Design of gRNA Expression Systems and Characterization of CRISPR Transcriptional Networks in *Saccharomyces cerevisiae*.

William E. Voje Jr.

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

James Carothers, Chair

Mary Lidstrom

Jesse Zalatan

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Design of gRNA Expression Systems and Characterization of CRISPR Transcriptional Networks in *Saccharomyces cerevisiae*.

William E. Voje Jr.

Chair of the Supervisory Committee:
Assistant Professor James Carothers
Chemical Engineering

The incredibly complex, nuanced, and robust processes happening within cells allow them to adapt to their environment, proliferate, and communicate with their neighbors. These processes are mediated by a network of bio-molecules that regulate each others' activity and expression. The sophistication of these natural systems make them intractable to recapitulate, let alone fully understand. An engineerable platform for the creation of predictable in vivo regulatory networks would make it significantly easier to access the potential biology has for human health, chemical manufacturing, and biosensing. This thesis discusses advances in engineering interconnected regulatory gene networks in *Saccharomyces cerevisiae.* These networks employ a version of CRISPR/Cas9, isolated from *Streptococcus pyogenes*, that has been optimized for transcriptional repression. This work develops a modular genetic unit composed of an engineered gRNA responsive promoters and a computationally designed gRNA expression platform. This thesis first discusses the computational methods used for RNA design. Then, how the CRISPR/Cas9 platform is used to build large networks composed of upwards of seven gRNA interactions. Finally, this work details how the platform was used to quantitatively characterize CRISPR/Cas9 systems, gRNA expression platforms, and to develop a novel ligand sensitive gRNA expression platform. The advances and insights gained in this work have impacts in transcriptional network design, Pol II gRNA expression and multi-gRNA CRISPR applications.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many people have helped me get through my graduate school experience, both technically and emotionally. First, thank you James Carothers for trusting me to start your lab. The freedom you gave me to explore the field of synthetic biology and RNA design has made me a robust and resourceful researcher. Thank you Rodrigo Correa-Rojas for teaching me microbiology and for your patience with me as the more eager and more arrogant person I was. Thank you Jason Stevens, you have enhanced my graduate school experience more than I could ever articulate. It was a pleasure to work next to you for 5.5 years. Thank you David Sparkman-Jager for your constant, if sometimes annoying, curiosity and intelligence. I'm glad to have been able to learn as much from you as I have. Thank you Chuhern Hwang, your positive nature and passion are inspiring and I look forward to following you and your future travels. Thank you Cassandra Burke, your support and guidance got me through some of the hardest times. Thank you Jason Fontana, it was great to work with you the little I did. Thank you Ian Faulkner, I would have not survived TAing without your positive outlook. A huge thanks to both Miles Gander and Justin Vrana. I shudder to think what my gradate career would have looked like without having walked into a collaboration with you both. Thank you to my undergraduate assistants Misha Yatsevitch and Elizabeth Glenski, I wish you all the best! Thank you to Cami Cordray and Michelle Parks and the rest of the UW BIOFAB team for making my experimental research work. Thank you to all of the amazing chemical engineering and bioengineering peers I was fortunate enough to interact with. You have made my graduate years all the more formative. If you haven't made it out yet, just hold on… Or don't.

Finally, I would like to thank my committee members for their feedback and support.

DEDICATION

This work is dedicated to my mother and father.

Without you, I wouldn’t be here.

Without you, I wouldn’t be as curious as I am.

Without you, I wouldn’t be as capable as I am.

Chapter 1. Introduction

1.1 Motivation

Cells make discrete and population-uniform decisions through sensing environmental and chemical cues. This is mediated by a process of genetic 'computation' where information is communicated through complex genetic interaction networks[1]. These are highly connected networks made of small molecules, DNA, protein, and RNA on a scale that make them currently intractable to design *de novo*[2,3]. The ability to engineer genetic decision networks as robust and precise as natural systems, would enable major advances in tissue engineering[4], cell-based analytics[5], and metabolic engineering[6]. The field has made significant advances toward this goal by taking an electronics approach in which characterized, orthogonal, and modular components are composed with few interconnections to realize complex system behavior. In this vein many natural genetic parts, which activate or repress gene expression, have been identified, isolated, and characterized. Many classes of genetic parts have been characterized including native[7] and engineered transcription factors[8], DNA recombinases[9], trans-acting RNA regulators[10,11], and cis-acting RNA regulators[12–15] with many being assembled into larger systems[16–18]. Unlike electrical systems however, biological circuits are subject to the complex and potentially resource limited environment of their host and additionally few classes of genetic parts are "portable" and only function in specific species or even specific strains[19,20].

A well characterized, modular, and portable framework for constructing biological circuits would be a boon for the field. This thesis will cover the advances that have been made to build engineered logical systems with CRISPRi transcriptional networks. This work merges computational design of portable RNA parts for the application of building a robust and extensible transcriptional network framework based on CRISPRi. The progress that has been made has implications for structured computational RNA design and CRISPR applications.

1.2 RNA as a Functional Molecule

Within the classic biological paradigm RNA served solely as a carrier of information from DNA to functional protein. But, since the 1980s this convention has been consistently challenged with the field accepting RNA as a dynamic, complex and structured genetic element that is essential for mediating biological activities, capable of catalysis and is used for cellular control schemes[21]. Non-coding RNA is responsible for mediating mRNA splicing in eukaryotes(group I and II introns), the regulation of translation in prokaryotes (RBS) and eukaryotes (IRES), inhibiting and activating gene expression (sRNA), mediating chemical reactions (ribozymes), binding small molecules and proteins with high affinity (aptamers), efficiently linking sensing and actuating functionalities to dynamically regulate gene expression (riboswitches) and forming functional protein complexes (ribosome, Cas9)[21]. These functions are encoded in a given RNA by both its sequence and structure[22]. RNA forms complex tertiary structures due to non-covalent molecular interactions between its component nucleotides[23]. Consequently, the formation of specific structures are dictated by the molecule's thermodynamic landscape. This landscape, and an RNAs ability to traverse it, is influenced by the rate of its polymerization[24], the environmental conditions[25,26], and its interaction with other molecules. It is for these reasons that the influence RNA has on biological processes is actively studied in the field of systems and synthetic biology.

