. This is especially exciting because many complex circuits such as binary, oscillators and toggle switches require ultrasensitive responses [172, 173]. With transcriptional systems, ultrasensitivity, or cooperativity, is often accomplished due to the fact that transcription factors must often multimerize to function. However, previous studies suggest that these endoRNases do not multimerize [160], so the ultrasensitivity observed here could serve as a building block when ultrasensitivity is required. A useful circuit that requires ultrasensitivity is a bistable toggle switch. The first synthetic toggle switch was built in bacteria [172] in 2000 and propelled the field of synthetic biology. However toggle switches in mammalian systems have been underwhelming. Here, we first construct a bistable toggle switch using two different endoRNases as nodes. These endoRNases can each activate themselves as well as repress the other endoRNase (Figure 4-21a). Separate constructs containing fluorescent proteins (mKO2 and EYFP) that can be repressed by each of the endoRNases were constructed to track cell state. As a first performance test, transfection of these constructs uninduced should lead to two states: a high-mKO2/low-EYFP state and low-mKO2/high-EYFP state with few cells expressing both. Unlike previous studies where uninduced toggles express both fluorescent proteins ([61] and Appendix B.1 this feature would give confidence that the system is indeed bistable. The system was evaluate by poly-transfection to allow for many ratios of each node to be sampled. Promisingly, despite the wide range of ratios samples, only 3% of cells expressed both fluorescent proteins hinting that this system is indeed bistable. The system needs to be further evaluated for its

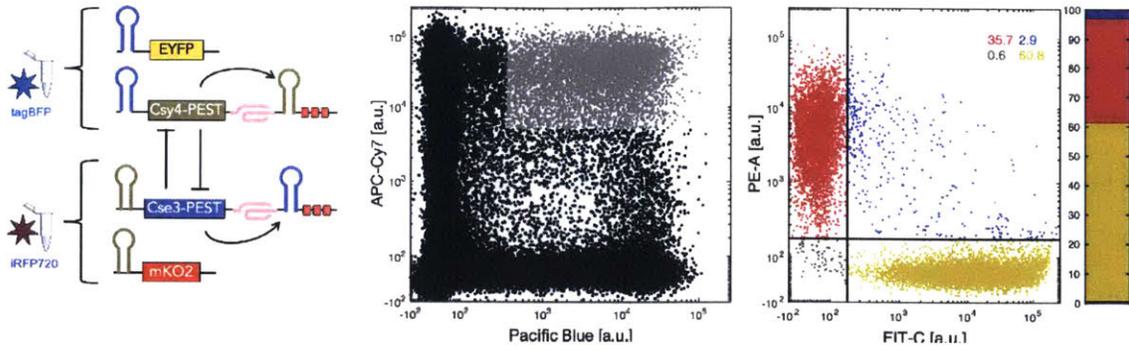


Figure 4-21: PERSIST-based bistable toggle switch: A toggle switch was constructed using the same single-node positive feedback + repression motifs in Figure 4-20 where the endoRNases were also able to mutually repress each other. Separate reporters (EYFP and mKO2) were transfected to track the cell state. Each arm of the toggle was poly-transfected containing it's own transfection marker. High transfection values for both arms of the toggle were evaluate (middle) and plotted in the scatter plot on the right. Cutoff value was chosen as 200 AU for EYFP and mKO2. Percentage of the evaluate population is plotted in the bar on the far right.

ability to maintain memory and ability to switch between states. If so, this toggle could serve as a useful tool for studying cell fate decision making or development of cell type classifiers.

4.4 Discussion and Future Directions

Overall the PERSIST platform makes use of cleavage as a method of regulating RNA for synthetic biology applications. While we adapt 9 CRISPR-specific endonucleases here, we are excited by the scalability of this platform to make use of other CRISPR-specific endonucleases that continue to be discovered [149, 174].

A worthwhile effort would be to evaluate and engineer the PERSIST ON-switch as a positive readout for endogenous cleaver detection. Several strategies could be undertaken to improve the performance of the switch as an miRNA sensor. These include improving the design of the miRNA itself to favor cleaving of transcripts, improving the design of the target site to favor cleavage [175], or a positive feedback strategy. The switch could also be used to detect endogenous disease-specific endonucleases such as IRE1- α which is expressed under cell stress conditions [176] Finally, as described in Chapter 5 the ON-switch could serve as a positive readout for ribozyme activation which could be useful for strand displacement-based mRNA

sensing applications.

Another next step would be to adapt endoRNases in platforms that currently use transcription factor output in order to ensure the PERSIST platform's compatibility. For instance internal protein sensors [177] or external protein sensors/receptors such SynNotch [56] that currently utilize transcription factor output could be adopted to release an endonuclease instead.

And finally it would be quite useful to engineer these endoRNases to be inducible (especially via small molecule). The ability to control the activity of the endoRNases remotely would ensure proper control of circuits built with the PERSIST switch platform and would facilitate the design and characterization of useful time-varying behaviors.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

mRNA sensing in mammalian cells using RNA Strand Displacement

An important challenge for therapeutic synthetic biology not yet mentioned here is that the footprint required for gene circuits can inhibit circuit delivery. Genetic encoding of multi-gene circuits requires gene cassettes that easily exceed several ten thousand bases. For example, the 6-input HeLa classifier circuit [178] is around 45kb. Efficient delivery of a 45kb payload is difficult in cell culture and almost impossible in animals. Current gene therapy efforts have been able to take advantage of the low inflammatory potential, high tropism, and high infection efficiency of AAV delivery [179]. In fact, AAV delivery has proven to be so efficient and its tropism so broad that groups have turned to creating reliable tumor-targeting AAV vectors [180]. However the packaging capacity of AAV is limited to about 4.5kb, which makes it an infeasible option for most transcriptional or translation circuits. Even the truncated Cas9 protein from *Staphylococcus aureus* (3.3kb) [181], which is a helpful decrease in size from the traditionally used Cas9 (4.2kb) [105,182], leaves very little room for any other protein such as a therapeutic output. Finally, there has been a large interest to move away from DNA-encoded devices due to the risk of genomic integration as well as barriers involved with cellular import and genetic silencing. Therefore, for certain clinical applications, it will be necessary to use a platform that avoids DNA encoding or interaction entirely and also avoids using proteins for computational modules due

to their large footprint.

It is possible to address all three of these challenges with RNA strand displacement-based circuits, which are highly designable, have an extremely small footprint, and use only RNA-RNA interactions for information processing. Strand displacement is a competitive hybridization reaction whereby an incoming “input” nucleic acid strand can displace a previously bound “output” from a complementary binding partner (Figure 5-1). Importantly, this reaction is initiated at a short complementary single-stranded region called the “toehold”. In this scenario, the input completely displaces the entire sequence of the output strand, therefore the last step is considered to be essentially irreversible. However, when an additional toehold is left at the departing end, the reaction becomes reversible and is called “toehold exchange” (Figure 5-1). The output strand can now act as an input strand for another strand displacement reaction now that its toehold is exposed. Also, leaving a toehold at the departing end allows for signal amplification using a “fuel” strand. The fuel displaces the original input strand leaving it free to bind to other strand displacement gates. Additionally a “threshold” strand can be used which is complementary to the input strand and acts as a sequestering agent. The combination of the fuel and threshold allow for the programming of digital behavior.

Such reactions have been used with DNA in cell-free settings to create complex multi-input logic circuits [183, 184], catalytic amplification mechanisms [185–187], synthetic molecular motors [188–191], and programmable nanostructures [192]. More recent demonstrations of neural networks [193] and logic circuits [183] using tens of logic gates and hundreds of oligonucleotides demonstrate the impressiveness of this mechanism. There are three main reasons for the remarkable success of strand displacement as an engineering technology in cell-free settings: first, a single basic building block, strand displacement, underlies every operation. The biophysics of this reaction is quantitatively understood since it relies solely on the interaction of nucleic acid strands. Second, we can create an essentially unlimited number of orthogonal sequences, leading to large libraries of orthogonal parts each with a very small nucleotide footprint. And third, any digital or analog circuit can in principle be

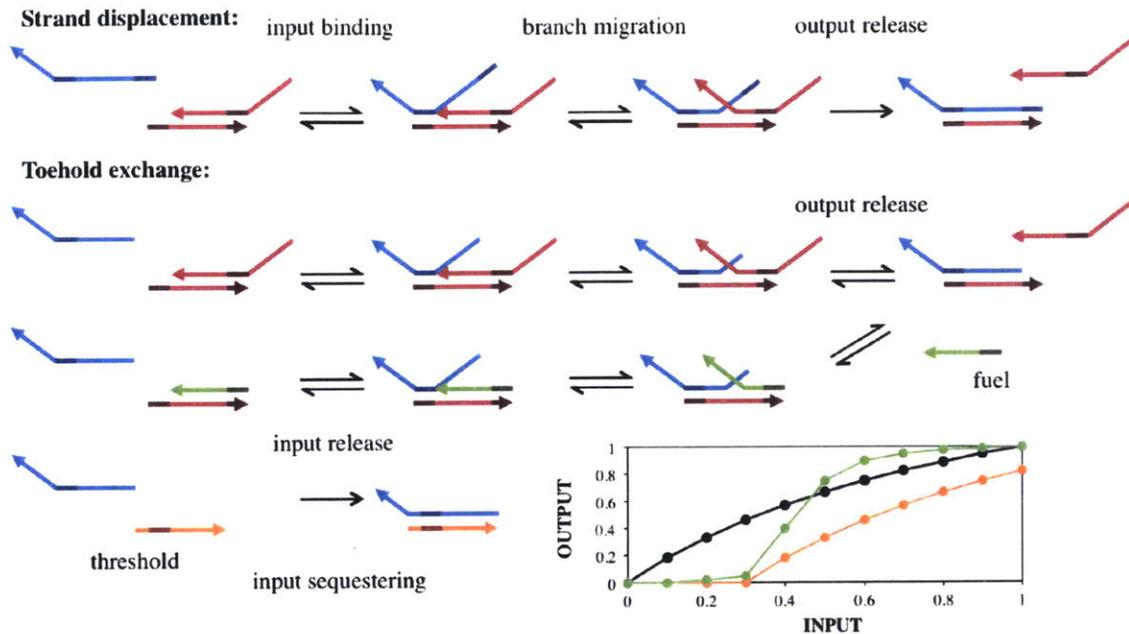


Figure 5-1: Strand Displacement mechanism: Strand displacement is a form of competitive hybridization and provides the mechanism underlying all sensors, gates and actuators used here. The input strand (blue) can be cellular RNA or output of a different gate. When the input extends the entire length of the gate, release of the output (red) is essentially irreversible. When the input does not extend the entire length of the gate, a toehold is left on the departing end and the reaction is called “toehold exchange”. The output now has a toehold domain to affect expression of a reporter gene or interact with the next gate. Importantly the toehold resulting from a shorter input allows the fuel mechanism to be used, where a complementary fuel strand can displace the original input from the gate only after strand displacement has occurred. This input can now interact with another gate which leads to the release of many output strands per input strand and thus an amplification of the signal. Additionally, a threshold mechanism can be used which acts to sequester input strands. Input strands will preferentially bind the threshold so that only input strands above the number of threshold strands will undergo strand displacement. The addition of the threshold and fuel strands makes this reaction much more digital

implemented with DNA strand displacement components, making it easy to go from a desired function to a molecular implementation. We hope to harness this technology as a synthetic biology platform for performing computation in mammalian cells, which has not yet been demonstrated. The strand displacement project is divided into three sectors: sensing, processing, and actuation. Sensing refers to the building of strand displacement sensors for the quantifiable detection of cellular RNAs. Processing is the construction of intermediate reactions, often involving the fuel and threshold mentioned above, used to evaluate the input signals. Actuation is the production of some output product that can then interact with the cell. While this

thesis chapter will touch on each of these areas, the primary focus is the predictable detection of cellular RNA transcripts.

5.1 FRET-based mRNA sensing

The FRET-based strand displacement reporter was originally designed by the Seelig lab at University of Washington [194]. It is a double-stranded RNA duplex, where one end has a short, 6 base pair, single-stranded toehold region for binding target input strands. On this same end, the other strand is functionalized with an Alexa Fluor 488 dye (IDT) to use as a transfection marker for the reporter. At the other end of the duplex, one strand is functionalized with a ROX dye (IDT) and the other strand is functionalized with an Iowa Black quencher (IDT). This quencher suppresses ROX fluorescence until strand displacement occurs which separates the two strands of the reporter. Therefore strand displacement signal is measured in the red channel. The bound stem region of the reporter is 16 base pairs in length. The reporter has several 2'OMe RNA nucleotides and phosphorothioate (PS) bonds throughout. The 2'OMe modifications increase binding affinity and help protect the reporter from ribonucleases and DNases (IDT). The PS bonds prevent endonuclease and exonuclease degradation.

5.1.1 Proof-of-concept detection of endogenously-transcribed mRNA

The reporter mentioned above is difficult and expensive to make. Therefore, rather than creating a new reporter for every mRNA input, instead we an engineered mRNA containing the input of interest to the same reporter will be sensed (Figure 5-2). Since output fluorescence of a protein is related to the concentration of its corresponding mRNA, a fluorescent protein transcript, tagBFP, was chosen as the base mRNA. gBlocks were ordered from IDT to create new gene entry vectors for the Gateway reaction using the Golden Gate reaction. Five different versions were made: tagBFP

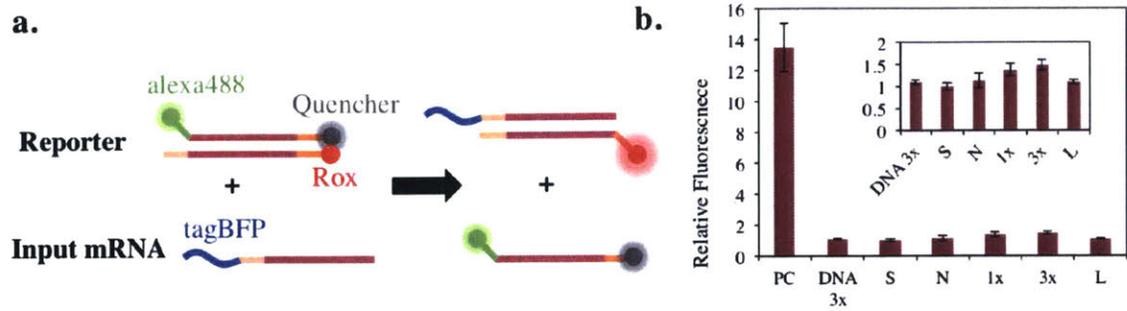


Figure 5-2: FRET reporter sensing of an engineered transcript: (a) Schematic of strand displacement-based sensing of engineered mRNA. The strand displacement-based FRET reporter contains a green fluorescent dye as a transfection marker. The opposite end is functionalized with a red ROX dye and a quencher which absorbs the ROX emission due to its close proximity. When strand displacement occurs with the input mRNA, the top strand is fully displaced, separating the ROX from the quencher and permitting red fluorescence. (b) Different variants of the tagBFP input DNA were transfected in HEK293FT cells, followed by the same RNA reporter. Flow cytometry analysis was performed on 20,000 cells per sample and the data analyzed as follows: the mean red fluorescence value was taken for all highly transfected cells in the blue and green channels (>80%). Input variants include: the positive control where the reporter was incubated with a synthetic input strand prior to transfection (PC), 3 input repeats were appended and transfected without a promoter to test for strand displacement with DNA (DNA 3x), a scrambled input (S), tagBFP without an appended input (N), 1 input sequence appended (1x), 3 input repeats appended (3x), and 1 input sequence at the end of a loop appended (L).

with no input attached, tagBFP with a scrambled version of the input sequence appended directly after the CDS, one input sequence appended directly after the CDS, three repeats of the input sequence appended directly after the CDS, and one input sequence in a loop structure appended directly after the CDS to promote availability. These new entry vectors were combined with a strong Pol II promoter (hEfla) to increase the copy number of transcripts.

Experiments were carried out in HEK293FT cells in a 24-well format and then analyzed using flow cytometry. DNA plasmids containing the various input sequences were transfected using lipofectamine 2000 at high levels (usually 1 μ g) into 150,000 cells in suspension. The cells were allowed to recover and adhere at 37C for 1 day, while the RNA input was transcribed, at which time the cells were resuspended and the reporter was transfected with lipofectamine (1-5 pmol). Cells were plated again and allowed to recover for 24 hours, at which time they were then resuspended again for flow cytometry analysis. The positive control used in these experiments consisted of the RNA reporter pre-incubated with a synthetic input strand in excess

for approximately 15 minutes prior to transfection. This insured the separation of all reporter molecules. The mean fluorescence in the red channel for highly transfected cells in the blue and green channels (>50-90 percentile) was used as the output signal. As seen in Figure 5-2b, fluorescence was analyzed for each input type and normalized to the signal from the sample with a scrambled input. As expected, the sample with no input also exhibits low fluorescence as well as the control sample which contains the tagBFP with 3x input appended but without a promoter. This control ensures that strand displacement is occurring with the transcript only and not the DNA. The samples with 1 repeat of the input and 3 repeats of the input both show increased strand displacement signals (40 and 50%), however the looped-input sample does not.

5.1.2 Increasing the strand displacement signal

While a consistent 50% increase in strand displacement fluorescence (as shown in Figure 5-2) is promising, it is not enough to perform characterization experiments such as those to assess sensitivity and is definitely not sufficient for realistic cases where mRNA transcript numbers are much lower. I aimed to increase the strand displacement signal by using the design space available by several methods: (1) increasing the toehold length, (2) increasing the number of repeats per transcript, (3) optimizing the transfection protocol to achieve maximum signal.

Development of the NxGG protocol

In order to construct plasmids with variable inputs appended at the 3' end required the development of a new system for appending any number of desired short sequences, here called the "NxGG" protocol since it required iterative steps of Golden Gate (GG). This protocol was used with the "Multiple Assembly Vector" created by Jeremy Gam. This vector is a destination vector for Gateway which can accept promoter and gene entry vectors. In addition, however, it contains a Golden Gate cassette at the 3' end for additional 3' UTR modifications. This would allow insertion of any sequence at the end of the mRNA of interest. The NxGG protocol is described in detail

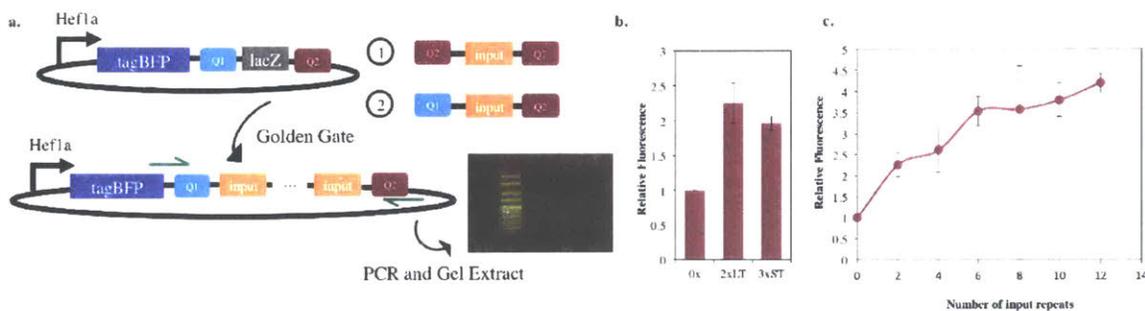


Figure 5-3: Methods for increasing the signal from strand displacement: (a) A new protocol was developed for the simple construction of a plasmid containing any promoter and gene with a desired number of sequence repeats in the 3' UTR. This protocol involves two rounds of Golden Gate, where the first round involves a reaction with the plasmid of interest and an sequence duplex containing Golden Gate sites with identical overhangs on both ends. This allows the addition of many repeats onto the plasmid backbone. The reaction is then quenched with a second matching sequence duplex, however one over hang matches that of the plasmid backbone instead. The addition of this final duplex creates a complete plasmid. PCR is then performed on the plasmid backbone to amplify the 3' UTR containing the sequence repeats. When run through gel electrophoresis, bands are separated based on the number of repeats. A desired repeat number can be selected, excised and re-ligated back into the plasmid. (b) A new reporter was created whereby the toehold length was increased from 6 to 8. A plasmid was created with two repeats of an input containing a complementary longer toehold. The strand displacement signal from this construct is greater than that from a construct with 3 repeats of an input with a shorter toehold. (c) Constructs were created with input repeats ranging from 2 to 12. Each construct was transfected with the same amount of reporter. The strand displacement signal from the reporter increases with increasing numbers of input repeats.

in Appendix A.1.2, but briefly it involves first performing golden gate with an input duplex containing flanking golden gate cut sites with identical overhangs, which allows multiple repeats to be added on to the vector. Then an identical input duplex with overhangs matching the other duplex on one end and the vector backbone on the other can be added to quench the reaction and complete the plasmid (see Figure 5-3a). When PCR is performed on the plasmid, the input repeat count can be identified through gel electrophoresis and further purified and inserted back into the backbone.

Increasing the toehold length

To see if increasing the length of the toehold could increase the strand displacement rate, an additional two bases were added to the end of the reporter to increase the toehold length from 6 to 8 (Giulio Alighieri). New mRNA constructs were developed using the NxGG method mentioned above so that the input repeats contained these

additional two complementary base pairs. As shown in Figure 5-3b just two repeats of an input with a long toehold produces a stronger strand displacement signal than 3 input repeats with a shorter toehold. This "long-toehold" input was used for all subsequent experiments.

Increasing the number of input repeats

Next we increased the number of input repeats on each transcript. Constructs were created with the NxGG protocol with the number of inputs ranging from 2 to 12. As shown in Figure 5-3c the strand displacement signal increases with the number of inputs added. Interestingly, however, the proportional increase in input repeats added does not result in the same fold increase of strand displacement signal.

Determining optimal transfection time

For a transient transfection using lipofectamine, the highest signal would most likely be obtained at an optimal point where the reporter has been incubated for enough time to bind with the input, but not enough time to undergo spontaneous dissociation which would increase the background signal. Also a high signal would occur when mRNA input level is the highest. To address this first point, a kinetic experiment was performed where the cells were monitored after lipofectamine transfection of the reporter using flow cytometry. An ODE model of strand displacement was fit to this data to find the rate constants associated with the experiment and will be discussed below. As can be seen in figure 5-4a, the strand displacement signal for the sample containing the input increases over time, but the background signal in the negative sample also increases, most likely due to reporter dissociation. Therefore there is an optimal window of detection only 3-4 hours after reporter transfection. This short incubation time allowed for subsequent temporal analysis of mRNA levels during the transient transfection of the input (Figure 5-4b). The input DNA was transfected at $t=0$ hrs for 5 different time-point samples and RNA was transfected in each of those samples at different time points ($t=0, 7.25, 31, 46, 64.75$ hrs) and analyzed by flow cytometry 4 hours after this RNA transfection. The trend of the

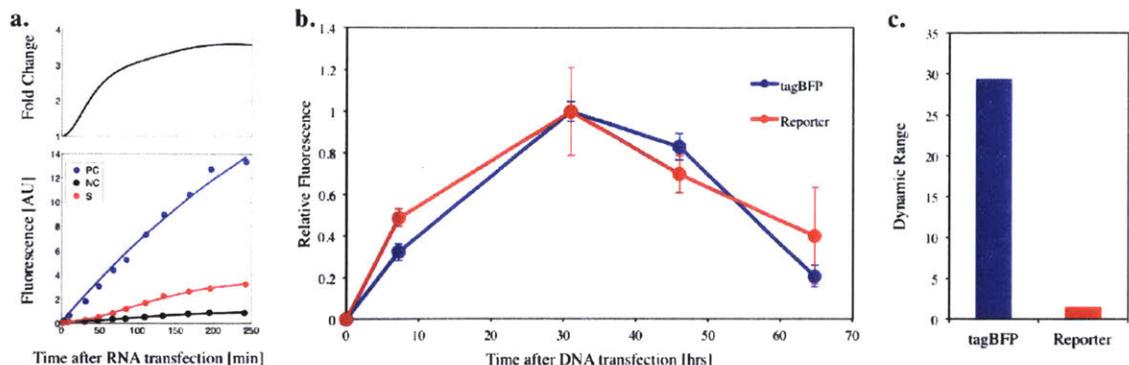


Figure 5-4: Kinetic experiments for transfection optimization: (a) Three different input cell samples were generated: tagBFP DNA without input signals appended transfected with the RNA reporter (NC), tagBFP DNA with 12 input sequences appended transfected with the reporter (S), and tagBFP without input signals appended transfected with the reporter which has been pre-incubated with a synthetic input strand (PC). Cells were resuspended before RNA reporter transfection, and subsequently analyzed by flow cytometry at various times after RNA transfection. Time 0 indicates RNA reporter transfection. Curves were generated from a computational model (discussed below) and the fold change is the the curve from sample S divided by the curve from sample NC. The fold change peaks 3-4 hours after reporter transfection. (b) The concentration of mRNA was evaluated as a function of time during a transient transfection using the strand displacement reporter. Reporter was added at various times after DNA input transfection and flow cytometry analysis was performed 4 hours after each addition of reporter complex. The strand displacement output signal is compared to that of the tagBFP fluorescence which correlate well. The values are normalized to their highest and lowest values, however the bar chart (right) shows a drastic difference between the dynamic ranges of these measurements.

fluorescence level of the tagBFP protein correlates well with the strand displacement signal output, however the dynamic range of the reporter is 20 fold less than that of the fluorescence protein. The mRNA transcript level peaks at around 30 hrs, which is when all subsequent RNA transfection were done unless otherwise noted.

All of these changes allowed for a large increase in the strand displacement signal, which can reach up to 5 fold above the negative control (Figure 5-5) and an observable difference between the population transfected with and without input (Figure 5-5 inset).

5.1.3 Strand displacement characterization

The high signal achieved through the optimizations above enabled characterization of strand displacement with the FRET-based reporter. I studied the kinetics and sensitivity of this method as well as the effect that strand displacement has on the

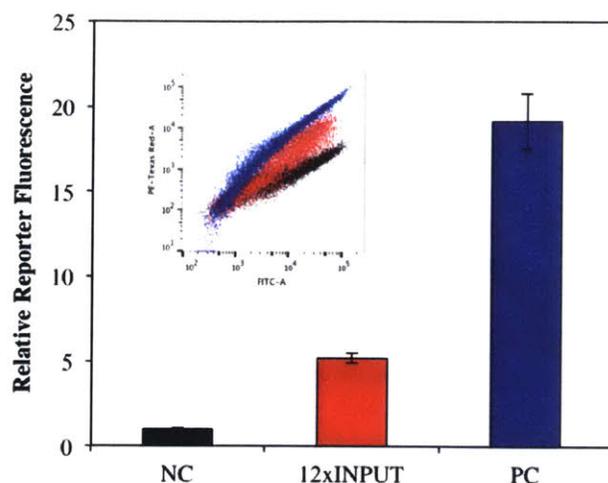


Figure 5-5: Sensor re-evaluation after optimization: Using the results from figure 5, strand displacement signal was re-evaluated using the optimized protocol timeline. Reporter RNA was transfected in suspension 30 hours after input DNA transfection. The cells remained in suspension for 4 hours after the addition of the reporter and were transferred to a FACS tube and analyzed by flow cytometry. The inset shows a large difference in red fluorescence in samples containing 12 repeats of the input (red) compared to samples containing DNA without the input (blue). The mean red fluorescence value of highly transfected (>60%) cells in the blue (input transfection marker) and green (reporter transfection marker) results in a 5 fold change due to input addition.

mRNA input. These experiments allowed for the development of a computational model for strand displacement.

Kinetics and thermodynamics

One of the greatest advantages of strand displacement is the tunability of the reaction. Researchers have extensively studied the kinetics of DNA strand displacement *in vitro* [195], and recently, efforts have emerged that specifically study the *in vitro* biophysics of RNA strand displacement [196]. This is an exciting step forward that I would like to pursue in living cells so that output signal may be a predictable result of cellular inputs using a computational model. This will allow the design of strand displacement sensors and gates with desired kinetics based on input nucleotide sequence. *in vitro*, the kinetics of strand displacement reactions can be accurately modeled and predicted from the length and sequence of the toehold domain, where the rate constant varies from $1 \text{ M}^{-1}\text{s}^{-1}$ to $6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ [197]. Here, I have taken steps to understand the dynamics of RNA strand displacement in mammalian cells

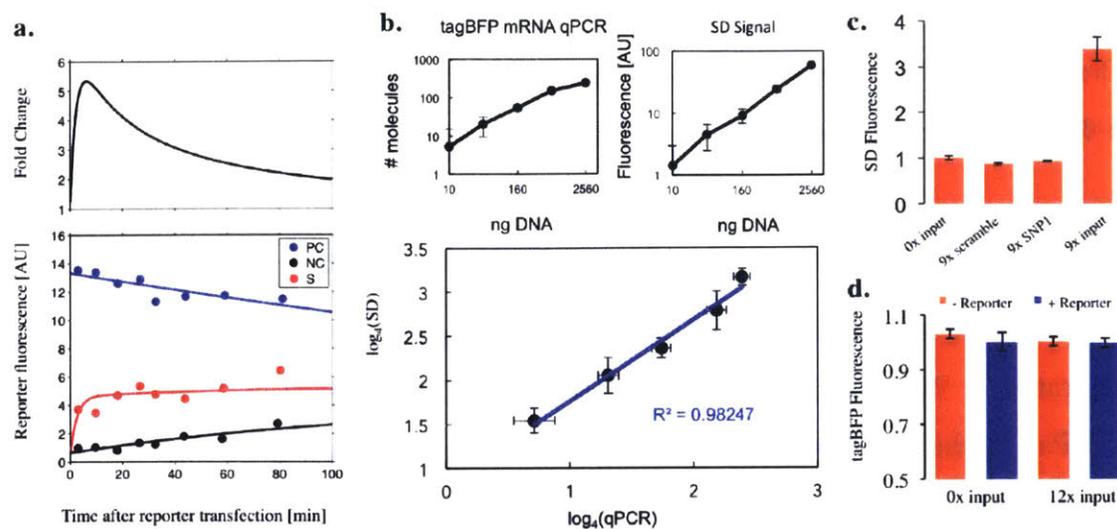


Figure 5-6: Strand displacement characterization of mRNA sensing inside living cells.:

(a) Kinetics of strand displacement evaluated by reporter nucleofection. The reporter alone (NC, S) or pre-incubated with synthetic reporter (PC) was transfected 30 hours after DNA transfection containing no input sequence (NC) or 12 input repeats (S, PC). The same protocol was followed as in figure 5a except the RNA reporter was transfected by nucleofection to separate the rates of strand displacement from those of transfection. (b) Strand displacement was quantified by comparing the signal to values obtained from bulk qPCR. Varying amounts of input DNA were transfected into HEK293FT cells. After 30 hours, cells were either transfected with the reporter or were lysed for RNA purification, reverse transcription and RT-qPCR using TRIzol, QuantiTect reverse transcription kit, and KAPA SYBR fast qPCR kit respectively. I compare the strand displacement signal from flow cytometry to the bulk qPCR values. (c) Constructs were assembled using the NxGG protocol with 9 repeats of either the input, the input with a scrambled toehold, or the input with a single base pair mutation at the end of the toehold closet to the double-stranded reporter sequence. After transfection with the RNA reporter, these were compared to mRNA without any input. (d) The fluorescence of tagBFP was analyzed 24 hours after the addition of reporter compared to cells without reporter transfection to test for the effects of strand displacement on mRNA stability and translation.

using the high signal accomplished above. As can be seen in Figure 5-6a, following nucleofection of 10 pmol of the reporter complex into cells that have been previously transfected with the input DNA, high fluorescent read-out is reached on the order of minutes, which implies through simple calculations that the rate is comparable in cells. The rapidness with which this reaction occurs makes RNA strand displacement a promising technology for monitoring cells in real time. The curves in Figure 5-6a represent the output of my strand displacement computational model which was fitted to the data.

Quantification

Benchmarking the strand displacement sensing method against qPCR (see Figure 5-6b) yields very linear results over 3 orders of magnitude.

Base pair sensitivity

Strand displacement is very sensitive to the base pairs involved in binding and branch migration. We hoped to test whether creating single nucleotide polymorphisms (SNPs) in the input strand would effect its strand displacement rate with the reporter and thus decrease its output signal. The NxGG protocol was used to make constructs with 9 repeats of either the scrambled sequence, the correct input sequence, or the input sequence with a single base pair change. Sulc et al. [196] created a computational model of RNA strand displacement that shows the energy landscape of binding and branch migration. This model shows a decrease in free energy as the binding progresses through the toehold and then an energy barrier as the strand progresses through branch migration. The first thought was to increase the free energy of the lowest point which occurs at the last base of the toehold. As can be seen in Figure 5-6c this prevents strand displacement from occurring: the output signal for the input version with the SNP is essentially equivalent to the scrambled input case.

Cell effects

Another important consideration for our sensors is that they directly interact with cellular miRNA and mRNA, which could have an effect on the dynamics and steady-state levels of these species and thus on cell state [198]. The use of the fuel molecule described above displaces the input (cellular miRNA or mRNA) from our circuits so that it will not be continuously bound and here reduce the "observer effect". However, even though this observer effect is not a likely property of our system, it is important that we test for it through proper controls, which includes quantifying RNA levels in cells with and without our sensors. As a first step, I have monitored the BFP fluorescence from the engineered tagBFP mRNA appended with our input in the

presence and absence of the reporter (see Figure 5-6d). Here, I show that there is little to no change in tagBFP fluorescence, which may indicate that the reporter does cause degradation of the mRNA to which it is bound.

Computational model of strand displacement

I have created an ODE model of strand displacement in MATLAB as shown in Figure 5-7 and detailed in Appendix B.3. The model was first fit to the transfection time-course shown in Figure 5-6a. First the green fluorescence of the negative control was used to determine transfection rate. Then, assuming all reporter is dissociated in the positive control, this sample was fit to the model to determine the relationship between the ROX and Alexa dye fluorescence. The negative control sample could then be used again to determine spontaneous reporter dissociation inside the cell when input is set to 0. Interestingly this spontaneous dissociation accounted for the shape of the negative control curve and therefore the background signal. Additionally, this may be due to off-target effects. Finally the sample containing input can be fit to determine strand displacement rate. These curves are shown in Figure 5-6a. This model was also used to fit the data from the nucleofection kinetic experiment as shown by the curves in Figure 5-6a, which align well.

5.2 Endogenous reporters for mRNA sensing

Using a quick back-of-the-envelope calculation, I find that I likely cannot use the FRET-based reporter mentioned above to detect native endogenous mRNAs due to its limited sensitivity: generally, I transfected 1 μg of input DNA which is transcribed from a very strong promoter. This means on the order of hundreds of plasmids are being transfected into the cell. Additionally, each transcript has 12 copies of the input sequence. In the native case scenario, there are usually one or two copies of DNA for a given gene and only one input sequence per transcript. This would decrease the overall strand displacement rate by 2-3 orders of magnitude. While exploring increased toehold length and GC content mentioned above may change this

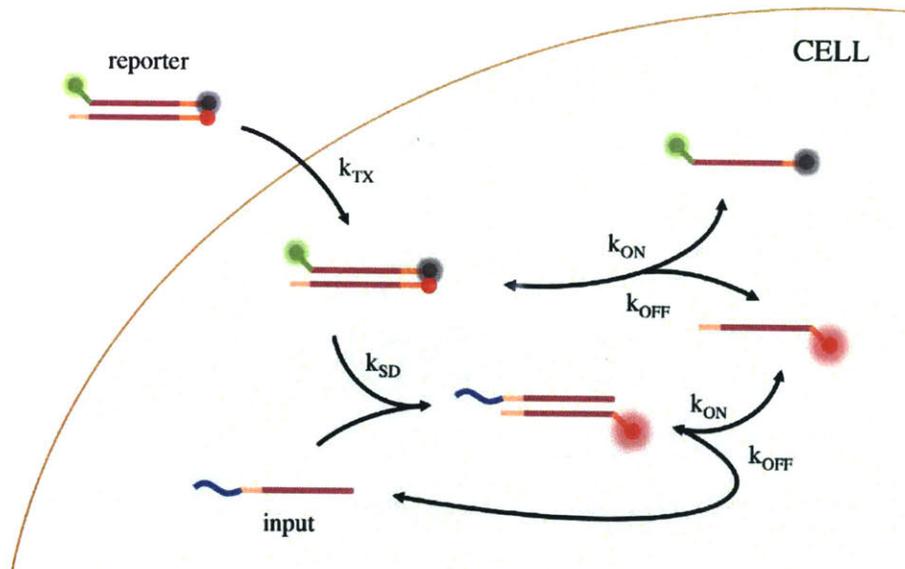


Figure 5-7: ODE model for strand displacement.: For simplicity, this model assumes a constant concentration of mRNA input throughout the 2-3 hour time-span of the experiment. It also assumes complete dissociation of the reporter in the positive control case. k_{TX} refers to the rate of transfection of the reporter into the cell, k_{SD} is a lumped strand displacement term assuming that the dissociation rate of the reporter and input once bound is greater than the rate of branch migration (which is likely true in a crowded cellular environment). Both the new input-reporter complex and the reporter itself can dissociate and associate spontaneously where the rates (k_{ON} and k_{OFF}) are assumed to be the same. Red fluorescence is the sum of the input-reporter complex and the free reporter strand.

estimation slightly, I need to develop an amplification strategy in order to detect native mRNAs at their typical levels. Here I discuss the development of a specific amplification option involving a transcribed mRNA reporter.

Due to the nature of the FRET synthetic reporter method, the dynamic range of the output signal is limited by the fluorescent intensity of the fluorophore itself. As shown in Figure 5-8 the activated reporter (positive control) reaches a level 20 fold above that of the reporter without input. In actuality, however, the strand displacement reaction likely exhibits a much greater dynamic range that is not captured by the fluorescence measurement. In cell-free settings, Zhang and Winfree predict that DNA strand displacement occurs at a rate 6.5 orders of magnitude faster in the case with a toehold than without [199]. We expect RNA strand displacement to follow a similar trend. In fact, a recent publication highlights an mRNA-based toehold-mediated strand displacement reporter with a fluorescent protein output in *E. coli*

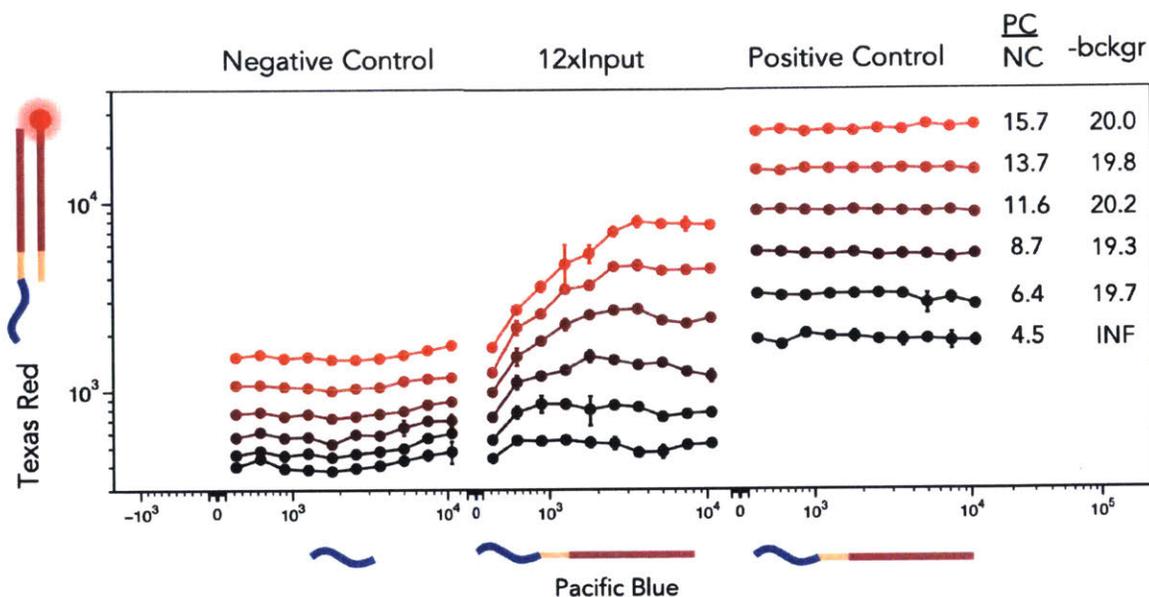


Figure 5-8: Dynamic range of FRET sensor limits its applicability: Reporter RNA was transfected in suspension 30 hours after input DNA transfection. The cells remained in suspension for 4 hours after the addition of the reporter and were transferred to a FACS tube and analyzed by flow cytometry. This method results in a poly-transfection-like space where the data could be binned on reporter amount (FIT-C) and the titration curve for each bin plotted as a function of input (blue) fluorescence. FIT-C bins are designated by colored red lines where red indicated higher fluorescence values. PC/NC indicates the average of the positive control sample divided by the negative control sample for each bin and -bckgr value is calculated by first subtracting the lowest bin values (black) from both data sets.

which can attain over 300 fold protein expression in response to strand displacement with a mRNA of interest [200]. In this section my goal was to develop a similar method for mammalian systems which regulates protein expression that may display a similar dynamic range in cells. This may allow a more accurate representation and prediction of strand displacement dynamic range. Also importantly, for synthetic biology applications, a requirement of sensors is that they convert an input signal into a signal that can be interpreted by the circuit. While a FRET-based fluorescent readout would not accomplish this, a protein readout could be replaced with any regulatory protein of interest which could in turn work within a gene circuit. I describe some efforts towards candidate designs, which are all based on toehold-mediated strand displacement.

5.2.1 Endogenous reporter modeling

When I set out to engineer an mRNA-based mRNA sensor (i.e. endogenous reporter), I first abstractly thought about the ways that this reporter could interact with a mRNA of interest through strand displacement and provide a readout of that interaction. For example, in the toehold reporter developed by Green et al. that is mentioned above, translation of a reporter transcript is inhibited by the transcript's own secondary structure which is relieved through interaction with the mRNA of interest. It is conceivable that an mRNA of interest could interact with the reporter mRNA at any stage of its life-cycle which could effect the final output fluorescence (Figure 5-9): it could affect splicing by interacting with pre-mRNA, it could effect export from the nucleus through interaction with nuclear mRNA, it could affect translation by interacting with cytoplasmic mRNA, or it could interact with any of these species and affect their degradation rates. I considered that many of the interesting mRNAs are exported from the nucleus in 30min and spend a majority of their half-life in the cytoplasm. While it might be interesting to detect splice variants and other nuclear-localized RNAs including lncRNAs, I focused on creating an mRNA reporter that would function in the cytoplasm. Therefore, in interacting with native mRNAs, we are restricted to the effects shown in orange in Figure 5-9. Specifically the mRNA of interest can interact with the reporter and turn it's translation off or on or through its interaction it could turn the degradation of the reporter off or on. All of these mechanisms would cause a change in output fluorescent protein levels which could be detectable.

The first step was to model each of these interactions to understand which mechanisms hold the most promise. Modeling is described in detail in Appendix B.4; Figure 5-10 shows fold change for each reporter mechanism where each mechanism was allowed the same dynamic range. There are a few notable observations. First, ON switches, or mechanisms where interactions with the mRNA of interest increase fluorescent protein expression tend to have a better fold change at intermediate levels between the completely OFF and completely ON states. While the overall dynamic

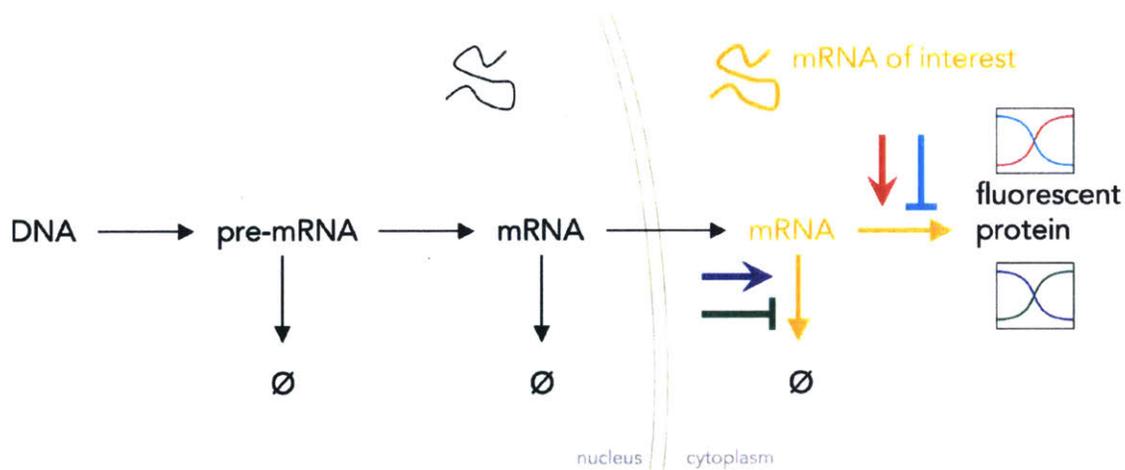


Figure 5-9: Endogenous mRNA reporter: An endogenously-transcribed mRNA reporter can have its fluorescence altered through interacting with an mRNA of interest at any stage in its expression. In detecting cytoplasmic mRNA we consider four options: turning degradation of the reporter off (green), turning degradation on (blue), turning translation of the reporter off (turquoise), and turning translation on (red).

range remains the same this effect is actually simply due to math: for example, if a process has a dynamic range of 100 which has an OFF state of 50 and ON state of 150, if the OFF or ON process is halfway to the ON or OFF state respectively, both processes would be at 100; the fold activation for the ON process is $100/50 = 2$, while the fold repression for the OFF system is $150/100 = 1.5$. Despite this seemingly arbitrary reporting benefit of ON processes, they also have an important practical advantage in that they don't have false positives in a delivery scenario. For an OFF process, fluorescence would decrease if a mRNA of interest was detected by the reporter and it would therefore be impossible to tell the difference between a positive cell and an undelivered cell since both would have low fluorescence (this often necessitates a "transfection marker" control which might be infeasible in an *in vivo* setting). For an ON process a cell that is fluorescing must have been delivered and mRNA detected.

We therefore sought to evaluate both ON process mechanisms for the endogenous reporter: the mechanism where translation is turned on through input mRNA interaction and the mechanism where reporter degradation is turned OFF. The translation-ON mechanism has a slightly better response than the degradation-OFF mechanism for lower levels of input due to the fact that there are much lower reporter amounts

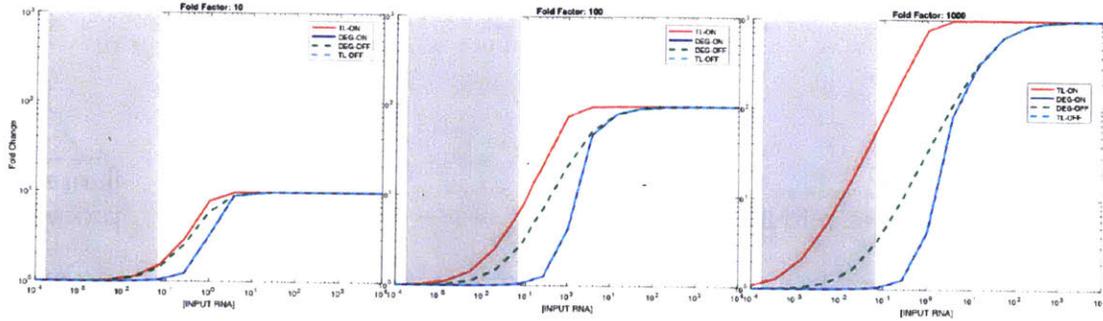


Figure 5-10: Model of various mRNA reporter mechanisms: We modeled the mRNA reporter for each mechanism using three different "Fold Factors": 10, 100, and 1000, implying that at most, each mechanism could activate or repress expression by a maximum of this fold factor. Gray shading indicates putative native mRNA levels calculated from [201]

so interaction is less favorable. Below I elaborate on two translation-ON designs (RNA binding protein aptamer interference and translation-blocking hairpin) and two degradation-OFF designs (miRNA sponge and ribozyme inactivation). It is important to note that none of these design were successful. In fact, this set of projects were the epitome of the "fail fast" methodology where enough experiments were performed to evaluate the promise of each design, which I hope will be valuable to readers hoping to design similar constructs.

5.2.2 RBP aptamer interference sensor

For this strategy, an RNA binding protein (RBP) binds to its aptamer located in the 5' UTR of the fluorescent reporter's transcript which prevents its translation. Through strand displacement with the input, the aptamer is disrupted which inhibits RBP binding and allows for reporter translation (Figure 5-11). As a first proof-of-concept experiment I used the RBP L7Ae and its associated RNA binding motif, the k-turn [35]. Insertion of the k-turn motif in front of EGFP prevents its translation in the presence of L7Ae as previously demonstrated (Figure 5-11). However addition of a plasmid encoding a transcript with a antisense k-turn motif does not relinquish the suppression by L7Ae. In fact it seems that, if anything, the antisense strand binds the reporter transcript and causes more repression. This is confounded by the fact that when a fluorescent protein (mKate2) is encoded in the same strand as the antisense

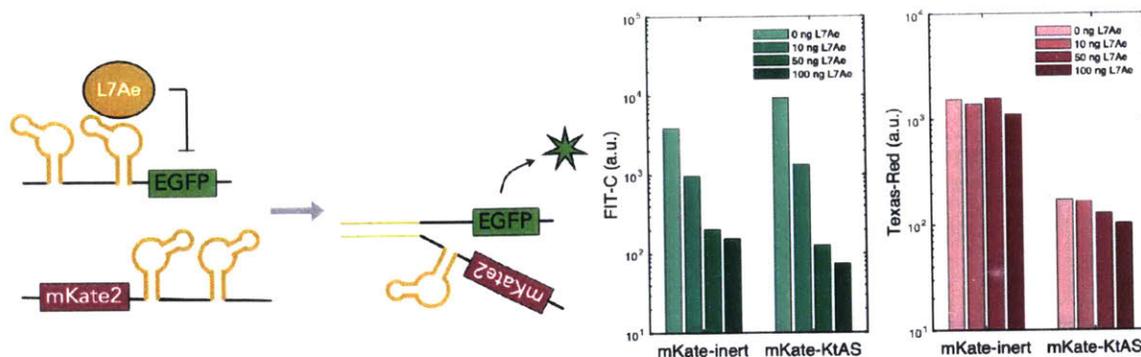


Figure 5-11: RBP aptamer interference sensor: Schematic of this "Translation-ON" design shown on left. L7Ae is prevented from repressing translation of the reporter through the reporter's interaction with an input mRNA. Bars show median FITC for +mKate cells (left) and median mKate for +mKate cells (right).

k-turn the mKate fluorescence decreases compared to a mKate2 transcript with no antisense k-turn motif.

5.2.3 Hairpin releasing sensor

This reporter design strategy shown in Figure 5-12A is most similar to the toehold sensor designed for *E. coli* [200]. But because mammalian systems do not have an RBS that functions like it does in prokaryotes, we had to devise a slightly different strategy. Here, a hairpin loop is placed directly upstream of the start codon which blocks translation initiation. An input mRNA is able to undergo strand displacement with the hairpin stem which removes the blockade. As a first step we evaluated whether hairpins placed directly upstream of the kozak sequence could significantly impede translation. We tried several hairpins with different free energies and GC content (Figure 5-12B) While several hairpins blocked translation quite well (Figure 5-12C,D), these required high GC content which is not compatible with strand displacement which requires GC content closer to 40% to work efficiently.