1.3 CRISPR-Cas

The discovery of the clustered regularly interspaced short palindromic repeats (CRISPR) adaptive bacterial immune system is arguably one of the most important biological discoveries in modern times[27]. CRISPR systems are generally composed of one or more proteins which function as an endonuclease that are targeted to specific DNA by one or more non-coding RNAs[28]. Natively this mechanism is used to destroy foreign DNA within bacterial hosts. However, the real power of this platform is its functionality and engineerability. For the first time researchers have had access to a endonuclease that can be specifically addressed to DNA

through re-coding the sequence of a non-coding RNA, instead of building proteins with DNA specificity. Of the two classes and four types of CRISPR systems the type II CRISPR-Cas9 derived from *Streptococcus pyogenes* (spCas9) has been the most studied. This system naively consists of a unique clustered repeat RNA (crRNA), containing a 20 bp target sequence, a common trans-activating crRNA (tracrRNA) and the spCas9 protein. After the tracrRNA and crRNA complex together they are bound by the spCas9. The target sequence of the crRNA specifically addresses the spCas9 to its target. The tracrRNA and crRNA have been successfully combined into a single guide RNA (gRNA)[29]. As a platform CRISPR-Cas9 has dramatically enhanced genome editing of mammalian cells[30], has served as a novel antibacterial platform[31] and has been used for high throughput phenotyping screens to identify functional genes[32,33]. spCas9 has additionally been mutated to remove its nuclease activity (dCas9) generating a protein scaffold that that can be specifically targeted to DNA. dCas9, when targeted near the transcription start site of a promoter reduces gene expression through a system of steric interference in bacteria[34] and mammalian cells[35] (CRISPRi). In addition to this functional domains have been fused to dCas9 giving it new function, including transcription factor domains gene activation[36] (CRISPRa), epigenetic modifying domains[37], fluorescent proteins to dynamically image genome loci[38] and endonuclease domains for higher fidelity genome modifications[39]. The applications of this technology is far reaching.

Of particular interest to this thesis is CRISPR-Cas9 being used a modular transcription factor (CRISPRi). Transcriptional networks can be made by using and expressed gRNA to target dCas9 to the promoter driving the transcription of another gRNA it. These CRISPRi repression networks have been built in bacteria[40], mammalian cells[41] and, in our hands, yeast cells[42]. These transcriptional networks are incredibly versatile in the fact that they are both regulated by and express gRNA. As a consequence it is possible to conceive of systems composed of hundreds of orthogonal connections all using the same genetic part, something unheard of only years ago.

1.4 Genetic circuits

An application of transcriptional networks is the creation of 'genetic circuits.' This is a framework which abstracts biological interactions so they may be used in more of an electrical engineering paradigm[43]. For example, the interaction between a transcriptional repressor and promoter could be viewed as a voltage attenuator, that decreases the "signal" (transcript) with increased expression of the repressor. Alternatively it could be viewed as single-input single-output Boolean NOT gate, where only if there is no repressor does "signal" (transcript) come from the promoter. In these kinds of frameworks the limits of expression, which could be transcription, translation, or gene-editing, are often considered to be "states." These ideas are compelling to bioengineering because if a robust biological framework is established, well established electrical engineering principles can be applied to create complex logical systems.

Many classes genetic parts exhibit characteristics that are applicable for creating genetic circuits have been established[44]. There are two general classes of parts, those that behave in a digital manner and those that behave and analog manner. These distinctions have to do with how much induction of a part is required to achieve a shift from low-to-high or high-to-low states. Digital systems achieve this transition as soon as they start, whereas analog systems have long transitions[45]. Digital systems are useful when attempting to create logical systems, which conditionally change states. Analog systems are useful for analog computation[18], or dynamically titrating a system output[46]. Significant progress has been in digital systems using recombinases to transpose DNA elements. These have been used to create a 16 input state machine in E. coli[9] and a modular system for building Boolean logic and arithmetic in mammalian cells[47]. Though these systems have incredible efficacy they often result in permanent gene edits making them unsuited to any dynamic systems that may need to switch back. Significant developments have been made in analog control systems through the design and optimization of promoters and transcription factors[44]. These works use either native or engineered transcription factors and generally require tuning the level of expression based on promoter engineering[48], characterizing the performance of many parts[16], and building sophisticated models of component behaviors[14,40,49]. Additionally

there have been efforts to combine analog and digital parts to create multi-stage analog systems and comparators[50,51]. Finally, the modular and easily reprogrammable CRISPRi platform has been used to construct genetic circuits in *E. Coli*[40], mammalian[41], and *S. cerevisiae*[42] by using gRNA to target the expression of other gRNA.

Despite the numerous tools and several impressive demonstrations of their utility genetic circuit design is still very challenging[52]. The creation of genetic circuits requires inputs and outputs of interacting biological parts to be perfectly matched[44,53]. Though there have been impressive efforts which characterize large libraries of parts, predict how they will interact with one another and then build them[16,48] this process is very time consuming and it is quite possible for the added components to react to one another, an effect referred to as retroactivity[54–56]. Methods have been developed to 'insulate' genetic parts from retroactive effects[49], but are on suitable to a small class of genetic circuits and require genetic parts with specific characteristics that are limited in number. Though many classes of genetic parts have been characterized and used there a few examples of combining their functionality[14,50,51]. This is because combining combining functionality requires very specific matching of component characteristics[14]. There are also issues of genetic stability, it is rarely in the interest of cells to express many exogenous genes and consequently they will often mediate the breaking of the engineered systems[57]. Finally, many genetic parts are not portable between organisms and in some cases between species. This means that excellent work done in a lab strain organism may forever be constrained to that specific organism context. On the whole, there is still much work left to do in the genetic design.