5.2.4 miRNA sponge sensor

In this reporter design, (shown in Figure 5-13) a synthetic miRNA is able to repress the translation of the reporter fluorescent protein. When an input strand is present,

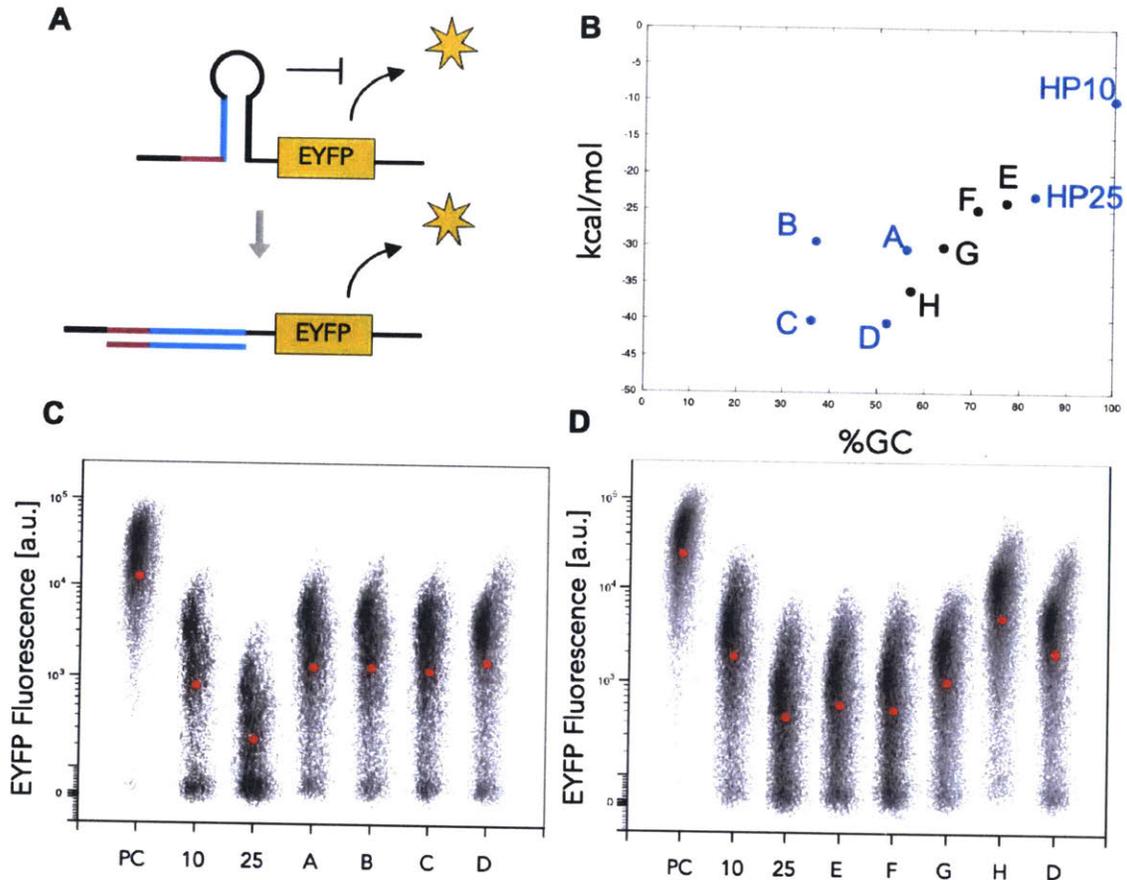


Figure 5-12: Hairpin releasing sensor: (A) Schematic of this "Translation-ON" design. (B) Design space of several candidate hairpin structures. (C-D) Evaluation of the hairpins. Scatter plots show EYFP values +Blue (transfection marker) cells and red dots indicate median.

strand displacement occurs with an intermediate gate, for which the output of the gate is an miRNA sponge. The seed region of the miRNA is exposed as the toehold. Sponging of the miRNA releases suppression of the reporter mRNA. Importantly, once the sponging of a certain miRNA is established, the domains of the intermediate gate can be changed to accept any RNA input, such as an upstream output signal or an endogenous mRNA of interest. Here we test this device using FF4 siRNA (Figure 5-14). When FF4 siRNA is added to cells containing a fluorescent reporter with 4xFF4 target sites, expression is repressed. However neither the gate or gate and input mRNA are able to sponge the siRNA to release repression (Figure 5-14A). This is likely due to the fact that the effect from sponging siRNA is quite small, less than 4 fold when even very large amounts of sponge (1000ng) are included. To further

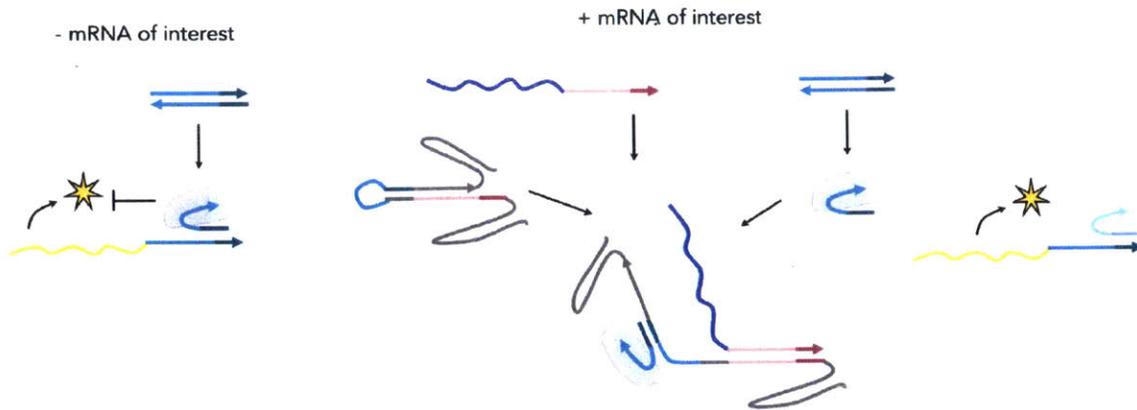


Figure 5-13: miRNA sponge sensor mechanism: Schematic of this "Degradation-OFF" design. When a U6-transcribed gate interacts with an input mRNA via strand displacement, a miRNA target site is revealed in the gate structure. An siRNA-RISC complex that was previously repressing a fluorescent reporter is sponged away by the gate therefore releasing the reporter for active expression.

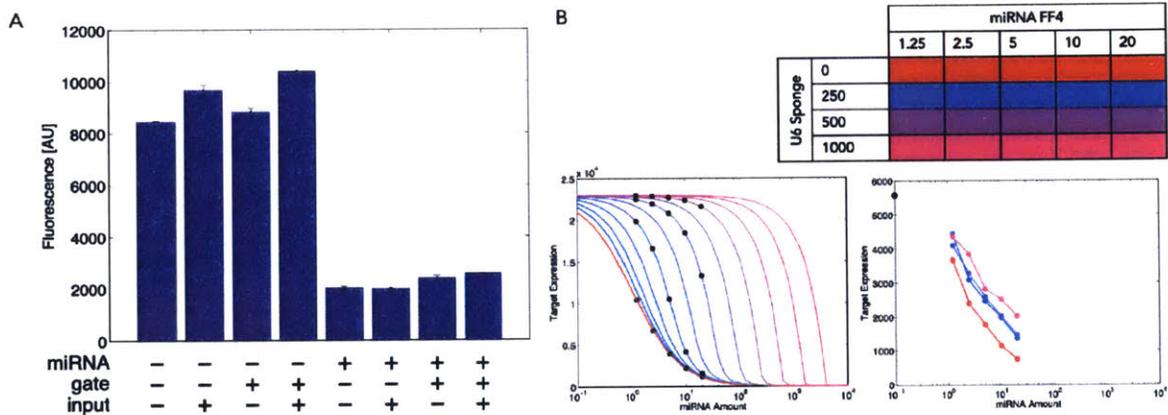


Figure 5-14: miRNA sponge is not able to relieve repression: (A) Median fluorescence output values are calculated for each combination of miRNA, gate, and input mRNA. (B) Fluorescence values are plotted across miRNA amount for various sponge amounts. Black dot indicates no miRNA. When compared to computational model we see that even very large amounts of sponge are unable to restore expression. Decreasing the amount of miRNA could make this possible however the dynamic range will be diminished.

test whether we could improve sponging capability by including a bulge in the target site, we transfected a replicon sponge into cells where an mKate transcript was being repressed by FF4 siRNA (Figure 5-15). In this study, when the siRNA is a perfect match to the target site, the reporter is repressed more efficiently, however the sponge recovers fluorescence slightly only when there is a bulge.

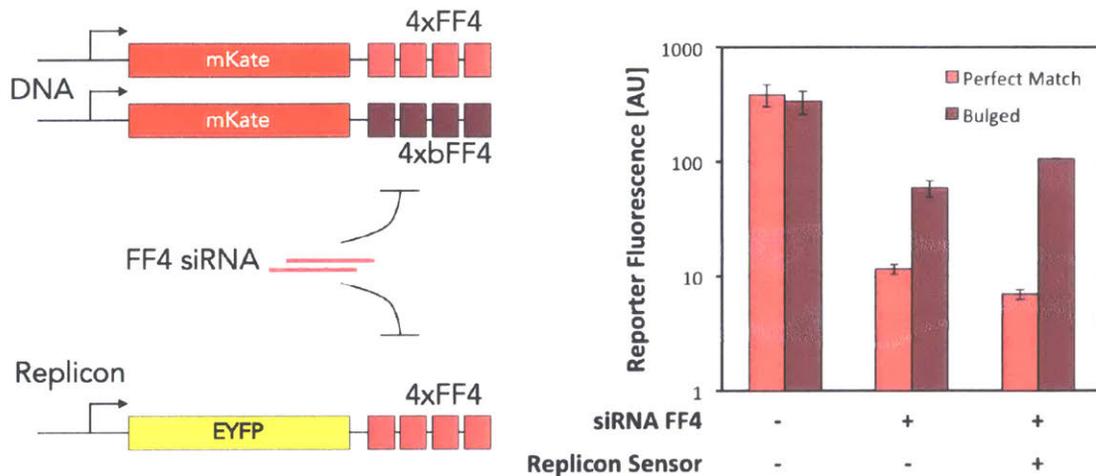


Figure 5-15: Evaluation of miRNA sponges with bulged target sequence: Bars show median mKate fluorescence in response to both the miRNA and the Replicon sponge. Pink indicates that the mKate reporter contains bulged target sites while dark red indicates the target sites are perfectly matched.

5.2.5 Ribozyme sensor

In this design a hammerhead ribozyme is embedded in the transcript of a fluorescent protein which causes RNA cleavage and degradation. The ribozyme stem can undergo strand displacement with the input which disrupts the hammerhead structure preventing cleavage and promoting mRNA stability and translation. As depicted in Figure 5-16, the strand displacement region is not associated with any sequence of the hammerhead ribozyme (shown in gray), which allows the sequence to be customized to any input of interest. Some aspects to consider here is that the interaction of the input with the reporter mRNA must occur in the cytoplasm so that the reporter can be translated. This may constrain the ribozyme choice to those with slightly slower rates to allow for nuclear export prior to cleavage. The N79 hammerhead ribozyme from Yen et al [202] is a good starting point since it has been extensively characterized and used by other labs [203, 204]. Different mutations can effect its cleavage rate allowing for tunability. Here, we started with the N107 hammerhead ribozyme from the same paper which removes upstream start codons and which causes strong repression of a fluorescent reporter when placed in the 5' UTR compared to a mutated version of the ribozyme. However, we found that adding an input complementary to

the stem did not inhibit the ribozyme. We hypothesized that the ribozyme might be cleaving too fast to be able to be transported into the cytoplasm to interact with the input. Simple calculations imply that this might be true. If we model a transcript that doesn't contain a ribozyme M_0 and a transcript that does contain a ribozyme M_X , we can write the simple ODEs to understand how quickly the ribozyme is cleaving inside cells: Since we know that mRNA half-life is approximately 2-12 hrs ($\gamma = 0.001 - 0.005 \text{ min}^{-1}$) this means the ribozyme cleavage rate, $k_x=0.075 - 0.375 \text{ min}^{-1}$, or in other words it cleaves every 1-9 minutes. This is slightly slower but comparable to *in vitro* ribozyme rates, however this is much faster than export times from the nucleus (approximately 30 min). For this reason we tried to mutate the ribozyme to allow for slower cleavage rates (Figure 5-16, Number 4) which worked well but resulted in a reduced dynamic range. The dynamic range could be restored by adding a longer stem to the mutated ribozyme, however using a scrambled input stem to simulate strand displacement resulted in only minor reductions in cleavage efficiency and low dynamic range. Overall, in order to create a ribozyme that was able to cleave outside of the nucleus would require mutations and diminishing the dynamic range of the device which prevents strand displacement detection. One solution would be to reverse the logic where interaction with a mRNA of interest activates the ribozyme. This would keep transcript levels high for interaction and only result in the irreversible cleavage as an output of the reaction. However, canonically this would be a "Degradation-ON" mechanism which we have shown in the model above (Figure 5-10) does not work well. Instead if ribozyme cleavage resulted in an increase of fluorescence it might work better. Such reversing of the logic of ribozyme cleavage is explored in the PERSIST ON-switch in Chapter 4.

5.3 Discussion and Future Directions

This Chapter helped to elucidate important determinants of strand displacement success as a platform inside mammalian cells through characterizations with a FRET reporter, computational modeling of various cellular reactions with RNA reporters,

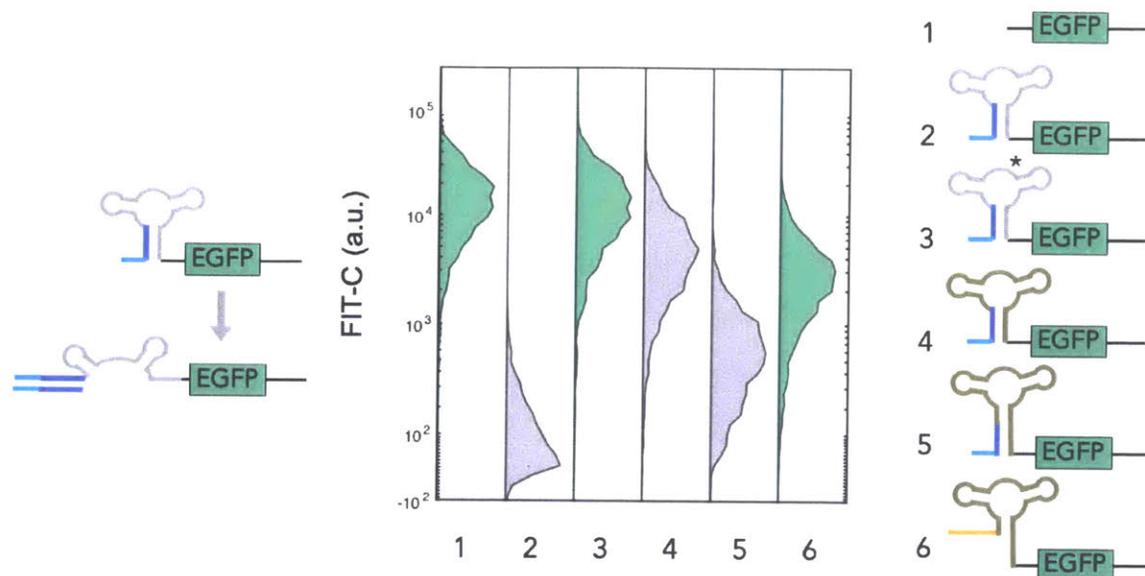


Figure 5-16: Ribozyme inactivation sensor.: Schematic of this "Degradation-OFF" design (left). Five different hammerhead ribozyme structures are evaluated and compared to a constitutive reporter. Bars indicate median of all +Red (mKate transfection marker) cells. Version 2 is N107, version 3 is the inactive mutant of N107, version 4 contains a mutation that decreases the ribozyme activity, version 5 is the same slower ribozyme with a longer stem, and version 6 contains the same additional stem structure with a scrambled lower stem to simulate strand displacement product.

and evaluation of several endogenous strand displacement reporter designs in living cells. While RNA has some useful design features such as Watson-Crick base pairing and secondary structure determination there is still quite a lot to learn about the predictability of RNA:RNA interaction inside living cells. An enzymatic output such as miRNA or gRNA (rather than mRNA used in the second half of this Chapter) could serve as an amplification method of reading out strand displacement reactions. Interactions with ribozymes appear to be the most promising route forward for affecting output based on RNA:RNA interactions. The ribozyme activation strategy described here could be further explored with the PERSIST ON-switch platform.

Appendix A

Appendix: Methods

A.1 Updated Golden Gate DNA Assembly Strategies

Here I develop new circuit assembly frameworks for updating all levels of cloning to use Golden Gate Assembly like similar assembly strategies used in our lab [205]. Golden Gate assembly is especially efficient because it uses Type II restriction enzymes which cut outside of their overhangs which means that their own restriction site is removed (irreversible) and it is scarless. I develop hierarchical Golden Gate-based assembly frameworks for several cloning requirements below.

A.1.1 New backbones and standardization for sub-pL0s and pL2

While Golden Gate is frequently used via the MoClo standard [206] for transcription unit generation ("plasmid level 1", pL1), other cloning methods are often used for other assembly requirements. For example, Gibson cloning [205] is typically used to assemble multiple pL1 transcription units onto the same "levels 2" (pL2) plasmid, however it is fairly expensive, requires a decent amount of hands-on time and is not as efficient as golden gate assembly. In Figure A-1, I outline my cloning framework for creating sub-level 0s as well as new backbones that can be used with SapI Golden

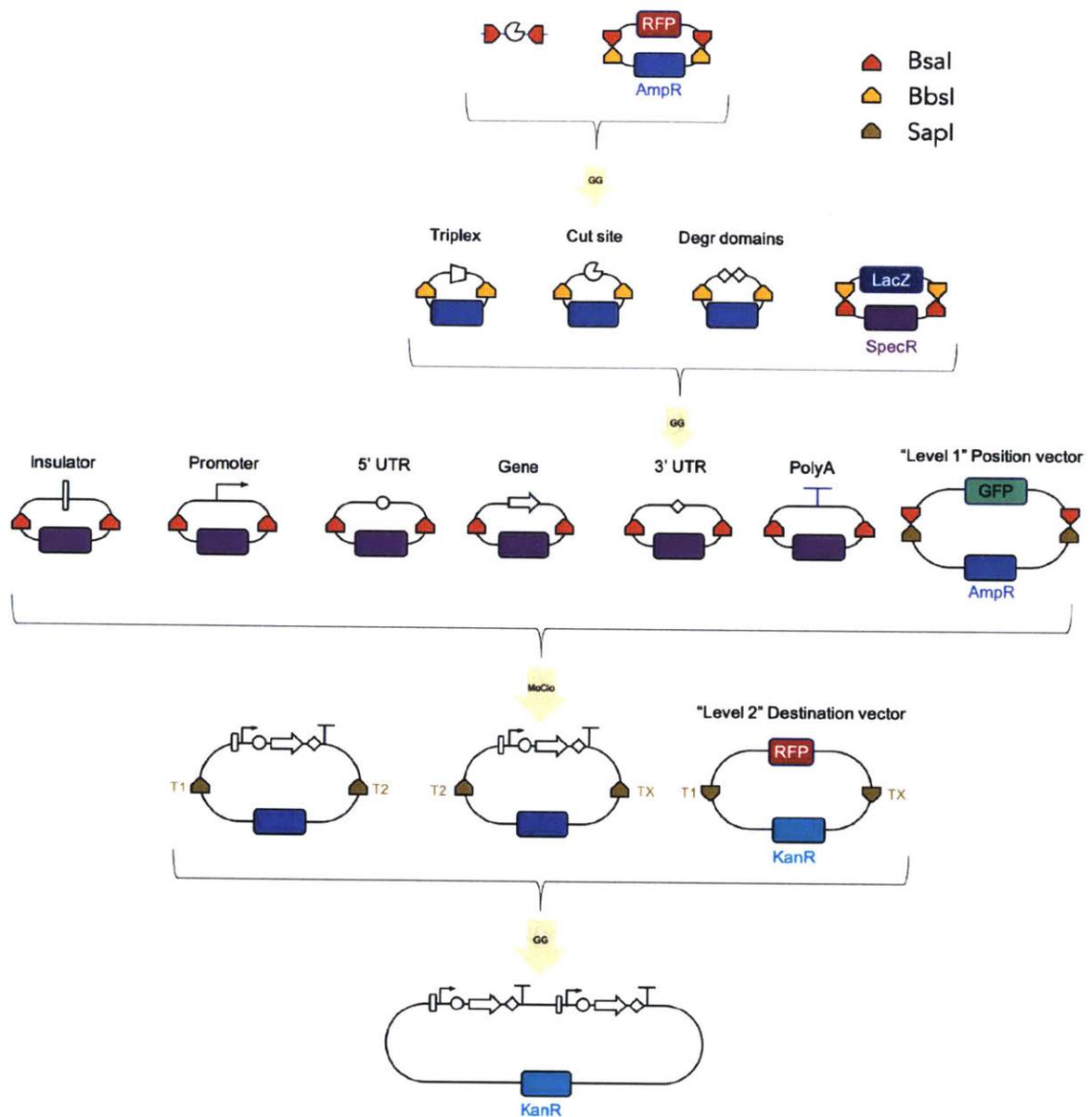


Figure A-1: An updated "All Golden Gate" assembly framework:

Gate to generate pL2 efficiently. Note that now all adjacent levels contain different colored screening markers (pL0s: LacZ, pL1: GFP, pL2:RFP) to facilitate creating "cassettes" by using lower level backbones in higher level cloning reactions. I highlight the sub-pL0-3 assembly strategy used in Chapter 4 to create the 3' PERSIST ON-motif that required swapping out different recognition sites and numbers of repeats, however this strategy can be used for any type of pL0. The standard overhangs for SapI pL2 assembly are listed in Table A.1 and plasmids that are part of this new

assembly framework are listed in Table A.2. The SapI overhangs are annotated with "T[N]" rather than the Gibson standard annotation "Seq[N]".

Overhang	Sequence
T1	ACC
T2	CGC
T3	TAC
T4	CAG
TX	AGA

Table A.1: caption

pBD Number	Description
pBD15.XaG	pL1:T1T2
pBD16.XaG	pL1:T2TX
pBD17.XaG	pL1:T2T3
pBD18.XaG	pL1:T3TX
pBD19.XaG	pL1:T3T4
pBD110.XaG	pL1:T4TX
pBD22.XcR	pL2:T1TX standard backbone
pBD23.XcR	pL2:T1TX landing pad backbone

Table A.2: caption

A.1.2 Multimerizing Golden Gate (NxGG) protocol for generating repeated elements

It is often useful to engineer many copies of a sequence into a plasmid. For example miRNA target sites are often repeated four times but other applications may require many more repeats which are difficult to synthesize. Here I describe a customizable method to generate many repeats of a sequence using Golden Gate (GG) that we call "NxGG" (Figure A-2).

The protocol works by designing a repeat element containing Type IIs restriction sites with the same overhangs on both sides ("part 1". A GG reaction step with the backbone will allow part 1 to concatenate on the backbone. Adding the quencher, "part2", binds to part1 and the backbone to close the plasmid and complete the

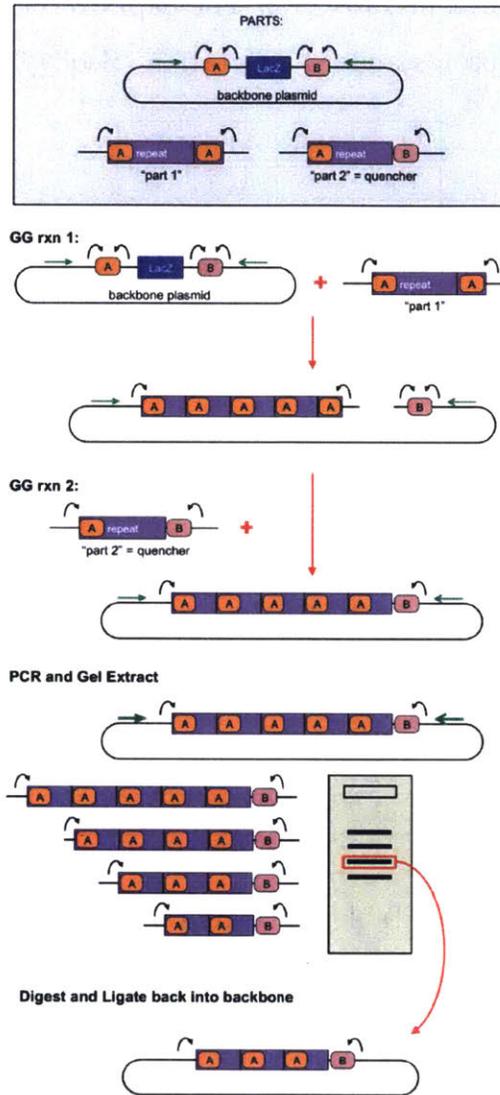


Figure A-2: NxGG protocol:

reaction. When directly transformed the reaction has significant bias towards only one repeat so generally it is best to PCR from the backbone of the plasmid, gel extract the band of choice with selected number of repeats and re-digest and ligate back into the backbone.