1.5 Organization of thesis

This thesis will cover the span of my research since joining the Carothers' lab in 2012. This is mostly in chronological order with the projects I have undertaken. All of these chapters are connected through the theme of engineering and characterizing biological parts and systems. When I started I was only concerned with building tools for computational RNA design and now

as I am leaving I have applied these tools and developed new tools to study transcriptional networks in yeast.

Chapter 2 provides and overview of computational RNA design and its role in advancing the engineerability of biological systems for systems and synthetic biology. This covers many of the biological components that can and have been designed, covers some of the computational methods that are available for RNA design, details the design conventions that have been developed for in-house RNA design and finally discusses the computational package I have developed during my tenure, pyrfold. This python package contains a set of objects for use in creating RNA parts, a set of APIs for popular RNA folding applications and a set of analysis tools for processing RNA folding information. This chapter closes with some general insights that have come from computational RNA folding designs.

Chapter 3 introduces a modular framework for creating genetic logic in yeast using CRISPRi. This work was in collaboration the Klavin's lab and was lead by Dr. Miles Gander. In this work an engineered gRNA responsive promoter containing two 20 bp target sites were combined with a ribozyme-gRNA-ribozyme (RGR) architecture to create a two-input one-output Boolean logic NOR gate. This modular NOR gate was used to build seven distinct other two-input one-output Boolean classifiers. Additionally these components were layered to create a seven layer repression cascade, the longest genetic cascade that has been constructed to date. Modeling and analysis of system dynamics revealed significant design features, including but not limited to, the time it takes to switch states and the systems sensitivity to input.

Chapter 4 uses the CRISPRi framework established in Chapter 3 as a platform to study the RGR Pol II gRNA expression platform and the CRISPRi transcriptional networks. Critical features of the RGR architectures were identified and eleven novel variations were created. A ligand inducible version of the RGR which substituted the 3' ribozyme with a ligand inducible aptazyme was introduced. This RGA architecture was able to achieve up to 10 fold reduction in the RGA mediated activity upon ligand induction. In quantitative studies of limiting dCas9 it was found that co-expressed gRNA can compete for the shared dCas9 resource resulting in significant perturbations to system behavior. Finally this work demonstrates that CRISPRi

transcriptional networks can be modeled and the performance of large systems can be numerically estimated. Finally Chapter 5 contextualizes this work in the field and presents ideas for how this work can be continued.

Chapter 2. Computational RNA design

2.1 Introduction

De novo design of biological components with programmed behavior is the ultimate demonstration of understanding biology[58]. Unfortunately, due to the immense design space and complicated nature of biology, humanity is still far away from commanding this kind of understanding. Though still in the distance, some biological molecules and mechanisms are emerging as engineerable platforms. One of these platforms is RNA. It is a structured molecule capable of diverse biological functions mediated by its sequence and structure. Through directed evolution[13,59] and rational design of functional RNA devices[15,60] great advances have been made in the fields of synthetic biology and metabolic engineering. Apart from one-off and specific algorithms for designing specific RNA parts, there is a lack of easily accessible computational tools for the design of RNA. This Chapter will discuss some of the progress that has been made in computational RNA design and a collection of in-house tools that have been developed to aide in RNA design.

2.1.1 RNA function is derived from its sequence and structure

RNA structure formation is a product of its sequence and the environment in which it is folded. RNA is composed of four ribonucleic acid bases A, U, C, and G. These bases fold onto themselves to form complex and functional three-dimensional structures. The ultimate structure formed is mediated by the thermodynamics of non-covalent chemical interactions between these bases. As the dominate force holding RNA molecules together is hydrogen bonding these structures are highly influenced by salt content, pH and temperature of the environment. The structure of RNA forms hierarchically. First, on the order of microseconds, linear RNA will form secondary structure. Then, on the order of milliseconds, the RNA will form tertiary structure

(Figure 2.1)[23]. Consequently and conveniently, the structure of RNA is dominantly dictated by its secondary structure and for the remainder of this Chapter we will exclusively consider secondary structure. As RNA folds onto itself it will form various secondary structure motifs, such as helix stems, loops, and junctions. These structural elements when considered together for a topology[61] and these topologies when combined with conserved sequence result in specific activity. After synthesis, RNA will attempt to fold into its most stable structure. Unfortunately, the finite lifespan of RNA (~10 minutes[62]) limit and RNAs ability to fold into this structure.

As an RNA molecule attempts to fold into its most thermodynamically favorable structure it will likely make "mistakes" causing it misfold and potentially get kinetically trapped in sub-optimal structures (Figure 2.1). The time required unfold and refold towards the correct structure is often not accessible given the short RNA lifespan. Even more confounding is that RNA is transcribed a single base at time DNA template. Moreover, the rate of polymerization can vary from 10 nt/s in mammalian systems to 300 nt/s in *in vitro* systems. Consequently, RNA sequence folds outside of its full context, an effect that potentially biases non-optimal RNA structure formation. Importantly, these seemingly confounding aspects of RNA folding have been overcome through natural evolution of RNA sequences in context. Theoretical and experimental work has been done which shows that the efficiency of the HDV ribozyme folding into its active structure is significantly influenced by rate of polymerase[63]. Additionally, new experimental RNA structure studies have shown that co-transcriptional processing enables aptamer binding to a ligand by allowing it to fold first opposed to it folding simultaneously[24] and have shown that RNA folding decisions are encoded by co-transnational processing[64].