Protocol:

1. Set up reaction as follows in 20uL

40 fmol backbone plasmid
200 fmol "part 1"
Ligase Buffer
Ligase
Type IIs Restriction Enzyme
Water

2. Cycle 1:

37°C 15 min

15x[37°C 2 min

16°C 5 min]

8°C pause

3. Add to reaction:

40 fmol "part 2 quencher"

Ligase Buffer

Ligase

Type IIs Restriction Enzyme

4. Cycle 2

37°C 15 min

15x[37°C 2 min

16°C 5 min]

50°C 5 min

80°C 5 min

8°C store

5. PCR 1uL of NxGG reaction

6. Run PCR reaction on gel to isolate band of desired size (2% gel at 50V works well to isolate small bands)

7. Gel extract band of desired repeat length

8. Digest and ligate back into original backbone

9. Transform into competent cells that support cloning repeats (e.g. NEB Stable cells)

A.2 Cell culture protocols

A.2.1 Lipid transfections: co-/poly-transfection optimization

All flow cytometry results that measure plasmid fluorescence expression in this thesis were accomplished via plasmid lipid transfection into mammalian cells. While co-transfections involve using the standard Lipofectamine 3000 protocol worked quite well, poly-transfections did require some optimization. Transfections using Lipofectamine 300 used a 2.2:1 P3000:DNA ratio and 2.2:1 Lipofectamine:DNA ratio. The following table summarizes transfection parameters used in transfection experiments:

	1-complex	2-complex	3-complex
Total DNA	500 ng	2 x 300 ng	3 x 250 ng
DNA of components (pL1 4kb)	50/500 ng	75/300 ng	250/500 ng
DNA of components (pL2 6kb)	75/500 ng	100/300 ng	375/500 ng
Format	24-well	24-well	12-well
Collect (cells)	20k	100k	1e6

Cells were transfected in suspension and analyzed after 48 hours.

A.2.2 Flow cytometry

Flow cytometry was performed using a BD LSRFortessa equipped with a 405nm laser with 450/50nm filter for measuring EBFP and tagBFP and other blue fluorescent proteins in the "Pacific Blue" channel, a 488 laser with 530/30 filter for measuring mNeonGreen, EGFP, EYFP, and other yellow/green fluorescent proteins in the "FIT-C" channel, a 561nm laser with 582/15nm filter for measuring mKO2 in the "PE" channel and 610/20 filter for measure mKate in the "PE-Texas-Red" channel, and a 637 laser with 710/50nm filter measuring iRFP720 in the "Alexa Fluor 700" channel.

A.2.3 Advanced flow cytometry analysis

I have developed a flow cytometry analysis pipeline for MATLAB that is freely available. It allows users to query and calculate various metrics of a multi-dimensional

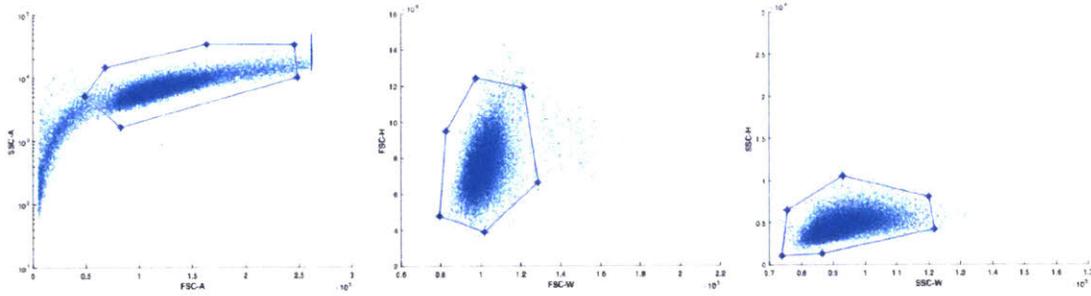


Figure A-3: Morphological Gating:

fluorescent parameter space after pre-processing their data. The analysis pipeline is performed for all data in this thesis as follows:

1. **Set-up:** Experiment header is defined which contains the path of the data folder, colors used (with associated color control files), non-transfected control file, beads file if needed and a defined binning schema if desired.
2. **Set-up:** Morphological gating is performed using a non-transfected cell sample with the function “creatMorphoGate”. The user can click a polygon on a plot of forward and side scatter to identify the cell population of interest and remove debris, dead cells, and aggregates (see Figure A-3).
3. **Set-up:** Compensation coefficients matrix and autofluorescence values are automatically calculated using single color controls with the function “calcCompCoeffs”. The fluorescence compensation step here helps to correct for any bleed-through from one channel into another unintended channel. (Figure A-4)
4. **Data analysis:** Raw data is extracted from samples of interest using “getflow-data” and requiring only the file name, experiment channels and any gates that have been created including the morphological gate.
5. **Data analysis:** Retrieved raw data is compensated with the function “compensate” using the previously calculated compensation parameters.
6. The data is now ready for use!

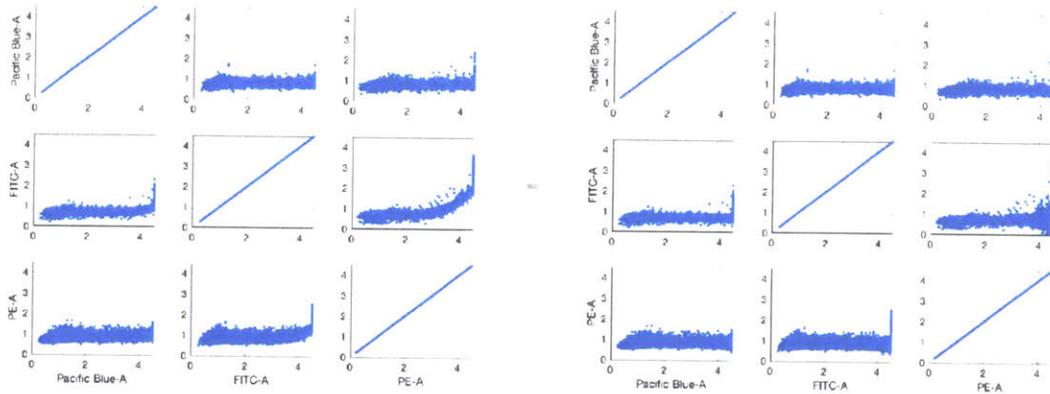


Figure A-4: Bleed-through correction:

These scripts are straightforward and easy to implement. The experimental set-up including gating and compensation calculation are performed only once with their values and associated plots stored within the data folder for future reference. The user can then define a binning scheme or perform any calculations or fitting algorithms to any portion of the data by indexing the matrix of compensated fluorescence values. My “evalbins” function takes in the defined binning schema and computes any one or number of (MATLAB or user-defined) functions that should be performed on each bin. Additional plotting functions have also been created to allow users to generate attractive graphs that reveal various aspects of their datasets. Some of my favorites include “plohiststacked” for histogram plotting (Figure 4-16 middle), “plotsurflines” for quickly generating 3D line plots, “plothmp” for heatmaps with non-integer axes (Figure 4-12), and “scatterplotbins” which plots scatterplots with points colored by associated bin.

Appendix B

Appendix: Theoretical modelling

"All models are wrong; some are useful."

- George E. P. Box

B.1 Theoretical model of the RNA-based switch circuit (replicon vs. pDNA)

This section contains work from the following published manuscript and therefore represents a collaboration between myself, Liliana Wroblewska, Tasuku Kitada, Kei Endo, Velia Siciliano, Hirohide Saito and Ron Weiss.

Wroblewska, L., Kitada, T., Endo, K., Siciliano, V., Stillo, B., Saito, H., Weiss, R. (2015) "Mammalian synthetic circuits with RNA binding proteins delivered by RNA" *Nature Biotechnology*, 33 (8): 839–841. doi: 10.1038/nbt.3301

In the absence of siRNA FF4 or FF5, the replicon-based and plasmid-based switch systems exhibit different behaviors (as seen in Figure B-1). The replicon-based system seems to fall into a more "mutually exclusive" distribution fairly soon after transfection, whereas the plasmid-based system appears to maintain a more unimodal population at an intermediate state for at least 48 hours.

To investigate these observations, simple computational models of the pDNA and replicon systems were implemented and analyzed. Stochastic simulations using the

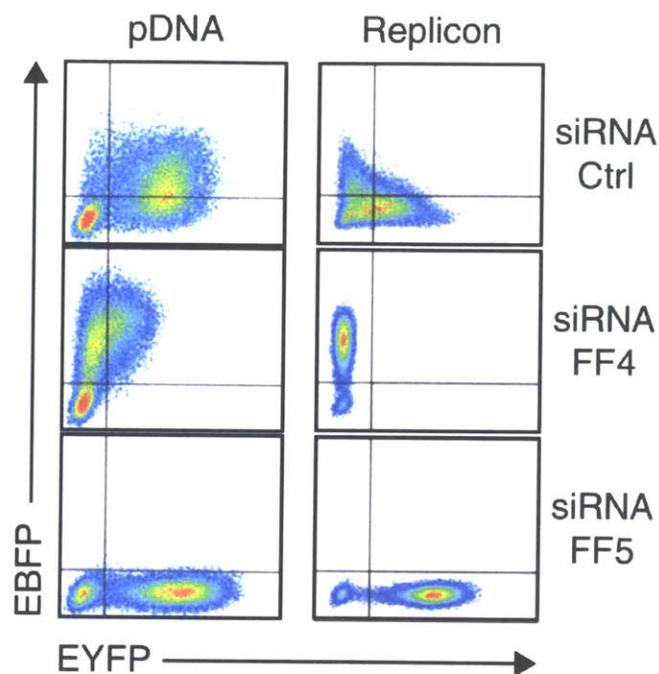


Figure B-1: pDNA and replicon toggle switch results: Corresponding representative two-dimensional flow cytometry plots for pDNA and replicon transfections of the RNA toggle switch.

Gillespie Algorithm were performed in MATLAB [207] using HTCondor queued computer cluster at MIT Computer Science and Artificial Intelligence Laboratory. The reaction equations and rates are reported below, and model schematic diagrams are displayed in Figure B-3. Unless otherwise stated, 96 cell simulations were performed for each parameter set. To assess the bimodality, or "mutual exclusivity", of the population of cells that results from these simulations, a Mutual Exclusivity (MEx) score was developed (described in Figure B-4). As demonstrated, this score provides useful information for the analysis of the two systems. The pDNA and replicon systems were simulated for 48 and 24 hours respectively with the parameter values listed in Table B.1. These simulations resulted in populations that were qualitatively very similar to Figure B-1 (see Figure B-5).

First, to better understand the nature of the unimodal state achieved by the pDNA system, several of the parameter values were varied. The resulting behavioral trends are shown in Figure B-6. When either the starting copy number (P_0) or the transcription rate (k_{TS}) is set to lower values, the system becomes more mutually

exclusive. This implies that this state might be due to a high and simultaneous burst of expression from both the L7Ae and MS2-CNOT7 plasmids, which express stable proteins. To investigate this, the variance-to-mean ratio (VMR), also known as the index of dispersion, of the distribution from which the starting plasmid copy number was selected was varied to allow for greater degree of initial bias. Since P0 follows a Poisson distribution with $\text{VMR} = 1$, this was achieved by selecting from a Poisson distribution with $\text{mean} = P0/\text{VMR}$ and then multiplying that value by the VMR. In this way, the mean of the distribution stays the same but the variance is increased. Increasing the VMR led to a higher degree of initial bias, causing a large increase in bimodality even when the initial copy number is kept high. In addition to the simultaneous burst of expression, this seemingly unstable state can be maintained for some time due to the slow switching rate, which is greatly influenced by the degradation rate of the proteins. To demonstrate this, we also increased the degradation rate, which causes an increased MEx score.

The next question to investigate was why the replicon system does not go through this high/high state. Based on the results from the pDNA analysis, we hypothesized that the replicon system either avoids the simultaneous burst of expression or it has a faster switching time due to the feedback mechanisms involved in the first few hours post-infection (ongoing negative strand synthesis). As depicted in Figure B-3, the computational model of the replicon system essentially mirrors the pDNA system after the 4 hour time point when negative strand synthesis ceases. Therefore, for computational simplicity, we chose to simulate the replicon system for just 4 hours post-infection in our analysis. Sample simulations were also run for 24 hours to verify this simplification. In our simulations, we carefully analyzed the parameters involved in both of these hypotheses: the starting replicon copy number ($R0$), the transport rate (k_{TR}), the transcription rate (k_{TS}), positive feedback (ϵ), and replication inhibition (β), and how the system responds to changes in each of these parameters.

We performed global sensitivity analysis by randomly sampling 2000 parameter sets from the log-transformed realistic parameter space (see Table B.1). Figure B-2a shows the MEx scores when each of the parameters is varied 1.5 decades within its

realistic set of values. Each point represents the distribution of 96 cell simulations for one random parameter set. Figure B-2b shows heat maps of the MEx scores for parameter pairs plotted together to identify parameter interactions.

From these results, it appears that the stochasticity of replicon transport and spherule formation plays a major role in the dynamics of the system. In fact, the feedback mechanisms would not even be possible without the independent nature of this process, which distinguishes it from the pDNA system. However, Figure B-2b indicates that the system is fairly insensitive to the strength of the feedback mechanisms (β and ϵ). Even at low kTR where replication inhibition (β) would have the most effect, we see no correlation with the MEx score. This is consistent with our experiments where we also find that L7Ae does not inhibit replication. To our knowledge there is no evidence that the binding of RBPs such as L7Ae to replicons affects formation of spherules. However, MS2-CNOT7 degradation of genomic RNA is likely. The positive feedback (ϵ) due to + genomic strand synthesis also has no effect on the bimodality of the system in our simulations.

There is, however, a strong relationship between mutual exclusivity and both R0 and kTR, the initial replicon copy number and the transport rate. Both relate to the independent and stochastic nature of spherule formation. Decreasing either R0 or kTR leads to an increase in MEx score. This occurs because stochasticity in the transport reaction increases, allowing an initial bias in replication. As expected, their effects are also correlated (Figure B-2b). These results are also biologically relevant. The transport of replicons to the plasma membrane for spherule formation and negative strand synthesis is carried out by the nonstructural protein P1234, which is translated only when the opal codon is read thru (about 10% of the time). This low level of active protein could lead to initially slow and stochastic transport events, especially when the number of transported species is low. Additionally, our electroporation experiments indicate that there is a delay in protein expression, when delivered gene is encoded on replicon as compared to pDNA (assuming same delivery method), which implies that spherule formation may take a significant amount of time. Lastly, our qPCR results and a related publication [208] suggest that the

starting number of replicons per cells in our electroporation experiments may be in the low tens while literature indicates that transfected pDNA copies are in the high tens to hundreds [209, 210]. Additionally, bimodality is further amplified by an increase in transcription rate for the replicon system (which is in contrast to the pDNA case where higher transcription rate decreases bimodality). Here, however, increased transcription serves as an amplification of the initial bias caused by transport delay. In general, alphaviruses are able to replicate very quickly [211], so this computational result is biologically realistic.

Overall, these results suggest that the individualized and stochastic nature of spherule formation and transport results in an initial bias in replication. The resulting bimodality can be realized in the first four hours post- infection. The effects are amplified by an increase in stochasticity through a decrease in replicon copy number, and by a fast replication rate (k_{TS}). These differences in dynamics are likely to have important implications when using replicons in synthetic biology circuits, especially when the expression timing of various species is important to circuit functionality

B.1.1 Computational Model

pDNA MODEL

Species:

pC	nuclear MS2-CNOT7 plasmid
pL	nuclear L7Ae plasmid
mC	MS2-CNOT7 mRNA
mL	L7Ae mRNA
LmC	L7Ae protein bound to cytoplasmic MS2-CNOT7 mRNA
CmL	MS2-CNOT7 protein bound to cytoplasmic L7Ae mRNA
L2mC	L7Ae protein doubly bound to cytoplasmic MS2-CNOT7 mRNA
C2mL	MS2-CNOT7 protein doubly bound to cytoplasmic L7Ae mRNA
C	MS2-CNOT7 protein
L	L7Ae protein

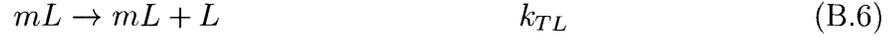
Reactions:

Transcription Transcription is assumed to be first-order upon cell division when the pDNA enters the cell nucleus.

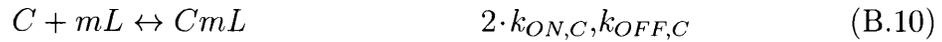
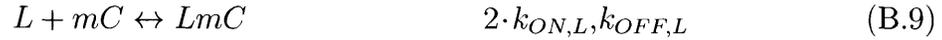


Translation Translation is assumed to be first-order. While MS2-CNOT7 binding does not have any steric effect on L7Ae translation, bound L7Ae greatly inhibits translation. When one L7Ae protein is bound to the RNA it inhibits translation by a factor, σ , and when two copies of L7Ae are RNA-bound translation is inhibited twice

as much.



Repressor Binding/Unbinding For simplicity, two binding sites were assumed for both MS2-CNOT7 and L7Ae RNA. A second-order association rate is used and first-order dissociation rate.



Degradation First-order degradation rates were assumed. When the deadenylase MS2-CNOT7 is bound to L7Ae RNA it increases the RNA's degradation rate by a factor, α . In addition to these reactions, all species (including plasmids) are diluted

by cell division.

$$mC \rightarrow 0 \quad \text{deg}_R \quad (\text{B.13})$$

$$mL \rightarrow 0 \quad \text{deg}_R \quad (\text{B.14})$$

$$LmC \rightarrow L \quad \text{deg}_R \quad (\text{B.15})$$

$$CmL \rightarrow C \quad \text{deg}_R \cdot \alpha \quad (\text{B.16})$$

$$LmC \rightarrow mC \quad \text{deg}_P \quad (\text{B.17})$$

$$CmL \rightarrow mL \quad \text{deg}_P \quad (\text{B.18})$$

$$L2mC \rightarrow 2L \quad \text{deg}_R \quad (\text{B.19})$$

$$C2mL \rightarrow 2C \quad \text{deg}_R \cdot 2 \cdot \alpha \quad (\text{B.20})$$

$$L2mC \rightarrow LmC \quad \text{deg}_P \quad (\text{B.21})$$

$$C2mL \rightarrow CmL \quad \text{deg}_P \quad (\text{B.22})$$

$$C \rightarrow 0 \quad \text{deg}_P \quad (\text{B.23})$$

$$L \rightarrow 0 \quad \text{deg}_P \quad (\text{B.24})$$

REPLICON MODEL

Species:

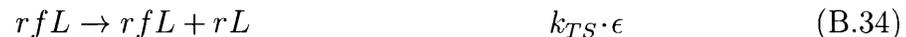
rC	cytoplasmic MS2-CNOT7 replicon (genomic)
rL	cytoplasmic L7Ae replicon (genomic)
rfC	MS2-CNOT7 replicon in spherule (replication factory)
rfL	L7Ae replicon in spherule
LrC	L7Ae protein bound to cytoplasmic MS2-CNOT7 replicon
CrL	MS2-CNOT7 protein bound to cytoplasmic L7Ae replicon
L2rC	L7Ae protein doubly bound to cytoplasmic MS2-CNOT7 replicon
C2rL	MS2-CNOT7 protein bound to cytoplasmic L7Ae replicon
mC	MS2-CNOT7 mRNA (subgenomic)
mL	L7Ae mRNA (subgenomic)
LmC	L7Ae protein bound to cytoplasmic MS2-CNOT7 mRNA
CmL	MS2-CNOT7 protein bound to cytoplasmic L7Ae mRNA
L2mC	L7Ae protein doubly bound to cytoplasmic MS2-CNOT7 mRNA
C2mL	MS2-CNOT7 protein doubly bound to cytoplasmic L7Ae mRNA
C	MS2-CNOT7 protein
L	L7Ae protein

Reactions:

Transport In this simplified model, the transport of replicons to the plasma membrane and the creation of spherules is assumed to be a first-order process. The transport of replicons into spherules depends on nonstructural proteins and other cellular factors and occurs independently for each replicon. In the replicon case, we also consider the inhibition of replicon transport through RBP binding, where β is a fraction (1 = no inhibition, 0 = complete inhibition).



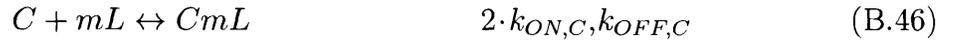
Transcription Transcription is assumed to be first-order upon the formation of spherules (replication factories). Spherules can also transcribe more genomic RNA (Equations 9 and 10). This positive feedback is tuned by the fraction ϵ



Translation Translation is assumed to be first-order as in the pDNA case.



Repressor Binding/Unbinding Second-order association rates and first-order dissociation rates were used as above. In the replicon system we assume RBPs can also bind the genomic RNA with the same efficacy (Equations above).



Degradation First-order degradation rates were assumed as above. We assume that the degradation factor for mRNAs bound by MS2-CNOT7 also applies to genomic replicon RNAs bound by MS2-CNOT7. Spherules are assumed to be stable

for the 4 hours simulated here and are only diluted through cell division.

$$rC \rightarrow 0 \quad \text{deg}_R \quad (\text{B.49})$$

$$rL \rightarrow 0 \quad \text{deg}_R \quad (\text{B.50})$$

$$LrC \rightarrow L \quad \text{deg}_R \quad (\text{B.51})$$

$$CrL \rightarrow C \quad \text{deg}_R \cdot \alpha \quad (\text{B.52})$$

$$LrC \rightarrow rC \quad \text{deg}_P \quad (\text{B.53})$$

$$CrL \rightarrow rL \quad \text{deg}_P \quad (\text{B.54})$$

$$L2rC \rightarrow 2L \quad \text{deg}_R \quad (\text{B.55})$$

$$C2rL \rightarrow 2C \quad \text{deg}_R \cdot 2 \cdot \alpha \quad (\text{B.56})$$

$$L2rC \rightarrow LrC \quad \text{deg}_P \quad (\text{B.57})$$

$$C2rL \rightarrow CrL \quad \text{deg}_P \quad (\text{B.58})$$

$$mC \rightarrow 0 \quad \text{deg}_R \quad (\text{B.59})$$

$$mL \rightarrow 0 \quad \text{deg}_R \quad (\text{B.60})$$

$$LmC \rightarrow L \quad \text{deg}_R \quad (\text{B.61})$$

$$CmL \rightarrow C \quad \text{deg}_R \cdot \alpha \quad (\text{B.62})$$

$$LmC \rightarrow mC \quad \text{deg}_P \quad (\text{B.63})$$

$$CmL \rightarrow mL \quad \text{deg}_P \quad (\text{B.64})$$

$$L2mC \rightarrow 2L \quad \text{deg}_R \quad (\text{B.65})$$

$$C2mL \rightarrow 2C \quad \text{deg}_R \cdot 2 \cdot \alpha \quad (\text{B.66})$$

$$L2mC \rightarrow LmC \quad \text{deg}_P \quad (\text{B.67})$$

$$C2mL \rightarrow CmL \quad \text{deg}_P \quad (\text{B.68})$$

$$C \rightarrow 0 \quad \text{deg}_P \quad (\text{B.69})$$

$$L \rightarrow 0 \quad \text{deg}_P \quad (\text{B.70})$$

Rate constant	Description	Value or range	Units	Source
k_{TS}	Transcription rate	1	min^{-1}	[201]
k_{TS}	Translation rate	8	min^{-1}	[201] [212]
$k_{ON,C}$	MS2 binding rate	4e-6	$\text{molec}^{-1}\text{s}^{-1}$	Assumed = L7Ae
$k_{ON,L}$	L7Ae binding rate	4e-6	$\text{molec}^{-1}\text{s}^{-1}$	[35]
$k_{OFF,C}$	MS2 dissociation rate	0.1	min^{-1}	[213]
$k_{OFF,L}$	L7Ae dissociation rate	0.01	min^{-1}	[35]
deg_R	RNA degradation rate	0.002	min^{-1}	[201]
deg_P	Protein degradation rate	5e-4	min^{-1}	[201]
α	CNOT7 degradation factor	400		this study
σ	L7Ae translational repression factor	3e-3		this study
P_0	Starting pDNA copy number (each)	100	molec	[214]
R_0	Starting replicon copy number (each)	$10^1:10^{2.5}$	molec	[215]
k_{TR}	Replicon transport and RF formation rate	$10^{-2.5}:10^{-1}$	min^{-1}	
β	RF transport inhibition fraction (1=no blocking, 0=complete blocking)	0:1		
ϵ	Genomic fraction of positive synthesized strands	$10^{-3.5}:10^{-2}$		this study

Table B.1: **Theoretical model: reaction rates:** For all calculations involving molar to molecule conversions, the cell volume is assumed to be 3e-12 L.