Of the many classes of RNA parts, its very likely that their activity is in someway determined by its sequence the structures it adopts. We consider an RNA part to be a sequence of RNA that can be characterized and that carries out a specific biochemical task. Consider the self-cleaving hammerhead ribozyme that mediates phosphodiester isomerization of itself resulting in practical terms to RNA cleavage. For many years after the hammerhead ribozyme was first discovered in 1982[65] research was mostly concerned with the "minimal sequence" that was required to achieve catalysis. However, in 2003 researchers figured out that additional RNA

sequence that encoded higher order structure greatly enhanced its rates of cleavage[66]. Only after considering the structure of the hammerhead ribozyme could practical applications be developed. RNA structure interactions are essential to RNA parts that function through trans RNA-RNA interactions, like sRNA and siRNA. These non-coding RNA are targeted to bind to RNA molecules by being having long stretches of highly complementary sequences. However, this process can be confounded by the *cis* structure each RNA has. This process will only happen of the net change of free energy between both RNA molecules unfolding and then complexing to eachother is negative. The role of RNA structure is so important that is is now common to search genomes looking for sequence which a likely to form a specific structural topologies which resemble known classes of RNA parts[67].

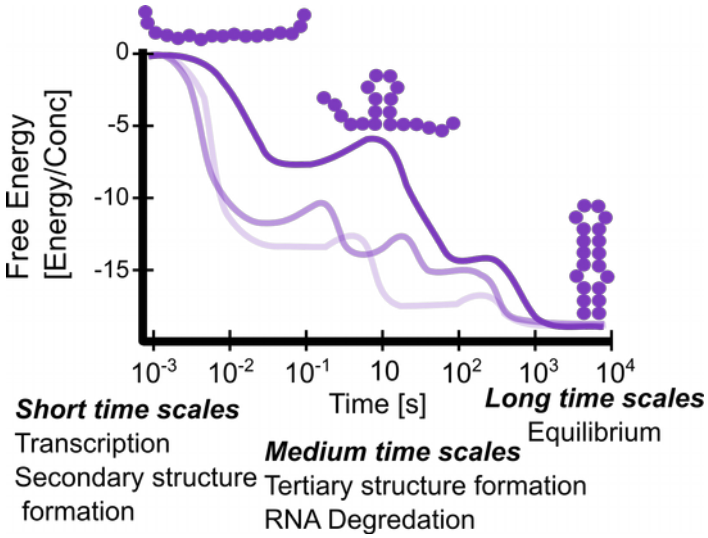


Figure 2.1: The structured and probabilistic life of RNA

An RNA molecule (purple balls) folds towards its lowest energy state during the course of its life. The purple lines represent the stability of its structure as a function of time. Each line represents a possible structure path that the same RNA sequence might take.

2.1.2 Computational methods for RNA structure prediction

The secondary structure of RNA can be predicted with three classes of tools: comparative, non-comparative and folding-path. Comparative algorithms predict the structure of a family of RNA sequences. These tools take a set of aligned RNA sequences and discern the RNA regions that commonly interact between all members. This analysis is useful when trying to identify the conserved structure that is associated within a family of RNA part[68]. Additionally these tools have been able to identify transient structure that RNA fold into during co-transcription that are conserved across species[69]. Though these methods are useful for identify the essential structural elements they cannot be used to predict the structure a given sequence of RNA would fold into give a sequence alone.

Non-comparative methods allow for the *de novo* estimation of RNA structure. These tools employ quantitative thermodynamic data to predict structures a given RNA sequence would fold into. They are commonly used to predict the most stable structure, referred to as the minimum free energy (MFE) structure. Though this metric has been historically useful, it is unlikely that an RNA molecule forms this structure, especially if it is long (>100 bp). This is due to the probabilistic nature of the RNA searching its accessible conformational space that is dictated its thermodynamics and the strong likelihood that the RNA will assume structures that it becomes kinetically trapped in. That being said more tools are being created which consider the partition function of all possible structures, the centroid structure, and structures that are potentially trapped in their non-ideal structure state[70–73]. Though these are effect the fail to consider any dynamics of structure formation.

To attempt to capture the kinetics of RNA folding a class of thermodynamic tools known as folding-path have been developed. These tools use thermodynamic rules to estimate the rates of kinetic transitions that a given RNA molecule can undertake. In general these tools account for co-transcriptional folding pathways and the possibilities that local structural elements will be ‘kinetically’ trapped in non MFE structures. The use of these have provided significant insight into the consequences of polymerase speed for RNA part functions[69,74] and have been used to

inform the creation of RNA parts[15]. Unfortunately these tools are mostly probabilistic in nature, requiring many simulations to be run per sequence to estimate the full scope of the folding landscape. This is compounded by the fact that these simulations are more computationally intensive to run.

2.1.3 Examples of computational RNA design

Computational RNA design is primarily concerned with cis or trans functionality, here we will discuss cis functionality. Cis RNA designs are concerned with incorporating all necessary elements of behavior into a single RNA part. In this vein, research groups were able to take characterized aptamers, an RNA part that binds a ligand with high affinity, and combine them with the known architecture of a terminator to create novel riboswitches[75]. Similarly, regions of a mammalian IRES that are critical for functional folding were made to only fold in the presence of a small molecule by incorporating a ligand sensing aptamer into the IRES sequence[76]. The ribosome binding site (RBS) has been designed using thermodynamic folding simulations to optimize the local structure that is formed is minimized and to facilitate interactions between Shine-Dalgarno sequence and the ribosome[60]. Additionally, thermodynamic simulations have been used to program interactions between the target sequence of gRNA and a cis RNA which is either severed by an endoribonuclease[77] or a cis encoded aptazyme[78]. Finally, cis RNA design has been applied to part insulation where sequences upstream and downstream of a characterized part are used to bias it to fold into its correct structure[15,42,79].

2.1.4 Challenges that still exist

Though RNA has been studied extensively and we possess the tools to predict the secondary structure that RNA forms, there are still many challenges. First, the RNA design space is massive, scaling with 4^n. This means that even with perfectly efficient computational folding simulations, searching all of design space is still prohibitively time consuming. Additionally the RNA folding landscape is rugged. Meaning that there are many local optimum and local minimum. There have been some impressive tools created that take a set of structural and

sequence constraints as inputs and then generate RNA sequences which satisfy these constraints[80], but it is still challenging to incorporate kinetic constrains into these systems.

2.2 Pyrfold a python package for RNA design

To address some of these issues I have developed Pyrfold, a python package for the design and characterization of RNA parts. This package has three primary functions: designing RNA parts with structural constrains for the purpose of design, simulating how those given sequence elements will fold, and analyzing the resulting data. Here I will detail some of the functionality.

2.2.1 RNA part design

As previously stated structured RNA parts consist of of structured and unstructured elements. Towards the goal of designing structured RNA parts I created a set of python objects that allow for the specification of single stranded RNA. In general, these RNA objects consider RNA at the level of sequence or at the level of a device. A device is multiple sequences stitched together to either form a functional part or to put a part in context.

2.2.1.1 design.RNA.Unpaired()

This the most basic RNA object. It allows for the specification of unpaired sequence. For the purpose of this and all subsequent tools unpaired means that it has no dependence on any other stretch of sequence. This object allows for the specification of the its length range and the range of GC content that are allowed. After specification of the object its sequence can be randomized using the .randomize() call. This will generate a new sequence based on the initialized constraints.

2.2.1.2 design.RNA.Helix()

This object allows for the specification of structured sequences. This object contains two stretches of linear RNA that have sequence that are dependent on each-other. The 5' helix is refereed to as 'helix0' and the 3' helix is refereed to as 'helix1.' Again the length of this helix can be specified and its GC content can be specified. Once specified this object can be used to generate either helix0 or helix1 from the sequence identity of the other using .generate_helix(). This function call has an option that allows for the random substitution of U bases for C bases, allowing the indirect design of G:U wobble base-pairing. Finally this object has a .randomize() function which will generate a random helix given the constraints it was initialized with.

2.2.1.3 design.RNA.Device()

Finally the above RNA objects are used within an RNA Device object. This object is used to define many RNA sequence elements needed to create a functional RNA part or an RNA part in sequence context. This object is first specified with a list of RNA sequences and a list of names which correspond to those sequences. After initializing the object, it is possible to define any of the RNA sequences to be unpaired or in a helix confirmation. When a specific RNA sequence is defined as such, the previously discussed constraints can be applied. Now the RNA.Device object has a global .randomize_parts() call where every sequence element that can be randomized is randomized. It is also possible to only randomize a single RNA sequence and to change the sequence of a part. Any section of RNA sequence within the device can be queried using .part_sequence(), which will return the sequence of the specified part or .windowed_sequence() which will return all sequence in the upstream and downstream window that is given. Finally these Device objects can be used to generate kinefold simulation specifications which will be discussed in more detail later.

2.2.2 RNA simulation

This package has a set of function calls which are used to fold RNA sequences. These include a suite of non-comparative simulations which use viennaRNA functions[72] and an extensive set of functions for deploying the folding-path simulation tool kinefold[81] on a computational cluster.

2.2.2.1 ViennaRNA wrapper

ViennaRNA is a fully featured RNA C based simulation tool. For this my work I have created a basic wrapper to access the RNAfold subroutine. This tool can be used to generate the minimum free energy structures (fold.vienna.mfe_plus()) or the maximum expected accuracy functions (fold.vienna.mea_plus()). Both functions return information on the entire partition of structures that can be formed.

2.2.2.2 Kinefold tools

Co-transcriptional folding simulations are very useful for RNA design. In my work I have dominantly used the tool Kinefold. The timescale of Kinefold simulations are on the order of 10s to 100s of seconds. Consequently, simulations must be run on a super computing cluster if they are to be used for design. We developed a suite of python code to parallelize kinefold folding simulations on a Terascale Open-source Resource and QUe manage (TORQUE) system. This consists of a collection of tools which convert raw kinefold output to robust python objects which map simulation time to standard dotbracket representations of structure. Though this package has only been tested on Hyak it should be possible to port this functionality to other compute clusters[79]. The basic framework for Kinefold is presented in Figure 2.2. A human readable .csv file of all requested simulations is used to create a folder hierarchy that is then processed on a cluster. A collection of processing scripts were created to convert the output of kinefold into a python accessible data structure. This involves aggregating many simulations together in such a way that they can be used to generate folding trace data.