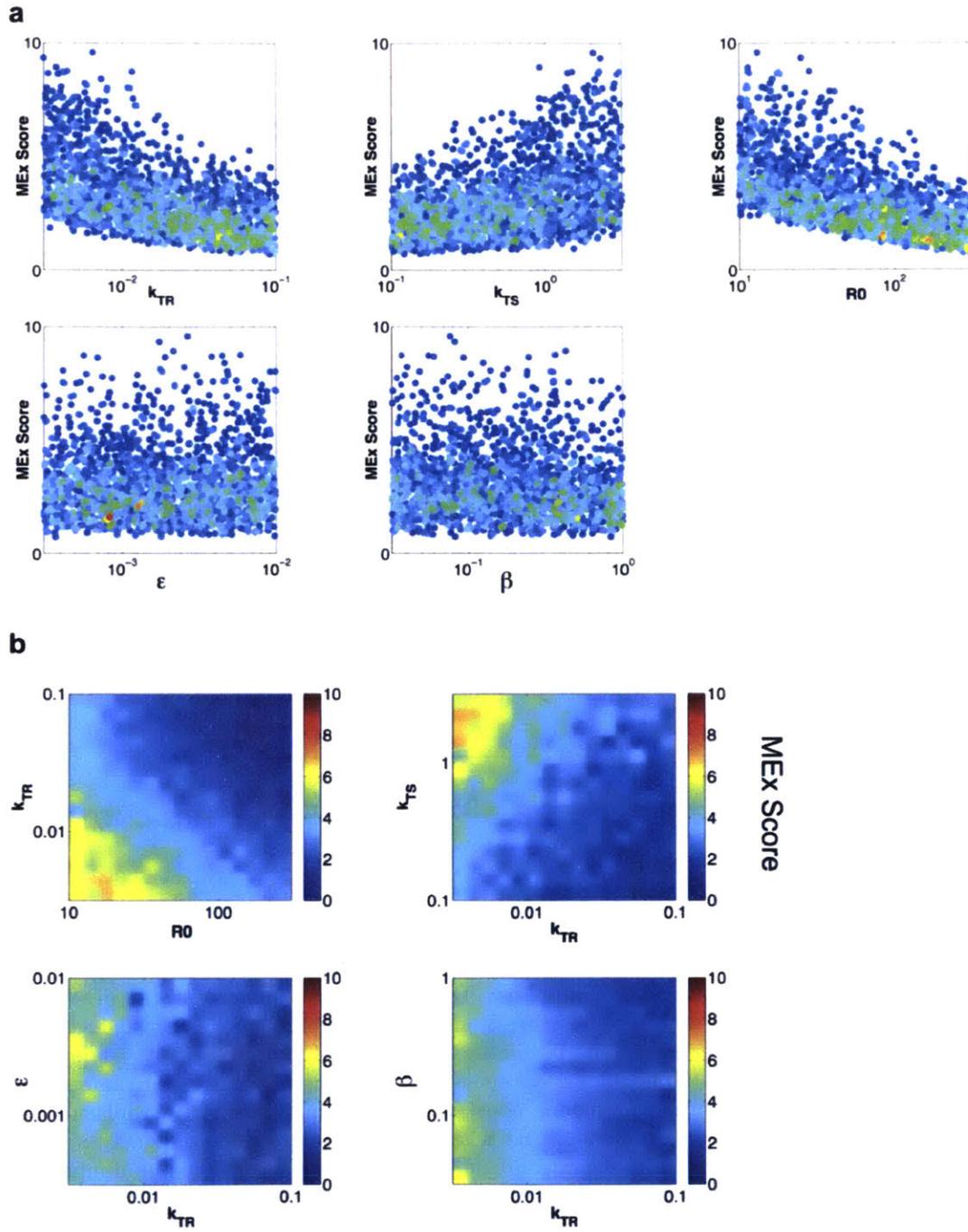
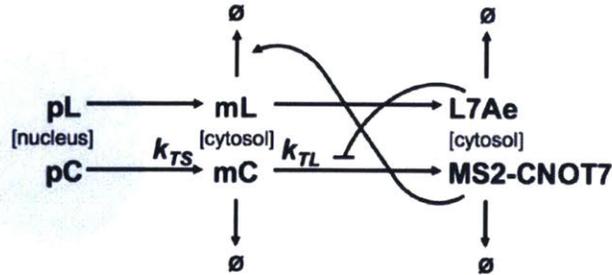
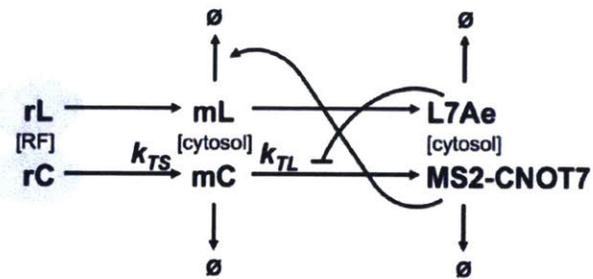


Figure B-2: Replicon system analysis of mutual exclusivity.: (a) MEx scores from distributions of 2000 simulated cell populations plotted for each parameter. Warmer colors indicate higher point density. (b) Heat maps for the identification of parameter interactions, with color intensity indicating MEx score. k_{TR} : transport rate, k_{TS} : transcription rate, $R0$: starting replicon copy number, ϵ : positive feedback, β : replication inhibition.

a pDNA schema



b Replicon schema
> 4 hours:



< 4 hours:

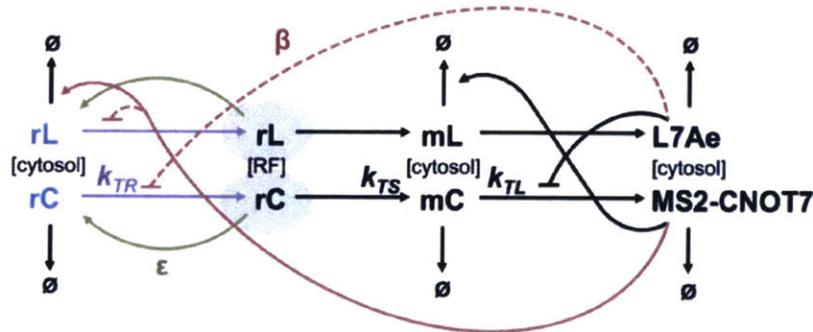


Figure B-3: Theoretical model diagrams: Diagrams of the models for (a) pDNA and (b) replicon systems. The primary species, pL/pC and rL/rC, represent the transfected pDNA or replicons respectively. 'L' pertains to L7Ae-related species and 'C' to MS2-CNOT7-related species (i.e. pL: pDNA L7Ae species). The pDNA system is inert until the plasmids enter the nucleus upon cell division, whereas for the replicon model, nonstructural proteins transport cytoplasmic replicon RNA strands to the plasma membrane (k_{TR}) where they form spherules that act as replication factories ('RF') and become double-stranded. The pDNA in the nucleus and the double-stranded RNA in the spherules both serve as templates for mRNA (mL/mC) transcription (k_{TS}). The mRNAs are translated (k_{TL}) into their respective RBPs, L7Ae or MS2-CNOT7. MS2-CNOT7 binds the L7Ae transcript and increases its degradation rate while L7Ae binds the MS2-CNOT7 transcript and blocks translation. Within the first few hours of the replicon system, while transcribing mRNAs, the spherules also transcribe more of the original genomic RNAs (green arrow, ϵ). Also, the RBPs not only interact with mRNA but with the replicons themselves. MS2-CNOT7 binds the L7Ae replicon and increases its degradation rate (red arrow). We also considered the possibility that binding of RBPs to the replicons can inhibit replication complex formation (red dashed lines, β).

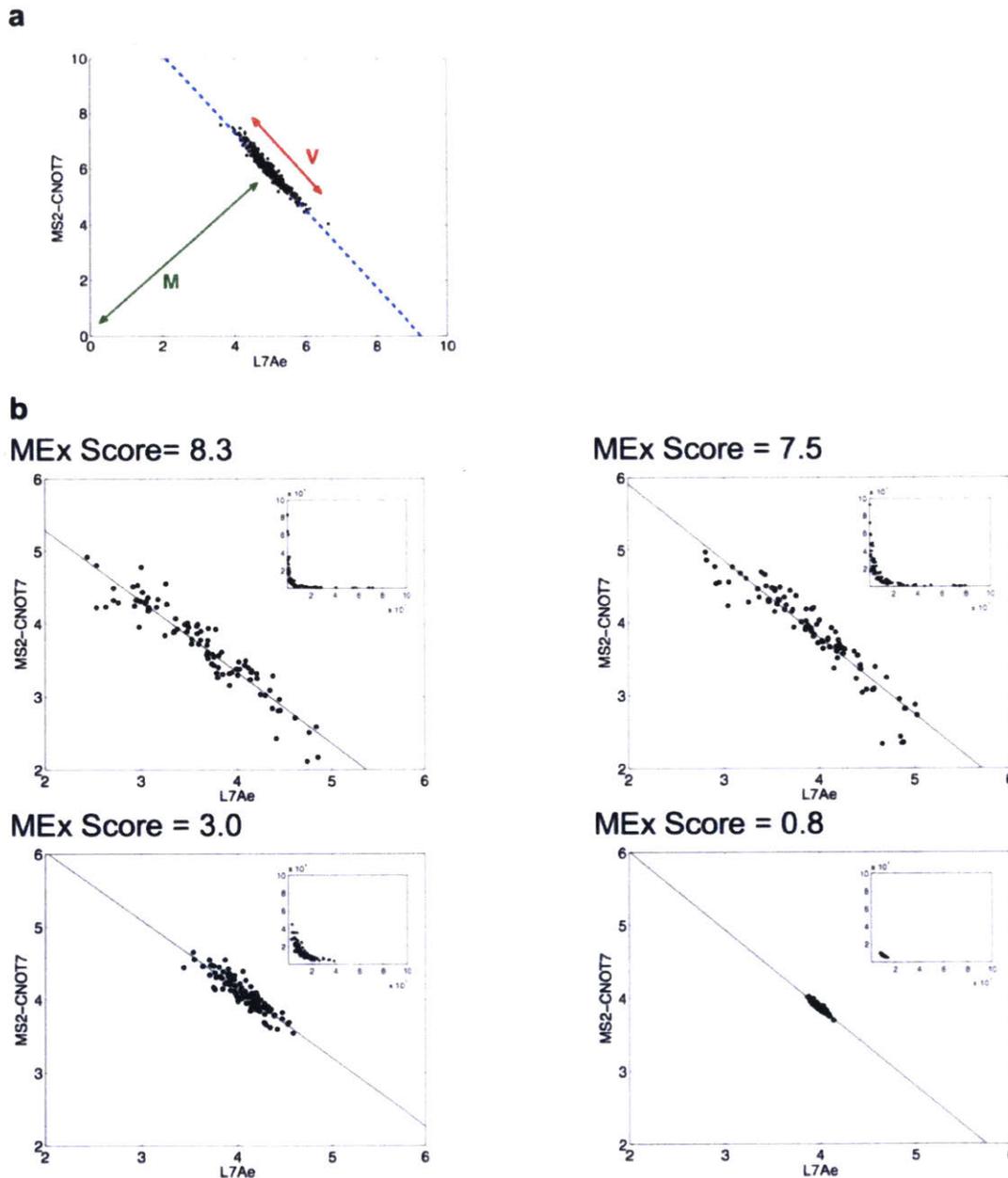


Figure B-4: Mutual Exclusivity (MEx) metric: (a) The MEx score was calculated by fitting the log-transform of the data to a line. The distribution is more mutually exclusive if cells enter the extreme regions of the plot (high L7Ae, low MS2-CNOT7 or low L7Ae, high MS2-CNOT7) and therefore have a large variance along the line (V). Normalizing V by the distance from the origin (M) gives higher scores to distributions with high variance that approach the x and y axes. The MEx score was thus calculated as V/M . Since cells with generally low expression (low transcription rate or low copy number) are close to the origin and receive artificially high scores, we normalized the data to the starting copy number (P0 or R0) and transcription rate (kTS) before performing this analysis. (b) Examples of cell distributions and their MEx scores. Insets show the same data on a linear scale.

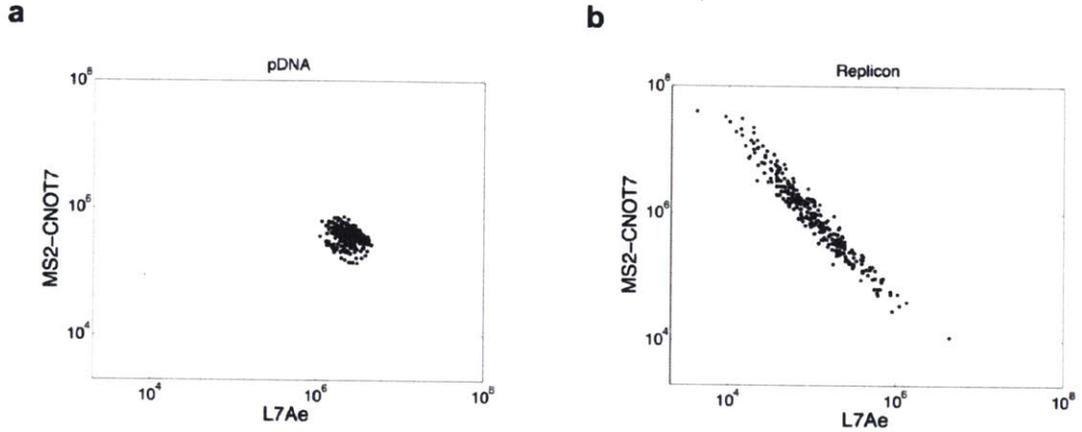


Figure B-5: Comparison of long-term behavior: Simulations of the long-term effects for both the pDNA (a) system and replicon (b) system for the example parameters listed in Table B.1. Both simulations involve 288 cells. The pDNA and replicon models were run for simulation times of 48 hours and 24 hours respectively to correspond to the experimental set-up of Figure B-1 (pDNA and replicon, siRNA Ctrl case).

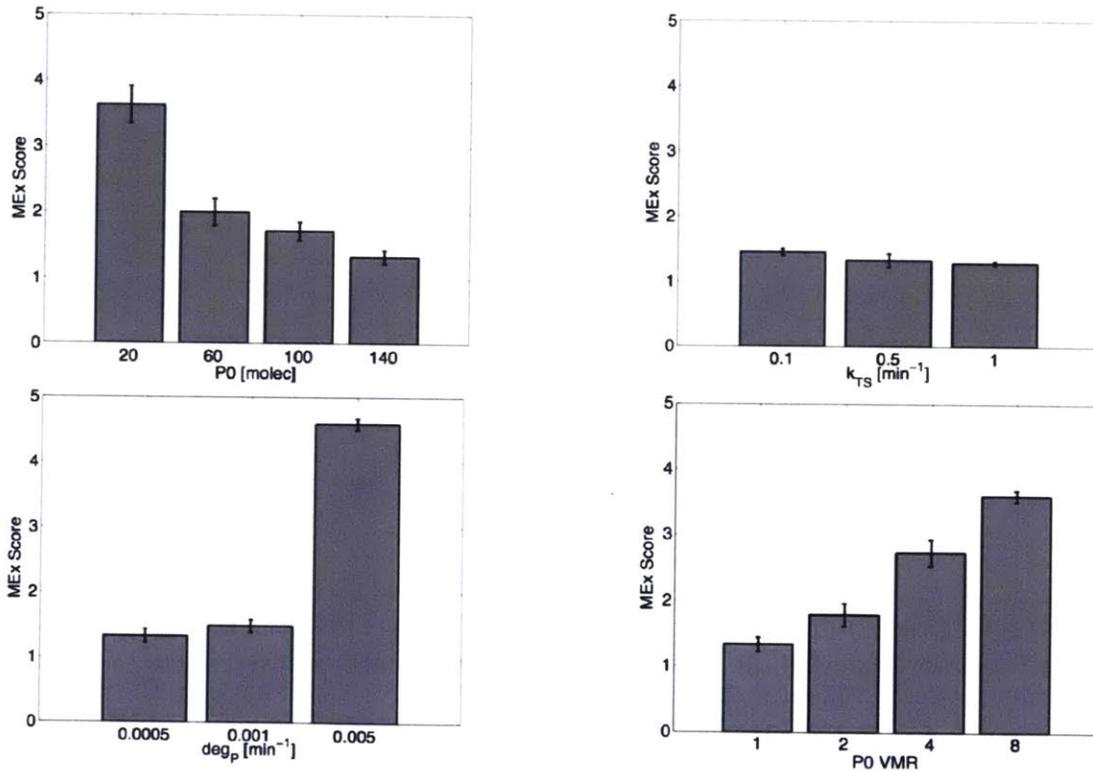


Figure B-6: pDNA system analysis of mutual exclusivity: Parameter perturbations for starting copy number (P0), transcription rate (k_{TS}), protein degradation rate (deg_P), and the starting copy number variance-to-mean ratio (P0 VMR). For each parameter set, 3 simulations were run with 96 cells each. Error bars are one standard deviation. Values from Table B.1 were used for parameters that are unperturbed.

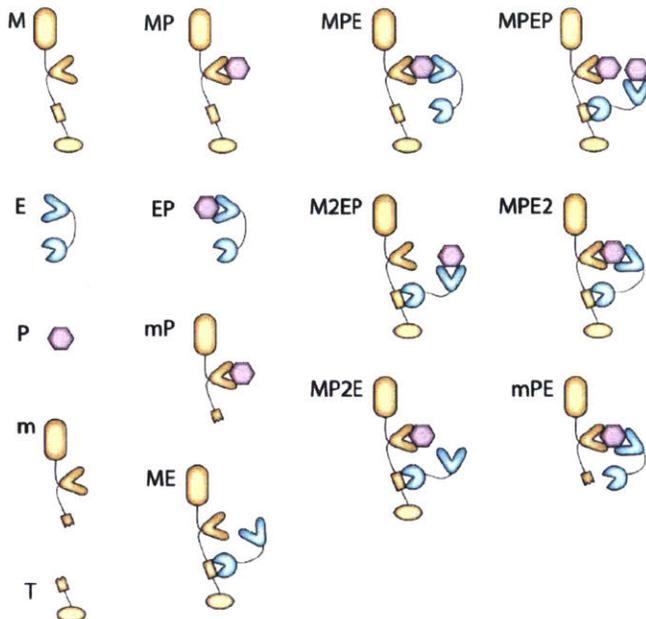


Figure B-7: Protein sensor model species:

B.2 Models for improving protein sensor performance

This section contains work from the following published manuscript and therefore represents a collaboration between myself, Velia Siciliano, Blandine Monel, Jacob Beal, Kiera Clayton, Liliana Wroblewsk, Annmarie McKeon, Bruce Walker and Ron Weiss.

Siciliano, V., DiAndreth, B., Monel, B., Beal, J., Huh, J., Clayton, K., Wroblewska, L., McKeon, A., Walker, B., Weiss, R. (2018) "Engineering modular intracellular protein sensor-actuator devices" *Nature Communications*, 9 (1):1881. doi: 10.1038/s41467-018-03984-5

To investigate the dynamics of the protein sensor topology and to uncover design principles to guide further engineering, we implemented and analyzed a simple computational model of the platform. A system of ordinary, first-order differential equations (ODEs) describing potential interactions of species involved in the protein sensor was constructed and solved using MATLAB [207].

The model includes 15 molecular species (see Figure B-7 and descriptions below), which interact through their various domains: the antibody (Ab) domains of both the membrane-bound sensor part (M) and the cytosolic sensor part (E) can bind the

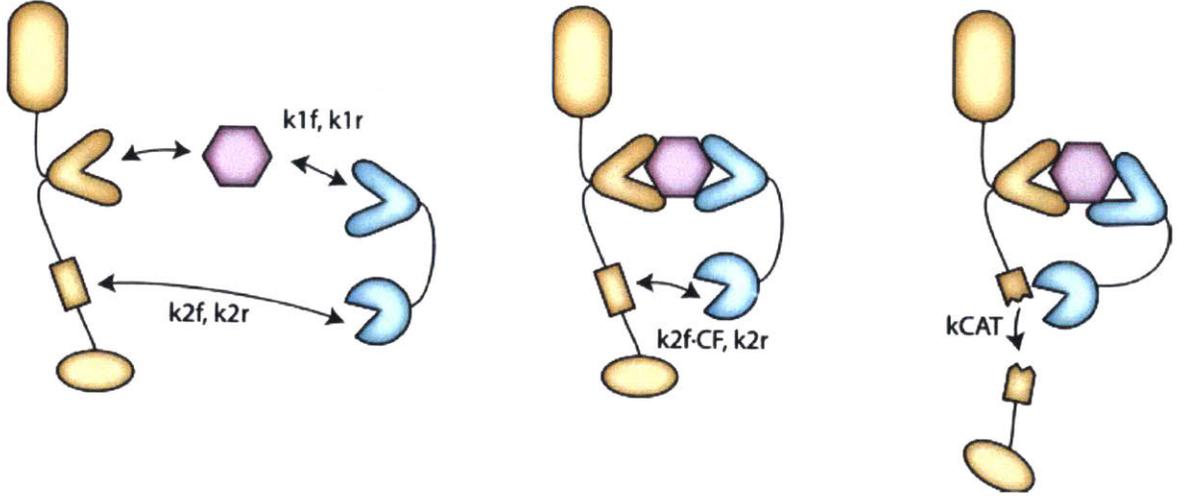


Figure B-8: Protein sensor model reactions: The sensor parts bind the protein of interest reversibly. TEV protease binds the TEV cleavage site (TCS) reversibly followed by irreversible cleavage of the TCS, which releases the transcription factor.

protein of interest (P) reversibly with the same forward and reverse rates, k_{1f} and k_{1r} respectively and dissociation constant K_D :

$$K_D = \frac{k_{1r}}{k_{1f}}$$

The TEV protease domain of E can bind the TEV cleavage site (TCS) of M reversibly with forward and reverse rates, k_{2f} and k_{2r} respectively, followed by irreversible cleavage of the TCS with rate k_{CAT} , releasing the transcription factor, T. These rates can be summarized by the Michaelis constant, K_M :

$$K_M = \frac{k_{2r} + k_{CAT}}{k_{2f}}$$

Reactions are shown in Figure B-8. We assume that when the two parts of the sensor are bound at either the TEV protease cleavage site (TCS) end or the antibody (Ab) end, subsequent binding at the unbound end is enhanced due to increased local effective concentration factor (CF) as described previously [216]. Each simulation was initialized with concentration values (C_0) for each sensor part (M and E). Fold change was calculated at $t=24$ hours by taking the concentration of cleaved transcription factor product, T, for a simulation initialized with the protein of interest, P, divided

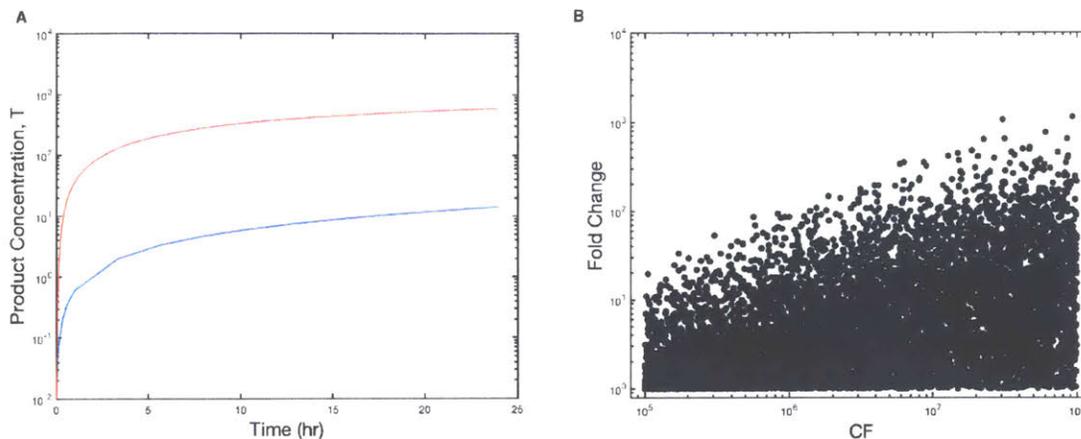


Figure B-9: Protein sensor model set-up: (a) Sample simulation trace for protein of interest = 0 pM (blue) and protein of interest = 1000 pM (red). Fold change is calculated at $t = 24$ hours for subsequent figures. (b) Fold change plots for 10,000 simulations with randomly sampled parameter sets from the log-transformed realistic parameter spaces.

by the concentration of T from a corresponding simulation with $P=0$ (see Figure B-9a). To understand how the sensor is affected by the kinetics of the antibody domains and TEV protease domains as well as molecular concentrations, we simulated various kinetic rate constants and initial concentrations (see Table B.2). The sensor performs better for increased values of binding enhancement, CF, after binding at one part of the sensor (Figure B-9b) and sets an upper limit for achievable fold change. Since the CF factor greatly confounds the results, in subsequent simulations its value was fixed. We also find improved sensor performance when the antibodies bind the protein of interest much stronger than TEV protease binds the TCS, which occurs at lower values of KD and higher values of KM (Figure B-10a). Intuitively, this allows for sensor complex to form only in the presence of the protein and also decreases the rate of cleavage in the case of off-target TEV:TCS binding in the absence of protein. This is in congruence with our experimental results where we find that the TCS with increased KM and slower kCAT (TCS-L) improves sensor performance. The fold change also increases for lower sensor concentrations and for optimal concentrations of protein of interest (Figure B-10b). Our experimental results, which show that lower levels of doxycycline (i.e. lower levels of sensor production) perform best, corroborate this finding. In practice, there is likely a lower limit for kCAT and sensor concentration (C_0) and an upper limit for KM due to background fluorescence. Note that the non-

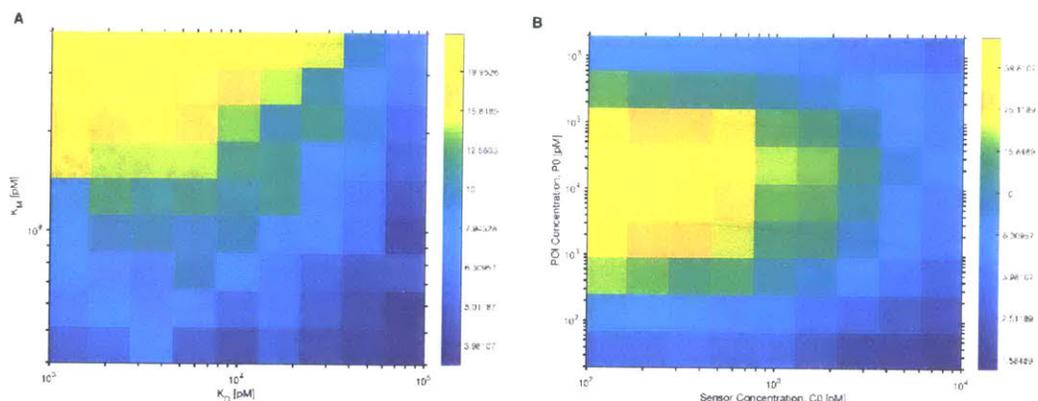


Figure B-10: Heatmaps for the identification of parameter interactions: Color intensity indicating the geometric mean of fold change for each bin. CF was fixed at $2e7$, while all other parameters were randomly sampled for 10,000 simulations.

monotonic response to protein of interest is not surprising due to the bipartite nature of the system. Such behavior has been previously described for reactions with similar topology such as receptor cross-linking [217] and supports our experimental results.