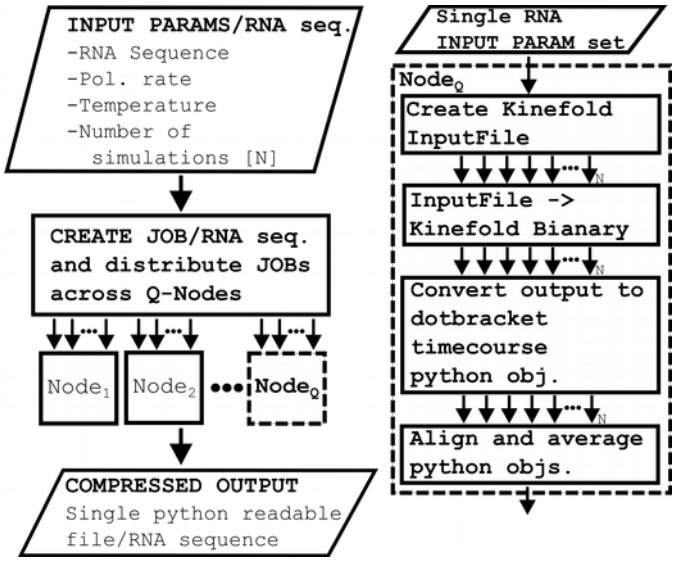


Figure 2.2: Computational structure pyrfold uses for kinefold simulations

Pyrfold contains a TimeCouresStructure() object that is used generate folding-trace data from kinefold simulations. Folding-trace data describes the structure a given sequence of RNA forms over time. As it is very difficult to visualize the many possible structures that are formed by an entire sequence of RNA over a period of time, a method of windowing a specific subset of sequence was developed. An application of this method would be to find all of the structures a given RNA part assumes during the course of the simulation. Consider the example folding trace data in Figure 2.3A, where during the course of the simulation there are two structures which represent 100% of the population. The metrics of maximum observed fraction folded, the average fraction folded, and the final fraction folded give rough estimates of these kind of dynamics (Figure 2.3B). Importantly when dealing with folding trace data many simulations are required to fully appreciate the folding dynamics (Figure 2.3C). These analysis methods make it easy to search for specific structures within sequences.

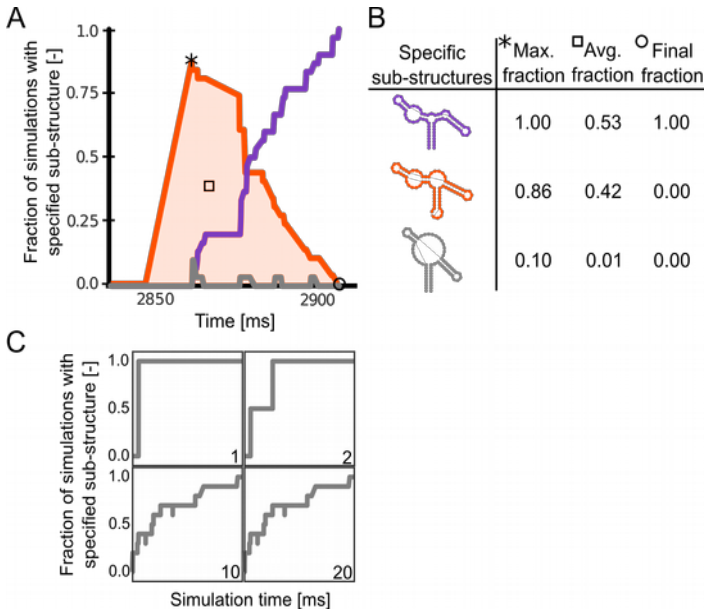


Figure 2.3: Overview of Timecourse analysis tools

(A) Folding traces for three substructures of interest (color indicated in panel B). (B) Summary metrics which describe the folding traces of panel A. (C) A look at the folding trace for a single sequence evolves as the number of averaged simulations (lower right corner) are increased.

2.3 Conclusions

Computational design of RNA can be used to great effect when designing new devices or moving a characterized RNA part from one context to another. The computational process to do this is not clearly defined. Consequently it is challenging to accurately specifying and modifying sequences, simulating the folds of the RNA, and processing the data. Folding path simulations are especially inaccessible. Here, the folding package pyrfold was introduced. Multiple objects have been created which aide in the design of structured RNA. These tools facilitate rapid and simple creation of candidate RNA sequences with user specified sequence variation. Additionally tools for simulating and subsequently analyzing the data are also introduced. On the whole, these contributions should help to accelerate someones ability to go from concept to computational RNA design. All code can be found in APPENDIX A.

Chapter 3. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates

3.1 Attribution

The following chapter has been adapted from:

> Gander, M.W., Vrana, J.D., Voje, W.E., Carothers, J.M., and Klavins, E. (2017). Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. Nat. Commun. 8, 15459.

My primary contributions to this work was computational RNA design for the creation of the ten RGR and ten iRGR devices. I performed sensitivity analysis on the impact parameter changes had on model performance, both for the predicted dynamic range of the system and the time to half max of the system.

3.2 Abstract

Natural genetic circuits enable cells to make sophisticated digital decisions. Building equally complex synthetic circuits in eukaryotes remains difficult, however, because commonly used components leak transcriptionally, do not arbitrarily interconnect or do not have digital responses. Here, we designed dCas9-Mxi1-based NOR gates in Saccharomyces cerevisiae that allow arbitrary connectivity and large genetic circuits. Because we used the chromatin remodeller Mxi1, our gates showed minimal leak and digital responses. We built a combinatorial library of NOR gates that directly convert guide RNA (gRNA) inputs into gRNA outputs,

enabling the gates to be 'wired' together. We constructed logic circuits with up to seven gRNAs, including repression cascades with up to seven layers. Modelling predicted the NOR gates have effectively zero transcriptional leak explaining the limited signal degradation in the circuits. Our approach enabled the largest, eukaryotic gene circuits to date and will form the basis for large, synthetic, cellular decision-making systems.

3.3 Introduction

Living cells make decisions based on information processing genetic programmes. Many of these programmes execute digital functions[82–89]. The capability to build synthetic digital systems in living cells could allow engineers to build novel decision-making regulatory networks for use in a variety of applications[90], ranging from gene therapies that modify cell state based on sensed information[91,92] to entirely new developmental programmes for tissue engineering[4,93]. In electronics a compositional approach has allowed the construction of digital circuits of great complexity to be quickly designed and implemented. Here, we have developed set of low-variability genetic parts that can be routinely composed to create large digital circuits in yeast cells.

Genetic components that implement simple logical operations, which in principle could be interconnected to form complex logic functions, have been demonstrated[16,40,41,94–100]. DNA-binding domains (DBDs) such as zinc fingers and TALEs (transcription activator- like effectors) have been used to construct libraries of transcription factors in eukaryotes [8,98,101–103]. However, scaling with DBDs in eukaryotes has been difficult because of challenges in synthesizing libraries of orthogonal parts[104,105]. Libraries of DBD-based parts have been shown in prokaryotes, but extensive part characterization and computer-aided design (CAD) was necessary to identify part combinations that yielded functional logic circuits[16]. Recently, programmable and orthogonal CRISPR-dCas9 transcription factors have been employed[34,35,40,41,106–108] to build up to five component circuits using dCas9-mediated repression in prokaryotes[40]. Transcriptional repression in these circuits is likely due to steric hindrance of RNA polymerase by dCas9.

Although dCas9 allows for programmable interconnections, its response function is leaky leading to signal degradation when layered[40]. Site-specific recombinases have been employed in genetic circuits as a means to reduce leak[51,109,110], but there are a limited number of such enzymes restricting the scalability of this approach. Here, we address these issues, advancing the art of engineering living digital circuits by focusing on two main engineering goals. First, we built a universal, single-gene NOR logic gate; the NOR gates are functionally complete[111] and as such can be composed to implement any logic function. Crucially, the input and output signals of our gates have the same molecular types while still being programmable so that, as in electronics, gates can be wired together. To achieve this, we made use of the CRISPR-dCas9 system: the signals in our framework are guide RNAs (gRNAs) whose sequences specifically match up to programmable target sequences on our NOR gate promoters.

Second, we required a consistent 'OFF' state for our NOR gates. To achieve this, we used the chromatin remodelling repression domain Mxi1 to take advantage of the eukaryotic cell's ability to repress gene expression, by fusing this domain to dCas9[35]. The Mxi1 domain is thought to recruit histone deacetylases[112,113], and with it we observed strong transcriptional repression in our circuits. The strong and consistent 'OFF' behaviour we observe with our NOR gates is a key factor that allows them to be composed into larger circuits by minimizing accumulation of transcriptional leak with every added layer. A mathematical model of our NOR gates predicts that they have effectively no transcriptional leak in their OFF states. We show that with low leak there exist parameters that allow our NOR gates to be composed without significant signal degradation. More importantly, we show experimentally that we can build a variety of digital logic circuits composed of up to five NOR gates and seven internal gRNA wires, as well as cascades of gates with up to seven layers that still have digital responses according to our specifications.

In summary, we developed low-variability single-gene NOR gates that can be regularly interconnected into arbitrary topologies that implement large digital circuits in yeast cells. Neither meticulous characterization of individual parts nor sophisticated design tools were necessary to find combinations of NOR gates that conferred functional circuits. Because the

technology is essentially generic and easy to rewire, it can in principle be used to implement arbitrary internal logic for a variety of synthetic cellular decision-making systems, such as those being explored for diagnostics[92,114], therapeutics[115], and development[115].

3.4 Results

3.4.1 NOR gate architecture

We built a universal, single-gene logic gate, in our case a NOR gate (Figure 3.1a). The NOR gate outputs are then gRNAs that match the target sequences on other NOR gate promoters (Figure 3.1b). Our NOR gates are genomically integrated into yeast cells (Figure 3.1c). We avoided using RNA polymerase (Pol) III promoters to express gRNAs[116–119] because they have low expression levels relative to Pol II promoters and are more difficult to engineer[120,121]. By programming the NOR gate input target sequences and output gRNA sequences in a set of gates, we were able to construct a variety of circuit topologies (Figure 3.1d).

Second, we required a consistent 'OFF' state for our NOR gates that corresponded to complete or near complete repression of the output promoter (Figure B.1). we used the chromatin remodelling repression domain Mxi1 to take advantage of the eukaryotic cell's ability to repress gene expression, by fusing this domain to dCas9[35] (Figure 3.2a). When compared with a number of repression domains, Mxi1 showed the strongest repression (Figure B.2). Our results suggest that such repression provides a significantly improved and more consistent 'OFF' signal compared with repression via steric hindrance (Figure 3.2b), in which dCas9 is interfering with transcriptional initiation, but is not remodelling chromatin. A mathematical model of our NOR gates, fit to both steady-state and time response data, predicts them have effectively zero transcriptional leak in their OFF states. Additionally, the model predicts that repression via steric hindrance leaks more than repression via dCas9-Mxi1 (Figure 3.2b).

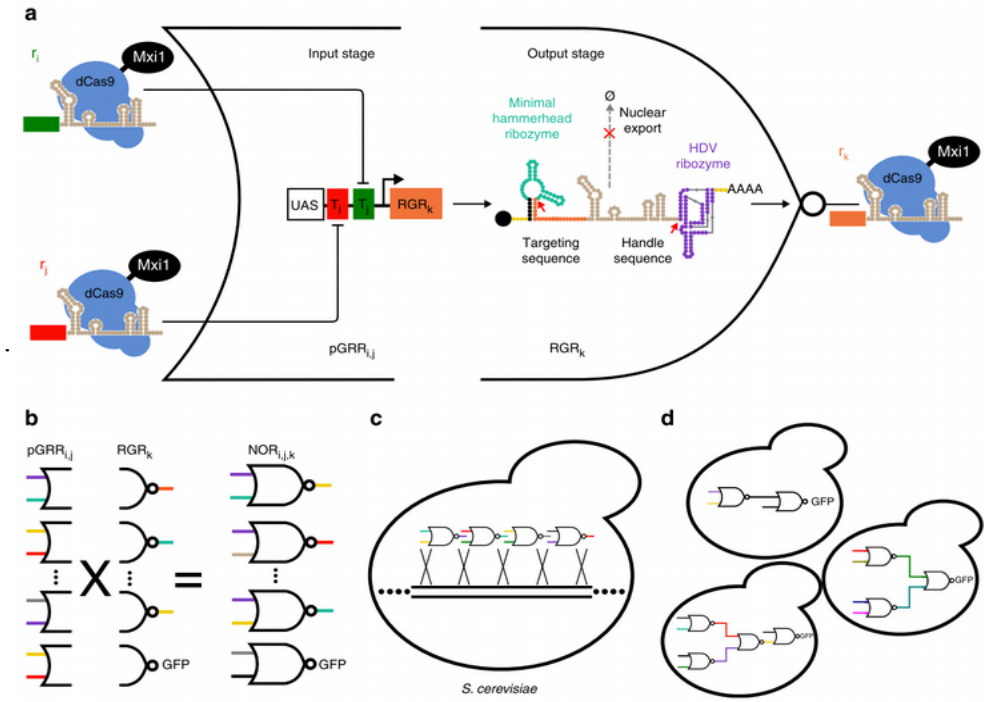


Figure 3.1: Schematic of the NOR gate architecture and circuit composition.