Overall, from these simulations we can draw five main conclusions, which we believe will be valuable in designing new protein sensors based on our platform (summarized in Figure B-11):

1. Careful design of sensor parts is necessary to ensure that binding the protein of interest by both parts greatly enhances binding of TEV to the TCS (i.e. CF is as high as possible).
2. Sensor performance improves when the antibodies bind the protein of interest strongly.
3. Sensor performance generally improves for weaker TEV protease binding and slower TEV protease cleavage rate (kCAT). This binding should be much weaker than binding of the antibodies to the protein of interest.
4. To a certain extent, lower sensor concentrations reduce off-target cleavage and therefore improve sensor performance
5. The sensor's response to protein of interest levels is not monotonic: while likely not physiologically relevant, too much protein can saturate the sensor.

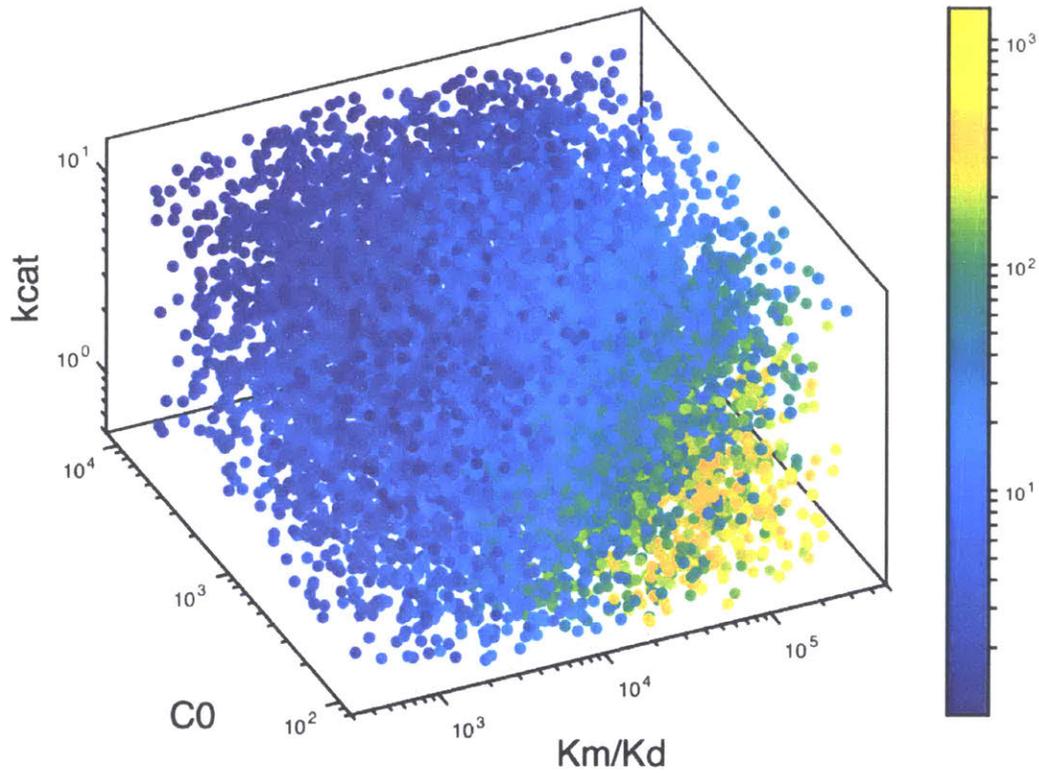


Figure B-11: Parameter effects summary: To elucidate interactions between the remaining engineerable parameters, CF was fixed at 2e7 and P0 was fixed at 1e3 pM, while all other parameters were randomly sampled for 10,000 simulations. Color represents fold change.

Species (see B-7):

M	Membrane-bound sensor part 1 E Cytosolic sensor part 2
P	Protein of interest (POI)
m	Cleaved sensor part 1
T	Cleaved transcription factor (product)
MP	Sensor part 1 bound to POI
EP	Sensor part 2 bound to POI
mP	Cleaved sensor part 1 bound to POI
ME	Sensor part 1 bound to sensor part 2 through TCS
MPE	Sensor part 1 bound to sensor part 2 through POI
M2EP	Sensor part 1 bound to POI-bound sensor part 2 through TCS
MP2E	POI-bound sensor part 1 bound to sensor part 2 through TCS
MPEP	POI-bound sensor part 1 bound to POI-bound part 2 through TCS
MPE2	Sensor part 1 bound to sensor part 2 through both TCS and POI
mPE	Cleaved sensor part 1 bound to sensor part 2 through POI

Reactions

$$\begin{aligned}\frac{dM}{dt} = & -k_{1f} \cdot M \cdot P + k_{1r} \cdot MP - k_{1f} \cdot M \cdot EP + k_{1r} \cdot MPE - k_{2f} \cdot M \cdot E + k_{2r} \cdot ME \\ & - k_{2f} \cdot M \cdot EP + k_{2r} \cdot M2EP\end{aligned}$$

$$\begin{aligned}\frac{dE}{dt} = & -k_{1f} \cdot E \cdot P + k_{1r} \cdot EP - k_{1f} \cdot E \cdot MP + k_{1r} \cdot MP - k_{2f} \cdot E \cdot M + k_{2r} \cdot ME \\ & - k_{2f} \cdot E \cdot MP + k_{2r} \cdot MP2E + k_{CAT} \cdot ME + k_{CAT} \cdot MP2E\end{aligned}$$

$$\begin{aligned}\frac{dP}{dt} = & -k_{1f} \cdot M \cdot P + k_{1r} \cdot MP - k_{1f} \cdot E \cdot P + k_{1r} \cdot EP - k_{1f} \cdot m \cdot P + k_{1r} \cdot mP \\ & - k_{1f} \cdot ME \cdot P + k_{1r} \cdot M2EP - k_{1f} \cdot ME \cdot P + k_{1r} \cdot MP2E - k_{1f} \cdot M2EP \cdot P \\ & + k_{1r} \cdot MPEP - k_{1f} \cdot MP2E \cdot P + k_{1r} \cdot MPEP\end{aligned}$$

$$\begin{aligned}\frac{dm}{dt} = & -k_{1f} \cdot m \cdot P + k_{1r} \cdot mP - k_{1f} \cdot m \cdot EP + k_{1r} \cdot mPE + k_{CAT} \cdot ME \\ & + k_{CAT} \cdot M2EP\end{aligned}$$

$$\begin{aligned}\frac{dT}{dt} = & k_{CAT} \cdot ME + k_{CAT} \cdot M2EP + k_{CAT} \cdot MP2E + k_{CAT} \cdot MPEP \\ & + k_{CAT} \cdot MPE2\end{aligned}$$

$$\begin{aligned}\frac{dMP}{dt} = & k_{1f} \cdot M \cdot P - k_{1r} \cdot MP - k_{1f} \cdot MP \cdot E + k_{1r} \cdot MPE - k_{2f} \cdot MP \cdot E + k_{2r} \cdot MP2E \\ & - k_{2f} \cdot MP \cdot EP + k_{2r} \cdot MPEP\end{aligned}$$

$$\begin{aligned}\frac{dEP}{dt} = & k_{1f} \cdot E \cdot P - k_{1r} \cdot EP - k_{1f} \cdot EP \cdot M + k_{1r} \cdot MPE - k_{1f} \cdot EP \cdot m \\ & + k_{1r} \cdot mPE - k_{2f} \cdot EP \cdot M + k_{2r} \cdot M2EP - k_{2f} \cdot EP \cdot MP + k_{2r} \cdot MPEP + k_{CAT} \cdot M2EP \\ & + k_{CAT} \cdot MPEP\end{aligned}$$

$$\begin{aligned}\frac{dmP}{dt} = & k_{1f} \cdot m \cdot P - k_{1r} \cdot mP - k_{1f} \cdot mP \cdot E + k_{1r} \cdot mPE + k_{CAT} \cdot MP2E \\ & + k_{CAT} \cdot MPEP\end{aligned}$$

$$\begin{aligned}\frac{dME}{dt} = & k_{2f} \cdot M \cdot E - k_{2r} \cdot ME - k_{1f} \cdot ME \cdot P + k_{1r} \cdot M2EP - k_{1f} \cdot ME \cdot P \\ & + k_{1r} \cdot MP2E - k_{CAT} \cdot ME\end{aligned}$$

$$\begin{aligned}\frac{dMPE}{dt} = & k_{1f} \cdot M \cdot EP - k_{1r} \cdot MPE + k_{1f} \cdot MP \cdot E - k_{1r} \cdot MPE - k_{1f} \cdot CF \cdot MPE \\ & + k_{2f} \cdot MPE2\end{aligned}$$

Parameter	Description	Value	Units	Source
K_D	Antibody K_D	1e3:5e5	pM	
K_M	TEV K_M	4e7:4e8	pM	[218]
k_{CAT}	TEV k_{CAT}	0.5:15	min ⁻¹	[218]
k_{1f}, k_{2f}	Association rates	1e-6:1e-4	pM ⁻¹ min ⁻¹	[219]
$C0$	Sensor concentration	1e2:1e4	pM	
$P0$	POI concentration	1e2:2e6	pM	[220]
CF	Enhanced binding factor	1e5:1e8		

Table B.2: **Computational model reaction rates:**

$$\begin{aligned}
\frac{dM2PE}{dt} &= k_{1f} \cdot ME \cdot P - k_{1r} \cdot M2EP + k_{2f} \cdot M \cdot EP - k_{2r} \cdot M2EP - k_{2f} \cdot CF \cdot M2EP \\
&\quad + k_{1r} \cdot MPE2 - k_{1f} \cdot M2EP \cdot P + k_{1r} \cdot MPEP - k_{CAT} \cdot M2EP \\
\frac{dMP2E}{dt} &= k_{1f} \cdot ME \cdot P - k_{1r} \cdot MP2E + k_{2f} \cdot MP \cdot E - k_{2r} \cdot MP2E - k_{1f} \cdot CF \cdot MP2E \\
&\quad + k_{1r} \cdot MPE2 - k_{1f} \cdot MP2E \cdot P + k_{1r} \cdot MPEP - k_{CAT} \cdot MP2E \\
\frac{dMPEP}{dt} &= k_{1f} \cdot M2EP \cdot P - k_{1r} \cdot MPEP + k_{1f} \cdot MP2E \cdot P - k_{1r} \cdot MPEP \\
&\quad + k_{2f} \cdot MP \cdot EP - k_{2r} \cdot MPEP - k_{CAT} \cdot MPEP \\
\frac{dMPE2}{dt} &= k_{1f} \cdot CF \cdot MP2E - k_{2r} \cdot MPE2 + k_{1f} \cdot CF \cdot M2EP - k_{1r} \cdot MPE2 \\
&\quad + k_{2f} \cdot CF \cdot MPE - k_{2r} \cdot MPE2 - k_{CAT} \cdot MPE2 \\
\frac{dmPE}{dt} &= k_{CAT} \cdot MPE2 + k_{1f} \cdot m \cdot EP - k_{1r} \cdot mPE + k_{1f} \cdot mP \cdot E - k_{1r} \cdot mPE
\end{aligned}$$

B.3 Strand Displacement FRET model

The following ODEs describe the strand displacement-based FRET reporter depicted in Figure 5-7 and described in Section 5.1.3

$$\begin{aligned}
\frac{dR_m}{dt} &= -k_{tx} \cdot R_m \\
\frac{dR}{dt} &= k_{tx} \cdot R_m - k_{OFF} \cdot R + k_{ON} \cdot W \cdot S - k_{SD} \cdot R \cdot I
\end{aligned}$$

$$\begin{aligned}\frac{dS}{dt} &= k_{OFF} \cdot R - k_{ON} \cdot W \cdot S + k_{OFF} \cdot C - k_{ON} \cdot S \cdot I \\ \frac{dW}{dt} &= k_{OFF} \cdot R - k_{ON} \cdot W \cdot S + k_{OFF} \cdot C - k_{ON} \cdot S \cdot I \\ \frac{dI}{dt} &= -k_{OFF} \cdot C + k_{ON} \cdot S \cdot I - k_{SD} \cdot R \cdot I \\ \frac{dC}{dt} &= k_{SD} \cdot R \cdot I - k_{OFF} \cdot C + k_{ON} \cdot S \cdot I\end{aligned}$$

B.4 Endogenous reporter model

The following ODEs describe the possible interactions involved in the endogenously transcribed mRNA-based strand displacement reporter depicted in Figures B-12 and B-13 and described in Section 5.2.1.

$$\begin{aligned}\frac{dM}{dt} &= k_{ts} - k_{ON} \cdot M \cdot I + k_{OFF} \cdot MI - \gamma_M \cdot M \\ \frac{dI}{dt} &= k_{ts} - k_{ON} \cdot M \cdot I + k_{OFF} \cdot MI - \gamma_M \cdot I \\ \frac{dMI}{dt} &= k_{ON} \cdot M \cdot I - k_{OFF} \cdot MI - \gamma_M \cdot MI \\ \frac{dP}{dt} &= k_{TL} \cdot M + k_{TL} \cdot MI - \gamma_P \cdot P\end{aligned}$$

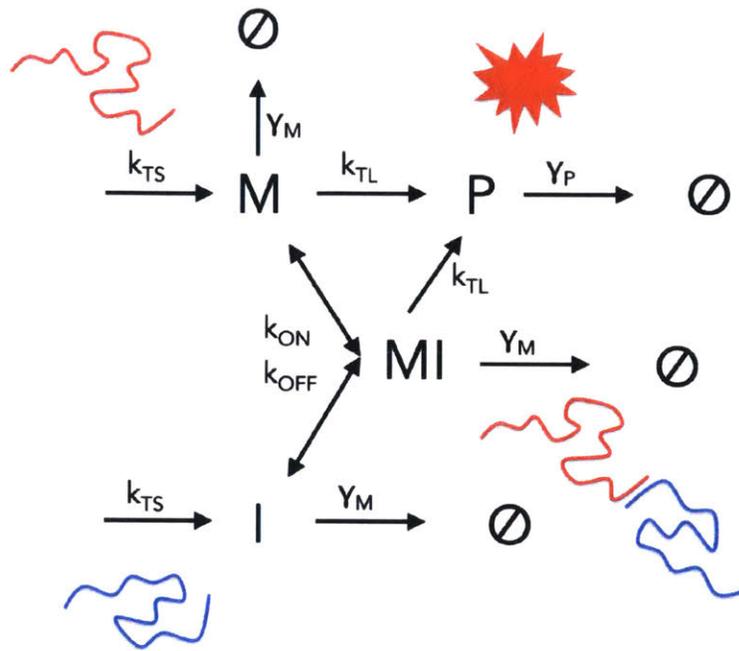


Figure B-12: Endogenous reporter model reactions:

degradation on degradation off translation on translation off

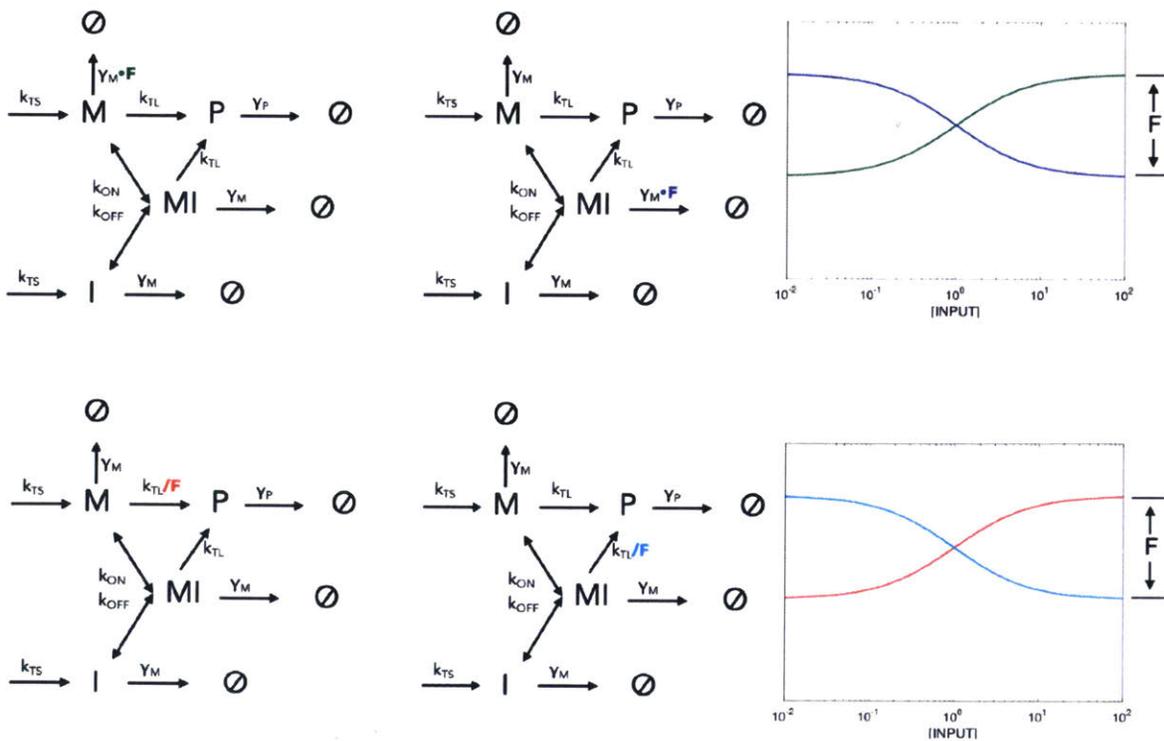


Figure B-13: Endogenous reporter model for various mechanisms:

Bibliography

- [1] M. Cavazzana-calvo, E. Payen, O. Negre, G. Wang, K. Hehir, F. Fusil, J. Down, M. Denaro, T. Brady, K. Westerman, R. Cavallesco, R. Giroto, R. Dorazio, G.-j. Mulder, A. Polack, A. Bank, J. Soulier, N. Cartier, N. Kabbara, B. Dalle, B. Gourmel, Y. Beuzard, and E. Gluckman, “Transfusion independence and HMGA2 activation after gene therapy of human b -thalassaemia,” vol. 467, no. September, pp. 2–7, 2010.
- [2] W. A. Lim and C. H. June, “Review The Principles of Engineering Immune Cells to Treat Cancer,” *Cell*, vol. 168, no. 4, pp. 724–740, 2017.
- [3] L. Naldini, “Gene therapy returns to centre stage,” *Nature*, vol. 526, pp. 351–360, 10 2015.
- [4] A. L. Slusarczyk, A. Lin, and R. Weiss, “Foundations for the design and implementation of synthetic genetic circuits.,” *Nature reviews. Genetics*, vol. 13, pp. 406–20, 6 2012.
- [5] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators.,” *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.
- [6] T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in *Escherichia coli*,” pp. 339–342.
- [7] W. C. Ruder, T. Lu, and J. J. Collins, “Synthetic biology moving into the clinic.,” *Science (New York, N. Y.)*, vol. 333, pp. 1248–52, 9 2011.
- [8] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson, “Multi-input RNAi-based logic circuit for identification of specific cancer cells.,” *Science (New York, N. Y.)*, vol. 333, pp. 1307–11, 9 2011.
- [9] M. Xie, H. Ye, H. Wang, G. Charpin-El Hamri, C. Lormeau, P. Saxena, J. Stelling, and M. Fussenegger, “ β -cell-mimetic designer cells provide closed-loop glycemic control.,” *Science (New York, N. Y.)*, vol. 354, pp. 1296–1301, 12 2016.
- [10] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, “From molecular to modular cell biology,” *Nature*, vol. 402, pp. C47–C52, 12 1999.

- [11] A. D. Ryding, M. G. Sharp, and J. J. Mullins, “Conditional transgenic technologies,” *The Journal of endocrinology*, vol. 171, pp. 1–14, 10 2001.
- [12] Y. Y. Chen, M. C. Jensen, and C. D. Smolke, “Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems,” vol. 107, no. 19, pp. 8531–8536, 2010.
- [13] M. Nizzardo, C. Simone, S. Dametti, S. Salani, G. Ulzi, S. Pagliarani, F. Rizzo, E. Frattini, F. Pagani, N. Bresolin, G. Comi, and S. Corti, “Spinal muscular atrophy phenotype is ameliorated in human motor neurons by SMN increase via different novel RNA therapeutic approaches,” *Scientific Reports*, vol. 5, p. 11746, 12 2015.
- [14] R. S. Finkel, C. A. Chiriboga, J. Vajsar, J. W. Day, J. Montes, D. C. De Vivo, M. Yamashita, F. Rigo, G. Hung, E. Schneider, D. A. Norris, S. Xia, C. F. Bennett, and K. M. Bishop, “Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study,” *The Lancet*, vol. 388, pp. 3017–3026, 12 2016.
- [15] L. Ye, J. Wang, Y. Tan, A. I. Beyer, F. Xie, M. O. Muench, and Y. W. Kan, “Genome editing using CRISPR-Cas9 to create the HPFH genotype in HSPCs: An approach for treating sickle cell disease and β -thalassemia,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. 10661–5, 9 2016.
- [16] E. A. Traxler, Y. Yao, Y.-D. Wang, K. J. Woodard, R. Kurita, Y. Nakamura, J. R. Hughes, R. C. Hardison, G. A. Blobel, C. Li, and M. J. Weiss, “A genome-editing strategy to treat β -hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition,” *Nature Medicine*, vol. 22, pp. 987–990, 9 2016.
- [17] M. Tabebordbar, K. Zhu, J. K. W. Cheng, W. L. Chew, J. J. Widrick, W. X. Yan, C. Maesner, E. Y. Wu, R. Xiao, F. A. Ran, L. Cong, F. Zhang, L. H. Vandenberghe, G. M. Church, and A. J. Wagers, “In vivo gene editing in dystrophic mouse muscle and muscle stem cells,” *Science (New York, N.Y.)*, vol. 351, pp. 407–411, 1 2016.
- [18] C. E. Nelson, C. H. Hakim, D. G. Ousterout, P. I. Thakore, E. A. Moreb, R. M. Castellanos Rivera, S. Madhavan, X. Pan, F. A. Ran, W. X. Yan, A. Asokan, F. Zhang, D. Duan, and C. A. Gersbach, “In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy,” *Science (New York, N.Y.)*, vol. 351, pp. 403–7, 1 2016.
- [19] C. Long, L. Amoasii, A. A. Mireault, J. R. McAnally, H. Li, E. Sanchez-Ortiz, S. Bhattacharyya, J. M. Shelton, R. Bassel-Duby, and E. N. Olson, “Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy,” *Science*, vol. 351, pp. 400–403, 1 2016.

- [20] A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu, “Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage,” *Nature*, vol. 533, pp. 420–424, 5 2016.
- [21] N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, and D. R. Liu, “Programmable base editing of A→T to G→C in genomic DNA without DNA cleavage,” *Nature*, vol. 551, pp. 464–471, 11 2017.
- [22] S. Morita, H. Noguchi, T. Horii, K. Nakabayashi, M. Kimura, K. Okamura, A. Sakai, H. Nakashima, K. Hata, K. Nakashima, and I. Hatada, “Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions,” *Nature Biotechnology*, vol. 34, pp. 1060–1065, 10 2016.
- [23] X. Xu, Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G.-l. Xu, and R. Hu, “A CRISPR-based approach for targeted DNA demethylation,” *Cell Discovery*, vol. 2, p. 16009, 12 2016.
- [24] S. D. Perli, C. H. Cui, and T. K. Lu, “Continuous genetic recording with self-targeting CRISPR-Cas in human cells,” *Science (New York, N.Y.)*, vol. 353, p. aag0511, 9 2016.
- [25] S. Kiani, A. Chavez, M. Tuttle, R. N. Hall, R. Chari, D. Ter-Ovanesyan, J. Qian, B. W. Pruitt, J. Beal, S. Vora, J. Buchthal, E. J. K. Kowal, M. R. Ebrahimkhani, J. J. Collins, R. Weiss, and G. Church, “Cas9 gRNA engineering for genome editing, activation and repression,” *Nature Methods*, vol. 12, pp. 1051–1054, 11 2015.
- [26] K. S. Manning and T. A. Cooper, “The roles of RNA processing in translating genotype to phenotype,” *Nature Reviews Molecular Cell Biology*, vol. 18, pp. 102–114, 2 2017.
- [27] D. Adams, “paper,” in *First European Meeting for ATTR Amyloidosis for Doctors and Patients, Paris, France, 2017*.
- [28] I. Sadowski, J. Ma, S. Triezenberg, and M. Ptashne, “GAL4-VP16 is an unusually potent transcriptional activator,” *Nature*, vol. 335, pp. 563–564, 10 1988.
- [29] F. Zhang, L. Cong, S. Lodato, S. Kosuri, G. M. Church, and P. Arlotta, “Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription,” *Nature Biotechnology*, vol. 29, pp. 149–153, 2 2011.
- [30] J. K. Bialek, G. A. Dunay, M. Voges, C. Schäfer, M. Spohn, R. Stucka, J. Hauber, and U. C. Lange, “Targeted HIV-1 Latency Reversal Using CRISPR/Cas9-Derived Transcriptional Activator Systems,” *PLOS ONE*, vol. 11, p. e0158294, 6 2016.

- [31] O. O. Abudayyeh, J. S. Gootenberg, P. Essletzbichler, S. Han, J. Joung, J. J. Belanto, V. Verdine, D. B. T. Cox, M. J. Kellner, A. Regev, E. S. Lander, D. F. Voytas, A. Y. Ting, and F. Zhang, “RNA targeting with CRISPR-Cas13,” *Nature*, vol. 550, pp. 280–284, 10 2017.
- [32] D. B. T. Cox, J. S. Gootenberg, O. O. Abudayyeh, B. Franklin, M. J. Kellner, J. Joung, and F. Zhang, “RNA editing with CRISPR-Cas13,” *Science*, vol. 358, pp. 1019–1027, 11 2017.
- [33] L. Mazzacurati, M. Marzulli, B. Reinhart, Y. Miyagawa, H. Uchida, W. F. Goins, A. Li, B. Kaur, M. Caligiuri, T. Cripe, N. Chiocca, N. Amankulor, J. B. Cohen, J. C. Glorioso, and P. Grandi, “Use of miRNA Response Sequences to Block Off-target Replication and Increase the Safety of an Unattenuated, Glioblastoma-targeted Oncolytic HSV,” *Molecular Therapy*, vol. 23, pp. 99–107, 1 2015.
- [34] M. S. Kinch, “An overview of FDA-approved biologics medicines,” *Drug Discovery Today*, vol. 20, pp. 393–398, 4 2015.
- [35] H. Saito, T. Kobayashi, T. Hara, Y. Fujita, K. Hayashi, R. Furushima, and T. Inoue, “Synthetic translational regulation by an L7Ae kink-turn RNP switch,” *Nature Chemical Biology*, vol. 6, pp. 71–78, 1 2010.
- [36] J. A. Stapleton, K. Endo, Y. Fujita, K. Hayashi, M. Takinoue, H. Saito, and T. Inoue, “Feedback Control of Protein Expression in Mammalian Cells by Tunable Synthetic Translational Inhibition,” *ACS Synthetic Biology*, vol. 1, pp. 83–88, 3 2012.
- [37] T. Quenault, T. Lithgow, and A. Traven, “PUF proteins: repression, activation and mRNA localization,” *Trends in Cell Biology*, vol. 21, pp. 104–112, 2 2011.
- [38] Y. Yagi, T. Nakamura, and I. Small, “The potential for manipulating RNA with pentatricopeptide repeat proteins,” *The Plant Journal*, vol. 78, pp. 772–782, 6 2014.
- [39] G. Tamulaitis, M. Kazlauskienė, E. Manakova, A. Venclovas, A. Nwokeoji, M. Dickman, P. Horvath, and V. Siksnys, “Programmable RNA Shredding by the Type III-A CRISPR-Cas System of *Streptococcus thermophilus*,” *Molecular Cell*, vol. 56, pp. 506–517, 11 2014.
- [40] R. Staals, Y. Zhu, D. Taylor, J. Kornfeld, K. Sharma, A. Barendregt, J. Koehorst, M. Vlot, N. Neupane, K. Varossieau, K. Sakamoto, T. Suzuki, N. Dohmae, S. Yokoyama, P. Schaap, H. Urlaub, A. Heck, E. Nogales, J. Doudna, A. Shinkai, and J. van der Oost, “RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of *Thermus thermophilus*,” *Molecular Cell*, vol. 56, pp. 518–530, 11 2014.

- [41] M. R. OâÄConnell, B. L. Oakes, S. H. Sternberg, A. East-Seletsky, M. Kaplan, and J. A. Doudna, “Programmable RNA recognition and cleavage by CRISPR/Cas9,” *Nature*, vol. 516, pp. 263–266, 12 2014.
- [42] O. O. Abudayyeh, J. S. Gootenberg, S. Konermann, J. Joung, I. M. Slaymaker, D. B. T. Cox, S. Shmakov, K. S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E. S. Lander, E. V. Koonin, and F. Zhang, “C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.” *Science (New York, N.Y.)*, vol. 353, p. aaf5573, 8 2016.
- [43] A. A. Smargon, D. B. Cox, N. K. Pyzocha, K. Zheng, I. M. Slaymaker, J. S. Gootenberg, O. A. Abudayyeh, P. Essletzbichler, S. Shmakov, K. S. Makarova, E. V. Koonin, and F. Zhang, “Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28,” *Molecular Cell*, vol. 65, pp. 618–630, 2 2017.
- [44] J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, and F. Zhang, “Nucleic acid detection with CRISPR-Cas13a/C2c2,” *Science*, vol. 356, pp. 438–442, 4 2017.
- [45] L. A. Banaszynski, L.-c. Chen, L. A. Maynard-Smith, A. G. L. Ooi, and T. J. Wandless, “A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules,” *Cell*, vol. 126, pp. 995–1004, 9 2006.
- [46] K. M. Bonger, L.-c. Chen, C. W. Liu, and T. J. Wandless, “Small-molecule displacement of a cryptic degron causes conditional protein degradation,” *Nature Chemical Biology*, vol. 7, pp. 531–537, 8 2011.
- [47] H. K. Chung, C. L. Jacobs, Y. Huo, J. Yang, S. A. Krumm, R. K. Plemper, R. Y. Tsien, and M. Z. Lin, “Tunable and reversible drug control of protein production via a self-excising degron,” *Nature Chemical Biology*, vol. 11, pp. 713–720, 9 2015.
- [48] L. A. Banaszynski, M. A. Sellmyer, C. H. Contag, T. J. Wandless, and S. H. Thorne, “Chemical control of protein stability and function in living mice,” *Nature Medicine*, vol. 14, pp. 1123–1127, 10 2008.
- [49] P. Saxena, G. Charpin-El Hamri, M. Folcher, H. Zulewski, and M. Fussenegger, “Synthetic gene network restoring endogenous pituitary-thyroid feedback control in experimental Graves’ disease.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, pp. 1244–9, 2 2016.
- [50] C.-Y. Wu, K. T. Roybal, E. M. Puchner, J. Onuffer, and W. A. Lim, “Remote control of therapeutic T cells through a small molecule-gated chimeric receptor,” *Science*, vol. 350, pp. aab4077–aab4077, 10 2015.

- [51] P. Saxena, B. C. Heng, P. Bai, M. Folcher, H. Zulewski, and M. Fussenegger, “A programmable synthetic lineage-control network that differentiates human iPSCs into glucose-sensitive insulin-secreting beta-like cells,” *Nature Communications*, vol. 7, p. 11247, 12 2016.
- [52] U. S. Bhalla and R. Iyengar, “Emergent Properties of Networks of Biological Signaling Pathways,” *Science*, vol. 283, pp. 381–387, 1 1999.
- [53] J. Meng, F. Muntoni, and J. E. Morgan, “Stem cells to treat muscular dystrophies — Where are we?,” *Neuromuscular Disorders*, vol. 21, pp. 4–12, 1 2011.
- [54] J. N. Kochenderfer, M. E. Dudley, S. A. Feldman, W. H. Wilson, D. E. Spaner, I. Maric, M. Stetler-Stevenson, G. Q. Phan, M. S. Hughes, R. M. Sherry, J. C. Yang, U. S. Kammula, L. Devillier, R. Carpenter, D.-A. N. Nathan, R. A. Morgan, C. Laurencot, and S. A. Rosenberg, “B-cell depletion and remissions of malignancy along with cytokine-associated toxicity in a clinical trial of anti-CD19 chimeric-antigen-receptor-transduced T cells,” *Blood*, vol. 119, pp. 2709–20, 3 2012.
- [55] K. Roybal, L. Rupp, L. Morsut, W. Walker, K. McNally, J. Park, and W. Lim, “Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits,” *Cell*, vol. 164, pp. 770–779, 2 2016.
- [56] L. Morsut, K. Roybal, X. Xiong, R. Gordley, S. Coyle, M. Thomson, and W. Lim, “Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors,” *Cell*, vol. 164, pp. 780–791, 2 2016.
- [57] C. A. Klebanoff, S. A. Rosenberg, and N. P. Restifo, “Prospects for gene-engineered T cell immunotherapy for solid cancers,” *Nature Medicine*, vol. 22, pp. 26–36, 1 2016.
- [58] K. T. Roybal, J. Z. Williams, L. Morsut, L. J. Rupp, I. Kolinko, J. H. Choe, W. J. Walker, K. A. McNally, and W. A. Lim, “Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors,” *Cell*, vol. 167, pp. 419–432, 10 2016.
- [59] N. Lapique and Y. Benenson, “Digital switching in a biosensor circuit via programmable timing of gene availability,” *Nature Chemical Biology*, vol. 10, pp. 1020–1027, 12 2014.
- [60] Y. Li, Y. Jiang, H. Chen, W. Liao, Z. Li, R. Weiss, and Z. Xie, “Modular construction of mammalian gene circuits using TALE transcriptional repressors,” *Nature Chemical Biology*, vol. 11, pp. 207–213, 3 2015.
- [61] L. Wroblewska, T. Kitada, K. Endo, V. Siciliano, B. Stillo, H. Saito, and R. Weiss, “Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery,” *Nature Biotechnology*, vol. 33, no. 8, 2015.

- [62] G. Mullokandov, A. Baccarini, A. Ruzo, A. D. Jayaprakash, N. Tung, B. Israelow, M. J. Evans, R. Sachidanandam, and B. D. Brown, “High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries,” *Nature Methods*, vol. 9, pp. 840–846, 8 2012.
- [63] B. C. Heng, D. Aubel, and M. Fussenegger, “Prosthetic gene networks as an alternative to standard pharmacotherapies for metabolic disorders,” *Current Opinion in Biotechnology*, vol. 35, pp. 37–45, 12 2015.
- [64] L. J. Aronne, A. E. Halseth, C. M. Burns, S. Miller, and L. Z. Shen, “Enhanced Weight Loss Following Coadministration of Pramlintide With Sibutramine or Phentermine in a Multicenter Trial,” *Obesity*, vol. 18, pp. 1739–1746, 9 2010.
- [65] K. Rössger, G. Charpin-El-Hamri, and M. Fussenegger, “A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice,” *Nature Communications*, vol. 4, p. 2825, 12 2013.
- [66] C. Kemmer, M. Gitzinger, M. Daoud-El Baba, V. Djonov, J. Stelling, and M. Fussenegger, “Self-sufficient control of urate homeostasis in mice by a synthetic circuit,” *Nature Biotechnology*, vol. 28, pp. 355–360, 4 2010.
- [67] K. Rössger, G. Charpin-El Hamri, and M. Fussenegger, “Reward-based hypertension control by a synthetic brain-dopamine interface,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 18150–5, 11 2013.
- [68] D. Ausländer, S. Ausländer, G. Charpin-El Hamri, F. Sedlmayer, M. Müller, O. Frey, A. Hierlemann, J. Stelling, and M. Fussenegger, “A Synthetic Multifunctional Mammalian pH Sensor and CO₂ Transgene-Control Device,” *Molecular Cell*, vol. 55, pp. 397–408, 8 2014.
- [69] P. Bai, H. Ye, M. Xie, P. Saxena, H. Zulewski, G. Charpin-El Hamri, V. Djonov, and M. Fussenegger, “A synthetic biology-based device prevents liver injury in mice,” *Journal of Hepatology*, vol. 65, pp. 84–94, 7 2016.
- [70] H. Ye, M. Xie, S. Xue, G. C.-E. Hamri, J. Yin, H. Zulewski, and M. Fussenegger, “Self-adjusting synthetic gene circuit for correcting insulin resistance,” *Nature Biomedical Engineering*, vol. 1, p. 0005, 12 2016.
- [71] W.-H. Boehncke and M. P. Schön, “Psoriasis,” *The Lancet*, vol. 386, pp. 983–994, 9 2015.
- [72] A. Menter, S. K. Tyring, K. Gordon, A. B. Kimball, C. L. Leonardi, R. G. Langley, B. E. Strober, M. Kaul, Y. Gu, M. Okun, and K. Papp, “Adalimumab therapy for moderate to severe psoriasis: A randomized, controlled phase III trial,” *Journal of the American Academy of Dermatology*, vol. 58, pp. 106–115, 1 2008.

- [73] K. Ghoreschi, P. Thomas, S. Breit, M. Dugas, R. Mailhammer, W. van Eden, R. van der Zee, T. Biedermann, J. Prinz, M. Mack, U. Mrowietz, E. Christophers, D. Schlöndorff, G. Plewig, C. A. Sander, and M. Röcken, “Interleukin-4 therapy of psoriasis induces Th2 responses and improves human autoimmune disease,” *Nature Medicine*, vol. 9, pp. 40–46, 1 2003.
- [74] K. Asadullah, W.-D. Döcke, M. Ebeling, M. Friedrich, G. Belbe, H. Audring, H.-D. Volk, and W. Sterry, “Interleukin 10 Treatment of Psoriasis,” *Archives of Dermatology*, vol. 135, pp. 187–192, 2 1999.
- [75] L. Schukur, B. Geering, G. Charpin-El Hamri, and M. Fussenegger, “Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis,” *Science translational medicine*, vol. 7, p. 318ra201, 12 2015.
- [76] J. A. N. Brophy and C. A. Voigt, “Principles of genetic circuit design,” *Nature Methods*, vol. 11, pp. 508–520, 5 2014.
- [77] D. Del Vecchio, “Modularity, context-dependence, and insulation in engineered biological circuits,” *Trends in Biotechnology*, vol. 33, pp. 111–119, 2 2015.
- [78] C. Pan, C. Kumar, S. Bohl, U. Klingmueller, and M. Mann, “Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions,” *Molecular & cellular proteomics : MCP*, vol. 8, pp. 443–50, 3 2009.
- [79] H. Clevers, “Modeling Development and Disease with Organoids,” *Cell*, vol. 165, pp. 1586–1597, 6 2016.
- [80] E. W. Esch, A. Bahinski, and D. Huh, “Organs-on-chips at the frontiers of drug discovery,” *Nature Reviews Drug Discovery*, vol. 14, pp. 248–260, 4 2015.
- [81] M. Ogawa, S. Ogawa, C. E. Bear, S. Ahmadi, S. Chin, B. Li, M. Grompe, G. Keller, B. M. Kamath, and A. Ghanekar, “Directed differentiation of cholangiocytes from human pluripotent stem cells,” *Nature Biotechnology*, vol. 33, pp. 853–861, 8 2015.
- [82] L. Schukur, B. Geering, and M. Fussenegger, “Human whole-blood culture system for ex vivo characterization of designer-cell function,” *Biotechnology and Bioengineering*, vol. 113, pp. 588–597, 3 2016.
- [83] J. T. MacDonald, C. Barnes, R. I. Kitney, P. S. Freemont, and G.-B. V. Stan, “Computational design approaches and tools for synthetic biology,” *Integrative Biology*, vol. 3, p. 97, 2 2011.
- [84] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt, “Genetic circuit design automation,” *Science*, vol. 352, pp. aac7341–aac7341, 4 2016.

- [85] N. Davidsohn, J. Beal, S. Kiani, A. Adler, F. Yaman, Y. Li, Z. Xie, and R. Weiss, “Accurate Predictions of Genetic Circuit Behavior from Part Characterization and Modular Composition,” *ACS Synthetic Biology*, vol. 4, pp. 673–681, 6 2015.
- [86] J.-Y. Dong, P.-D. Fan, and R. A. Frizzell, “Quantitative Analysis of the Packaging Capacity of Recombinant Adeno-Associated Virus,” *Human Gene Therapy*, vol. 7, pp. 2101–2112, 11 1996.
- [87] A. L. Epstein, “Progress and prospects: Biological properties and technological advances of herpes simplex virus type 1-based amplicon vectors,” *Gene Therapy*, vol. 16, pp. 709–715, 6 2009.
- [88] H. Yin, R. L. Kanasty, A. A. Eltoukhy, A. J. Vegas, J. R. Dorkin, and D. G. Anderson, “Non-viral vectors for gene-based therapy,” *Nature Reviews Genetics*, vol. 15, pp. 541–555, 8 2014.
- [89] U. Sahin, K. Karikó, and A. Türeci, “mRNA-based therapeutics – developing a new class of drugs,” *Nature Reviews Drug Discovery*, vol. 13, pp. 759–780, 10 2014.
- [90] M. A. Maier, M. Jayaraman, S. Matsuda, J. Liu, S. Barros, W. Querbes, Y. K. Tam, S. M. Ansell, V. Kumar, J. Qin, X. Zhang, Q. Wang, S. Panesar, R. Hutabarat, M. Carioto, J. Hettlinger, P. Kandasamy, D. Butler, K. G. Rajeev, B. Pang, K. Charisse, K. Fitzgerald, B. L. Mui, X. Du, P. Cullis, T. D. Madden, M. J. Hope, M. Manoharan, and A. Akinc, “Biodegradable Lipids Enabling Rapidly Eliminated Lipid Nanoparticles for Systemic Delivery of RNAi Therapeutics,” *Molecular Therapy*, vol. 21, pp. 1570–1578, 8 2013.
- [91] J. Hoggatt, “Gene Therapy for ‘Bubble Boy’ Disease,” *Cell*, vol. 166, p. 263, 7 2016.
- [92] C. Kloks, C. Berger, P. Cortez, Y. Dean, J. Heinrich, L. Bjerring Jensen, V. Koppenburg, S. Kostense, D. Kramer, S. Spindeldreher, and H. Kirby, “A fit-for-purpose strategy for the risk-based immunogenicity testing of biotherapeutics: a European industry perspective,” *Journal of Immunological Methods*, vol. 417, pp. 1–9, 2 2015.
- [93] A. C. Nathwani, E. G. Tuddenham, S. Rangarajan, C. Rosales, J. McIntosh, D. C. Linch, P. Chowdary, A. Riddell, A. J. Pie, C. Harrington, J. O’Beirne, K. Smith, J. Pasi, B. Glader, P. Rustagi, C. Y. Ng, M. A. Kay, J. Zhou, Y. Spence, C. L. Morton, J. Allay, J. Coleman, S. Sleep, J. M. Cunningham, D. Srivastava, E. Basner-Tschakarjan, F. Mingozzi, K. A. High, J. T. Gray, U. M. Reiss, A. W. Nienhuis, and A. M. Davidoff, “Adenovirus-Associated Virus Vector-Mediated Gene Transfer in Hemophilia B,” *New England Journal of Medicine*, vol. 365, pp. 2357–2365, 12 2011.
- [94] A. C. Nathwani, U. M. Reiss, E. G. Tuddenham, C. Rosales, P. Chowdary, J. McIntosh, M. Della Peruta, E. Lheriteau, N. Patel, D. Raj, A. Riddell,

- J. Pie, S. Rangarajan, D. Bevan, M. Recht, Y.-M. Shen, K. G. Halka, E. Basner-Tschakarjan, F. Mingozzi, K. A. High, J. Allay, M. A. Kay, C. Y. Ng, J. Zhou, M. Cancio, C. L. Morton, J. T. Gray, D. Srivastava, A. W. Nienhuis, and A. M. Davidoff, “Long-Term Safety and Efficacy of Factor IX Gene Therapy in Hemophilia B,” *New England Journal of Medicine*, vol. 371, pp. 1994–2004, 11 2014.
- [95] R. A. Maldonado, R. A. LaMothe, J. D. Ferrari, A.-H. Zhang, R. J. Rossi, P. N. Kolte, A. P. Griset, C. O’Neil, D. H. Altreuter, E. Browning, L. Johnston, O. C. Farokhzad, R. Langer, D. W. Scott, U. H. von Andrian, and T. K. Kishimoto, “Polymeric synthetic nanoparticles for the induction of antigen-specific immunological tolerance.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, pp. 156–65, 1 2015.
- [96] T. K. Kishimoto, J. D. Ferrari, R. A. LaMothe, P. N. Kolte, A. P. Griset, C. O’Neil, V. Chan, E. Browning, A. Chalishazar, W. Kuhlman, F.-n. Fu, N. Viseux, D. H. Altreuter, L. Johnston, and R. A. Maldonado, “Improving the efficacy and safety of biologic drugs with tolerogenic nanoparticles,” *Nature Nanotechnology*, vol. 11, pp. 890–899, 10 2016.
- [97] E. Sands, “No Title,” in *Annual European Congress of Rheumatology, Madrid, Spain*, 0.
- [98] R. M. Blaese, K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan, and W. F. Anderson, “T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years.,” *Science (New York, N.Y.)*, vol. 270, pp. 475–80, 10 1995.
- [99] A. Aiuti, M. G. Roncarolo, and L. Naldini, “Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products,” *EMBO Molecular Medicine*, vol. 9, pp. 737–740, 6 2017.
- [100] E. Dolgin, “Epic \$12 billion deal and FDA’s approval raise CAR-T to new heights,” *Nature Biotechnology*, vol. 35, pp. 891–892, 10 2017.
- [101] J. Yu, K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, I. I. Slukvin, and J. A. Thomson, “Human induced pluripotent stem cells free of vector and transgene sequences.,” *Science (New York, N.Y.)*, vol. 324, pp. 797–801, 5 2009.
- [102] L. Warren, P. D. Manos, T. Ahfeldt, Y.-H. Loh, H. Li, F. Lau, W. Ebina, P. K. Mandal, Z. D. Smith, A. Meissner, G. Q. Daley, A. S. Brack, J. J. Collins, C. Cowan, T. M. Schlaeger, and D. J. Rossi, “Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA,” *Cell Stem Cell*, vol. 7, pp. 618–630, 11 2010.

- [103] E. P. Papapetrou, M. J. Tomishima, S. M. Chambers, Y. Mica, E. Reed, J. Menon, V. Tabar, Q. Mo, L. Studer, and M. Sadelain, “Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 12759–64, 8 2009.
- [104] G. Nagamatsu, S. Saito, T. Kosaka, K. Takubo, T. Kinoshita, M. Oya, K. Horimoto, and T. Suda, “Optimal ratio of transcription factors for somatic cell reprogramming.,” *The Journal of biological chemistry*, vol. 287, pp. 36273–82, 10 2012.
- [105] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, “Multiplex Genome Engineering Using CRISPR/Cas Systems,” *Science*, vol. 339, pp. 819–823, 2 2013.
- [106] Y.-L. Min, H. Li, C. Rodriguez-Caycedo, A. A. Mireault, J. Huang, J. M. Shelton, J. R. McAnally, L. Amoasii, P. P. A. Mammen, R. Bassel-Duby, and E. N. Olson, “CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells,” *Science Advances*, vol. 5, p. eaav4324, 3 2019.
- [107] W. Fujii, K. Kawasaki, K. Sugiura, and K. Naito, “Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease,” *Nucleic Acids Research*, vol. 41, pp. e187–e187, 11 2013.
- [108] Y. Hori, C. Kantak, R. M. Murray, and A. R. Abate, “Cell-free extract based optimization of biomolecular circuits with droplet microfluidics,” *Lab on a Chip*, vol. 17, pp. 3037–3042, 9 2017.
- [109] S. Kosuri, D. B. Goodman, G. Cambray, V. K. Mutalik, Y. Gao, A. P. Arkin, D. Endy, and G. M. Church, “Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 14024–9, 8 2013.
- [110] L. Woodruff, T. E. Goroehowski, N. Roehner, T. S. Mikkelsen, D. Densmore, D. Gordon, R. Nicol, and C. A. Voigt, “Registry in a tube: multiplexed pools of retrievable parts for genetic design space exploration,” *Nucleic Acids Research*, vol. 45, p. gkw1226, 12 2016.
- [111] A. Ghodasara and C. A. Voigt, “Balancing gene expression without library construction via a reusable sRNA pool,” *Nucleic Acids Research*, vol. 45, pp. 8116–8127, 7 2017.
- [112] K. Woodruff and S. J. Maerkl, “A High-Throughput Microfluidic Platform for Mammalian Cell Transfection and Culturing,” *Scientific Reports*, vol. 6, p. 23937, 7 2016.

- [113] Y. Li, Y. Jiang, H. Chen, W. Liao, Z. Li, R. Weiss, and Z. Xie, “Modular construction of mammalian gene circuits using TALE transcriptional repressors,” *Nature Chemical Biology*, vol. 11, no. february, pp. 1–10, 2015.
- [114] A. S. L. Wong, G. C. G. Choi, A. A. Cheng, O. Purcell, and T. K. Lu, “Massively parallel high-order combinatorial genetics in human cells,” *Nature Biotechnology*, vol. 33, pp. 952–961, 9 2015.
- [115] A. S. L. Wong, G. C. G. Choi, C. H. Cui, G. Pregonig, P. Milani, M. Adam, S. D. Perli, S. W. Kazer, A. Gaillard, M. Hermann, A. K. Shalek, E. Fraenkel, and T. K. Lu, “Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM,” *Proceedings of the National Academy of Sciences*, vol. 113, pp. 2544–2549, 3 2016.
- [116] D. S. Kong, T. A. Thorsen, J. Babb, S. T. Wick, J. J. Gam, R. Weiss, and P. A. Carr, “Open-source, community-driven microfluidics with Metafluidics,” *Nature Biotechnology*, vol. 35, pp. 523–529, 6 2017.
- [117] G. Schwake, S. Youssef, J.-T. Kuhr, S. Gude, M. P. David, E. Mendoza, E. Frey, and J. O. Rädler, “Predictive modeling of non-viral gene transfer,” *Biotechnology and Bioengineering*, vol. 105, pp. n/a–n/a, 3 2010.
- [118] N. S. Bhise, R. B. Shmueli, J. Gonzalez, and J. J. Green, “A Novel Assay for Quantifying the Number of Plasmids Encapsulated by Polymer Nanoparticles,” *Small*, vol. 8, pp. 367–373, 2 2012.
- [119] M. Xie and M. Fussenegger, “Designing cell function: assembly of synthetic gene circuits for cell biology applications,” *Nature Reviews Molecular Cell Biology*, vol. 19, pp. 507–525, 8 2018.
- [120] T. S. Hawley, R. G. Hawley, and W. G. Telford, “Fluorescent Proteins for Flow Cytometry,” in *Current Protocols in Cytometry*, vol. 80, pp. 1–9, Hoboken, NJ, USA: John Wiley & Sons, Inc., 4 2017.
- [121] A. A. Dominguez, W. A. Lim, and L. S. Qi, “Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation,” *Nature Reviews Molecular Cell Biology*, vol. 17, pp. 5–15, 1 2016.
- [122] A. Chavez, J. Scheiman, S. Vora, B. W. Pruitt, M. Tuttle, E. P. R. Iyer, S. Lin, S. Kiani, C. D. Guzman, D. J. Wiegand, D. Ter-Ovanesyan, J. L. Braff, N. Davidsohn, B. E. Housden, N. Perrimon, R. Weiss, J. Aach, J. J. Collins, and G. M. Church, “Highly efficient Cas9-mediated transcriptional programming,” *Nature Methods*, vol. 12, pp. 326–328, 4 2015.
- [123] L. Wang, K. Qian, Y. Huang, N. Jin, H. Lai, T. Zhang, C. Li, C. Zhang, X. Bi, D. Wu, C. Wang, H. Wu, P. Tan, J. Lu, L. Chen, K. Li, X. Li, and D. Wang, “SynBioLGDB: a resource for experimentally validated logic gates in synthetic biology,” *Scientific Reports*, vol. 5, no. Figure 2, p. 8090, 2015.

- [124] B. D. Brown, B. Gentner, A. Cantore, S. Colleoni, M. Amendola, A. Zingale, A. Baccarini, G. Lazzari, C. Galli, and L. Naldini, “Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state,” *Nature Biotechnology*, vol. 25, pp. 1457–1467, 12 2007.
- [125] K. Shimizu, F. Sakurai, K. Tomita, Y. Nagamoto, S.-I. Nakamura, K. Katayama, M. Tachibana, K. Kawabata, and H. Mizuguchi, “Suppression of leaky expression of adenovirus genes by insertion of microRNA-targeted sequences in the replication-incompetent adenovirus vector genome.,” *Molecular therapy. Methods & clinical development*, vol. 1, p. 14035, 1 2014.
- [126] G. Lillacci, Y. Benenson, and M. Khammash, “Synthetic control systems for high performance gene expression in mammalian cells,” *Nucleic Acids Research*, vol. 46, pp. 9855–9863, 10 2018.
- [127] S. Mukherji, M. S. Ebert, G. X. Y. Zheng, J. S. Tsang, P. A. Sharp, and A. van Oudenaarden, “MicroRNAs can generate thresholds in target gene expression,” *Nature Genetics*, vol. 43, pp. 854–859, 9 2011.
- [128] J. J. Gam, J. Babb, and R. Weiss, “A mixed antagonistic/synergistic miRNA repression model enables accurate predictions of multi-input miRNA sensor activity,” *Nature Communications*, vol. 9, p. 2430, 12 2018.
- [129] M. D. Bullock, K. M. Pickard, B. S. Nielsen, A. E. Sayan, V. Jenei, M. Mellone, R. Mitter, J. N. Primrose, G. J. Thomas, G. K. Packham, and A. H. Mirnezami, “Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression,” *Cell Death & Disease*, vol. 4, pp. e684–e684, 6 2013.
- [130] J. P. Ferreira, R. W. S. Peacock, I. E. B. Lawhorn, and C. L. Wang, “Modulating ectopic gene expression levels by using retroviral vectors equipped with synthetic promoters,” *Systems and Synthetic Biology*, vol. 5, pp. 131–138, 12 2011.
- [131] J. P. Ferreira, K. W. Overton, and C. L. Wang, “Tuning gene expression with synthetic upstream open reading frames.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 11284–9, 7 2013.
- [132] L. Ordovás, R. Boon, M. Pistoni, Y. Chen, E. Wolfs, W. Guo, R. Sambathkumar, S. Bobis-Wozowicz, N. Helsen, J. Vanhove, P. Berckmans, Q. Cai, K. Vanuytsel, K. Eggermont, V. Vanslembrouck, B. Z. Schmidt, S. Raitano, L. Van Den Bosch, Y. Nahmias, T. Cathomen, T. Struys, and C. M. Verfaillie, “Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition,” *Stem Cell Reports*, vol. 5, no. 5, pp. 918–931, 2015.
- [133] P. Zhu, J. Sawinski, A. Cetin, M. T. Hasan, U. Baron, P. Osten, R. Sprengel, J. Herb, M. I. Aller, S. Cambridge, M. Bausen, S. Kügler, M. L. Nelson, and

- P. H. Seeburg, "Silencing and Un-silencing of Tetracycline-Controlled Genes in Neurons," *PLoS ONE*, vol. 2, no. 6, p. e533, 2007.
- [134] M. Kim, P. M. O'Callaghan, K. A. Droms, and D. C. James, "A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies," *Biotechnology and Bioengineering*, vol. 108, no. 10, pp. 2434–2446, 2011.
- [135] C. C. Hsu, H. P. Li, Y. H. Hung, Y. W. Leu, W. H. Wu, F. S. Wang, K. D. Lee, P. J. Chang, C. S. Wu, Y. J. Lu, T. H. Huang, Y. S. Chang, and S. H. Hsiao, "Targeted methylation of CMV and E1A viral promoters," *Biochemical and Biophysical Research Communications*, vol. 402, no. 2, pp. 228–234, 2010.
- [136] Y. Yang, Mariati, J. Chusainow, and M. G. Yap, "DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines," *Journal of Biotechnology*, vol. 147, no. 3–4, pp. 180–185, 2010.
- [137] E. Papadakis, S. Nicklin, A. Baker, and S. White, "Promoters and Control Elements: Designing Expression Cassettes for Gene Therapy," *Current Gene Therapy*, vol. 4, no. 1, pp. 89–113, 2005.
- [138] M. Yamada, Y. Suzuki, S. C. Nagasaki, H. Okuno, and I. Imayoshi, "Light Control of the Tet Gene Expression System in Mammalian Cells," *Cell Reports*, vol. 25, no. 2, pp. 487–500, 2018.
- [139] C. M. Akitake, M. Macurak, M. E. Halpern, and M. G. Goll, "Transgenerational analysis of transcriptional silencing in zebrafish," *Developmental Biology*, vol. 352, no. 2, pp. 191–201, 2011.
- [140] H. Ayyub, R. J. S. Murray, D. M. Jeziorska, B. Graham, R. Gaentzsch, J. N. Lund, D. R. Higgs, T. Chen, M. De Gobbi, A. J. H. Smith, E. Li, J. Telenius, J. R. Hughes, M. Lynch, C. Tufarelli, and D. Garrick, "DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease," *Proceedings of the National Academy of Sciences*, vol. 114, no. 36, pp. E7526–E7535, 2017.
- [141] I. Comet, B. Leblanc, K. Helin, J. Johansen, E. Riising, and X. Wu, "Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide," *Molecular Cell*, vol. 55, no. 3, pp. 347–360, 2014.
- [142] V. Flury, P. R. Georgescu, V. Iesmantavicius, Y. Shimada, T. Kuzdere, S. Braun, and M. Bühler, "The Histone Acetyltransferase Mst2 Protects Active Chromatin from Epigenetic Silencing by Acetylating the Ubiquitin Ligase Brl1," *Molecular Cell*, vol. 67, no. 2, pp. 294–307, 2017.
- [143] Y. Yu, M. M. Lowy, and R. C. Elble, "Tet-On lentiviral transductants lose inducibility when silenced for extended intervals in mammary epithelial cells," *Metabolic Engineering Communications*, vol. 3, pp. 64–67, 2016.

- [144] X. Wang, Z. Xu, Z. Tian, X. Zhang, D. Xu, Q. Li, J. Zhang, and T. Wang, “The EF-1 α promoter maintains high-level transgene expression from episomal vectors in transfected CHO-K1 cells,” *Journal of Cellular and Molecular Medicine*, vol. 21, no. 11, pp. 3044–3054, 2017.
- [145] J. Y. Qin, L. Zhang, K. L. Clift, I. Hulur, A. P. Xiang, B. Z. Ren, and B. T. Lahn, “Systematic comparison of constitutive promoters and the doxycycline-inducible promoter,” *PLoS ONE*, vol. 5, no. 5, pp. 3–6, 2010.
- [146] F. Cella, L. Wroblewska, R. Weiss, and V. Siciliano, “Engineering protein-protein devices for multilayered regulation of mRNA translation using orthogonal proteases in mammalian cells,” *Nature Communications*, vol. 9, p. 4392, 12 2018.
- [147] X. J. Gao, L. S. Chong, M. S. Kim, and M. B. Elowitz, “Programmable protein circuits in living cells,” *Science (New York, N.Y.)*, vol. 361, pp. 1252–1258, 9 2018.
- [148] T. Fink, J. Lonžarić, A. Praznik, T. Plaper, E. Merljak, K. Leben, N. Jerala, T. Lebar, A. Strmšek, F. Lapenta, M. Benčina, and R. Jerala, “Design of fast proteolysis-based signaling and logic circuits in mammalian cells,” *Nature Chemical Biology*, vol. 15, no. 2, pp. 115–122, 2019.
- [149] E. K. Borchardt, L. A. Vadoros, M. Huang, P. E. Lackey, W. F. Marzluff, and A. Asokan, “Controlling mRNA stability and translation with the CRISPR endoribonuclease Csy4,” pp. 1921–1930, 2015.
- [150] E. K. Borchardt, R. M. Meganck, H. A. Vincent, C. B. Ball, S. B. V. Ramos, N. J. Moorman, W. F. Marzluff, and A. Asokan, “Inducing circular RNA formation using the CRISPR endoribonuclease Csy4,” pp. 619–627, 2017.
- [151] K. Endo, K. Hayashi, T. Inoue, and H. Saito, “synthetic translational switches,” *Nature Communications*, vol. 4, pp. 1–9, 2013.
- [152] K. P. Adamala, D. A. Martin-Alarcon, and E. S. Boyden, “Programmable RNA-binding protein composed of repeats of a single modular unit,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 19, pp. E2579–E2588, 2016.
- [153] R. Geissler, A. Simkin, D. Floss, R. Patel, E. A. Fogarty, J. Scheller, and A. Grimson, “A widespread sequence-specific mRNA decay pathway mediated by hnRNP A1 and A2 / B1,” no. 607, pp. 1–23, 2016.
- [154] V. Haberle and A. Stark, “Eukaryotic core promoters and the functional basis of transcription initiation,” *Nature Reviews Molecular Cell Biology*, vol. 19, pp. 621–637, 10 2018.
- [155] J. E. Wilusz, C. K. Jnbaptiste, L. Y. Lu, W. F. Marzluff, C.-d. Kuhn, L. Joshua-tor, and P. a. Sharp, “A triple helix stabilizes the 3' ends of long noncoding

- RNAs that lack poly (A) tails,” *Genes & Development*, vol. 26, no. 21, pp. 2392–2407, 2012.
- [156] J. A. Brown, D. Bulkley, J. Wang, M. L. Valenstein, T. A. Yario, T. A. Steitz, and J. A. Steitz, “Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix,” *Nature Structural and Molecular Biology*, vol. 21, no. 7, pp. 633–640, 2014.
- [157] T. A. Yario, J. A. Steitz, M. L. Valenstein, J. A. Brown, and K. T. Tycowski, “Formation of triple-helical structures by the 3’-end sequences of MALAT1 and MENA noncoding RNAs,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 47, pp. 19202–19207, 2012.
- [158] L. Nissim, S. D. Perli, A. Fridkin, P. Perez-Pinera, and T. K. Lu, “Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells,” *Molecular Cell*, vol. 54, no. 4, pp. 698–710, 2014.
- [159] R. T. H. Lee, A. S. M. Ng, and P. W. Ingham, “Ribozyme mediated gRNA Generation for in vitro and in vivo CRISPR/Cas9 mutagenesis,” *PLoS ONE*, vol. 11, no. 11, pp. 1–12, 2016.
- [160] M. L. Hochstrasser and J. A. Doudna, “Cutting it close: CRISPR-associated endoribonuclease structure and function,” *Trends in Biochemical Sciences*, vol. 40, no. 1, pp. 58–66, 2015.
- [161] O. Niewoehner, M. Jinek, and J. A. Doudna, “Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases,” *Nucleic Acids Research*, vol. 42, pp. 1341–1353, 1 2014.
- [162] D. G. Sashital, M. Jinek, and J. A. Doudna, “An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3,” *Nature Publishing Group*, vol. 18, no. 6, pp. 680–687, 2011.
- [163] M. M. Jore, M. Lundgren, E. van Duijn, J. B. Bultema, E. R. Westra, S. P. Waghmare, B. Wiedenheft, A. Pul, R. Wurm, R. Wagner, M. R. Beijer, A. Barendregt, K. Zhou, A. P. L. Snijders, M. J. Dickman, J. A. Doudna, E. J. Boekema, A. J. R. Heck, J. van der Oost, and S. J. J. Brouns, “Structural basis for CRISPR RNA-guided DNA recognition by Cascade,” *Nature Structural & Molecular Biology*, vol. 18, pp. 529–536, 5 2011.
- [164] J. Carte, R. Wang, H. Li, R. M. Terns, and M. P. Terns, “Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes,” *Genes & Development*, vol. 22, pp. 3489–3496, 12 2008.
- [165] S. Konermann, P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev, P. D. Hsu, T. V.-d. C. Effectors, S. Konermann, P. Lotfy, N. J. Brideau, J. Oki, M. N. Shokhirev, and P. D. Hsu, “Transcriptome Engineering with RNA-Targeting

- Article Transcriptome Engineering with RNA-Targeting,” *Cell*, vol. 173, no. 3, pp. 665–668, 2018.
- [166] R. E. Haurwitz, M. Jinek, B. Wiedenheft, K. Zhou, and J. A. Doudna, “Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease,” *Science*, vol. 329, pp. 1355–1358, 9 2010.
- [167] V. Kunin, R. Sorek, and P. Hugenholtz, “Evolutionary conservation of sequence and secondary structures in CRISPR repeats,” *Genome Biology*, vol. 8, p. R61, 4 2007.
- [168] X. W. Wang, L. F. Hu, J. Hao, L. Q. Liao, Y. T. Chiu, M. Shi, and Y. Wang, “A microRNA-inducible CRISPR-Cas9 platform serves as a microRNA sensor and cell-type-specific genome regulation tool,” *Nature Cell Biology*, vol. 21, no. April, 2019.
- [169] S. Ausländer, D. Ausländer, M. Müller, M. Wieland, and M. Fussenegger, “Programmable single-cell mammalian biocomputers,” *Nature*, pp. 5–10, 2012.
- [170] B. H. Weinberg, N. T. H. Pham, L. D. Caraballo, T. Lozanoski, A. Engel, S. Bhatia, and W. W. Wong, “Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells,” *Nature Biotechnology*, vol. 35, no. 5, pp. 453–462, 2017.
- [171] R. Gaber, T. Lebar, A. Majerle, B. Šter, A. Dobnikar, M. Benčina, and R. Jerala, “construction of logic circuits in mammalian cells,” vol. 10, no. january, 2014.
- [172] J. J. Collins, T. S. Gardner, and C. R. Cantor, “Construction of a genetic toggle switch in *Escherichia coli*,” *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.
- [173] J. E. Ferrell and S. H. Ha, “Ultrasensitivity part I : Michaelian responses and zero-order ultrasensitivity,” *Trends in Biochemical Sciences*, vol. 39, no. 10, pp. 496–503, 2014.
- [174] M. L. Hochstrasser, D. W. Taylor, J. E. Kornfeld, E. Nogales, and J. A. Doudna, “DNA Targeting by a Minimal CRISPR RNA-Guided Cascade,” *Molecular Cell*, vol. 63, no. 5, pp. 840–851, 2016.
- [175] G. R. Chen, H. Sive, and D. P. B. Correspondence, “A Seed Mismatch Enhances Argonaute2-Catalyzed Cleavage and Partially Rescues Severely Impaired Cleavage Found in Fish,” *Molecular Cell*, vol. 68, pp. 1095–1107, 2017.
- [176] W. Tirasophon, K. Lee, B. Callaghan, A. Welihinda, and R. J. Kaufman, “The endoribonuclease activity of mammalian IRE1 autoregulates its mRNA and is required for the unfolded protein response,” *Genes and Development*, vol. 14, no. 21, pp. 2725–2736, 2000.

- [177] V. Siciliano, B. DiAndreth, B. Monel, J. Beal, J. Huh, K. L. Clayton, L. Wroblewska, A. McKeon, B. D. Walker, and R. Weiss, “Engineering modular intracellular protein sensor-actuator devices,” *Nature Communications*, vol. 9, p. 1881, 12 2018.
- [178] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson, “Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells,” *Science*, vol. 333, pp. 1307–1311, 9 2011.
- [179] C. E. Thomas, A. Ehrhardt, and M. a. Kay, “Progress and problems with the use of viral vectors for gene therapy,” *Nature reviews. Genetics*, vol. 4, no. 5, pp. 346–358, 2003.
- [180] S. Rajendran, S. Collins, J. P. Van Pijkeren, D. O’Hanlon, G. C. O’Sullivan, and M. Tangney, “Targeting of breast metastases using a viral gene vector with tumour-selective transcription,” *Anticancer Research*, vol. 31, no. 5, pp. 1627–1635, 2011.
- [181] F. A. Ran, L. Cong, W. X. Yan, D. a. Scott, J. S. Gootenberg, A. J. Kriz, B. Zetsche, O. Shalem, X. Wu, K. S. Makarova, E. V. Koonin, P. a. Sharp, and F. Zhang, “In vivo genome editing using *Staphylococcus aureus* Cas9,” *Nature*, vol. 520, no. 7546, pp. 186–190, 2015.
- [182] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, and G. M. Church, “RNA-Guided Human Genome Engineering via Cas9,” *Science*, vol. 339, no. 6121, pp. 823–826, 2013.
- [183] L. Qian and E. Winfree, “Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades,” *Science*, vol. 332, pp. 1196–1201, 6 2011.
- [184] G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree, “Enzyme-Free Nucleic Acid Logic Circuits,” *Science*, vol. 314, pp. 1585–1588, 12 2006.
- [185] R. M. Dirks and N. a. Pierce, “Triggered amplification by hybridization chain reaction.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 43, pp. 15275–15278, 2004.
- [186] G. Seelig, B. Yurke, and E. Winfree, “Catalyzed relaxation of a metastable DNA fuel,” *Journal of the American Chemical Society*, vol. 128, no. 37, pp. 12211–12220, 2006.
- [187] D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree, “Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA,” *Science*, vol. 318, pp. 1121–1125, 11 2007.
- [188] J. Bath and A. J. Turberfield, “DNA nanomachines.,” *Nature nanotechnology*, vol. 2, no. 5, pp. 275–284, 2007.

- [189] S. M. Douglas, I. Bachelet, and G. M. Church, “A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads,” *Science*, vol. 335, pp. 831–834, 2012.
- [190] R. a. Muscat, J. Bath, and A. J. Turberfield, “A programmable molecular robot,” *Nano Letters*, vol. 11, no. 3, pp. 982–987, 2011.
- [191] T. Omabegho, R. Sha, and N. C. Seeman, “A Bipedal DNA Brownian Motor with Coordinated Legs,” *Science*, vol. 324, pp. 67–71, 4 2009.
- [192] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce, “Programming biomolecular self-assembly pathways,” *Nature*, vol. 451, pp. 318–322, 1 2008.
- [193] L. Qian, E. Winfree, and J. Bruck, “Neural network computation with DNA strand displacement cascades,” *Nature*, vol. 475, pp. 368–372, 7 2011.
- [194] G. Chatterjee, Y.-J. Chen, and G. Seelig, “Nucleic Acid Strand Displacement with Synthetic mRNA Inputs in Living Mammalian Cells,” *ACS Synthetic Biology*, vol. 7, pp. 2737–2741, 12 2018.
- [195] N. Srinivas, T. E. Ouldridge, P. Sulc, J. M. Schaeffer, B. Yurke, A. a. Louis, J. P. K. Doye, and E. Winfree, “On the biophysics and kinetics of toehold-mediated DNA strand displacement,” *Nucleic acids research*, vol. 41, pp. 10641–58, 12 2013.
- [196] P. Šulc, T. Ouldridge, F. Romano, J. Doye, and A. Louis, “Modelling Toehold-Mediated RNA Strand Displacement,” *Biophysical Journal*, vol. 108, no. 5, pp. 1238–1247, 2015.
- [197] D. Y. Zhang and G. Seelig, “Dynamic DNA nanotechnology using strand-displacement reactions,” *Nature chemistry*, vol. 3, no. 2, pp. 103–113, 2011.
- [198] M. S. Ebert, J. R. Neilson, and P. A. Sharp, “MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells,” *Nat Meth*, vol. 4, pp. 721–726, 9 2007.
- [199] D. Y. Zhang and E. Winfree, “Control of DNA strand displacement kinetics using toehold exchange,” *J Am Chem Soc*, vol. 131, no. 47, pp. 17303–17314, 2009.
- [200] A. a. Green, P. a. Silver, J. J. Collins, and P. Yin, “Resource Toehold Switches : De-Novo-Designed Regulators of Gene Expression,” *Cell*, vol. 159, no. 4, pp. 925–939, 2014.
- [201] B. Schwanhausser, “Global quantification of mammalian gene expression control,” *Nature*, vol. 473, pp. 337–342, 2011.

- [202] L. Yen, M. Magnier, R. Weissleder, B. R. Stockwell, and R. C. Mulligan, "Identification of inhibitors of ribozyme self-cleavage in mammalian cells via high-throughput screening of chemical libraries.," *RNA (New York, N.Y.)*, vol. 12, no. 5, pp. 797–806, 2006.
- [203] Y. Laising, S. Jennifer, J.-S. Lee, J. T. Gray, M. Magnier, T. Baba, R. J. D'Amato, and R. C. Mulligan, "Exogenous control of mammalian gene expression through modulation of RNA self-cleavage Laising," *Nature*, vol. 431, no. September, pp. 471–476, 2004.
- [204] R. Penchovsky, "Computational design of allosteric ribozymes as molecular biosensors.," *Biotechnology advances*, vol. 32, no. 5, pp. 1015–27, 2014.
- [205] P. Guye, Y. Li, L. Wroblewska, X. Duportet, and R. Weiss, "Rapid, modular and reliable construction of complex mammalian gene circuits," *Nucleic Acids Research*, vol. 41, pp. e156–e156, 9 2013.
- [206] W. Weber and M. Fussenegger, "Emerging biomedical applications of synthetic biology," *Nature Reviews Genetics*, vol. 13, no. 1, pp. 21–35, 2011.
- [207] "MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.."
- [208] J. Beal, T. E. Wagner, T. Kitada, O. Azizgolshani, J. M. Parker, D. Densmore, and R. Weiss, "Model-Driven Engineering of Gene Expression from RNA Replicons," *ACS Synthetic Biology*, vol. 4, pp. 48–56, 1 2015.
- [209] R. N. Cohen, M. A. van der Aa, N. Macaraeg, A. P. Lee, and F. C. Szoka, "Quantification of plasmid DNA copies in the nucleus after lipoplex and polyplex transfection," *Journal of Controlled Release*, vol. 135, pp. 166–174, 4 2009.
- [210] L. Bleris, Z. Xie, D. Glass, A. Adadey, E. Sontag, and Y. Benenson, "Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template," *Molecular Systems Biology*, vol. 7, p. 519, 2011.
- [211] J. H. Strauss and E. G. Strauss, "The Alphaviruses: Gene Expression, Replication, and Evolution," Tech. Rep. 3, 1994.
- [212] N. Mittal, N. Roy, M. M. Babu, and S. C. Janga, "Dissecting the expression dynamics of RNA-binding proteins in posttranscriptional regulatory networks.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 20300–5, 12 2009.
- [213] D. S. Peabody, "The RNA binding site of bacteriophage MS2 coat protein," Tech. Rep. 2, 1993.
- [214] T. Middleton and B. Sugden, "Retention of plasmid DNA in mammalian cells is enhanced by binding of the Epstein-Barr virus replication protein EBNA1.," *Journal of virology*, vol. 68, pp. 4067–71, 6 1994.

- [215] J. Beal, R. Weiss, D. Densmore, A. Adler, E. Appleton, J. Babb, S. Bhatia, N. Davidsohn, T. Haddock, J. Loyall, R. Schantz, V. Vasilev, and F. Yaman, “An end-to-End workflow for engineering of biological networks from high-level specifications,” *ACS Synthetic Biology*, vol. 1, no. 8, pp. 317–331, 2012.
- [216] W. P. Jencks, “On the attribution and additivity of binding energies,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, pp. 4046–50, 7 1981.
- [217] J. J. Lauffenburger, D. A. & Linderman, “Receptors : models for binding, trafficking, and signaling,” *Oxford University Press*, 1993.
- [218] R. B. Kapust, J. Tözsér, T. D. Copeland, and D. S. Waugh, “The P1 specificity of tobacco etch virus protease,” *Biochemical and Biophysical Research Communications*, vol. 294, pp. 949–955, 6 2002.
- [219] S. H. Northrup and H. P. Erickson, “Kinetics of protein-protein association explained by Brownian dynamics computer simulation,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, pp. 3338–42, 4 1992.
- [220] D. W. Colby, Y. Chu, J. P. Cassady, M. Duennwald, H. Zazulak, J. M. Webster, A. Messer, S. Lindquist, V. M. Ingram, and K. D. Wittrup, “Potent inhibition of huntingtin aggregation and cytotoxicity by a disulfide bond-free single-domain intracellular antibody,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, pp. 17616–21, 12 2004.