(a) A NOR gate input stage consists of a Pol II pGRR promoter that is fully repressed by the binding of either one or both of its cognate gRNA-dCas9-Mxi1 complexes. The output stage of the NOR gate is a gRNA transcript, flanked by self-cleaving ribozymes (RGR). Cleavage sites are indicated by red arrows. The cleavage of the ribozymes prevents nuclear export of the gRNA, indicated by dotted grey arrow. (b) The process of NOR gate library construction. Our library consists of a set of 400 two-input pGRR promoters and 20 RGR outputs for a total of 8,000 possible NOR gates. (c) Genomically integrating NOR gates into *S. cerevisiae.* (d) Arbitrary circuits are constructed by integrating multiple NOR gates into a single strain.

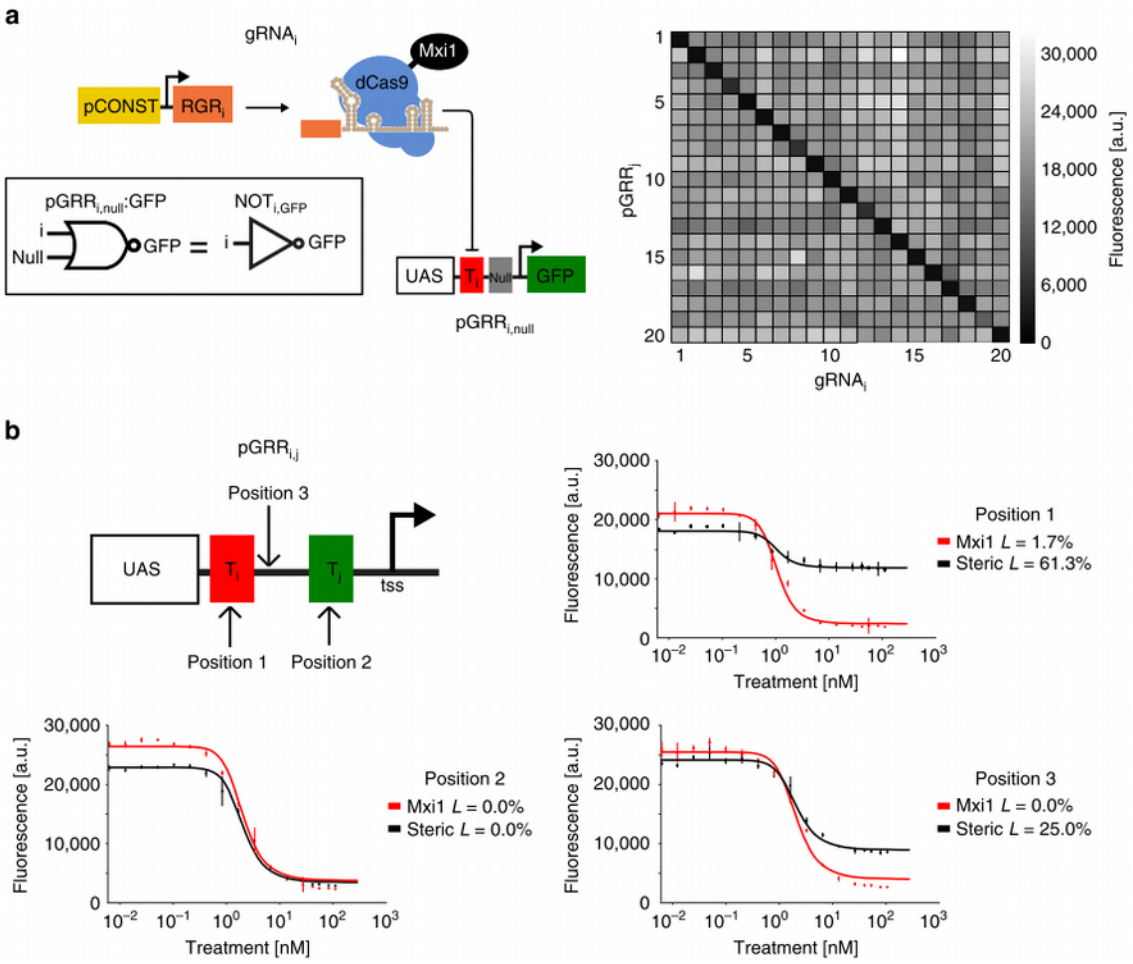


Figure 3.2: Orthogonality and repression via dCas9-Mxi1

(**a**) A constitutive promoter drives expression of gRNAs paired with a combinatorial library of cognate promoters. Orthogonality of the gRNA guide sequences was tested by crossing the 20 $pGRR_{i,null}$ promoters, each expressing GFP, with the 20 $gRNA_i$, creating 400 different strains of yeast. Fluorescence values of each strain were measured using flow cytometry. Fluorescence values from one biological replicate are displayed in the matrix. (**b**) Dose response curves are shown for repression via dCas9-Mxi1 and dCas9 repression via steric hindrance of pGRR driving GFP at three separate positions in the promoter. The three positions are annotated on the pGRR promoter representation. At all three positions, at maximal induction, dCas9-Mxi1 represses the promoter to a lower fluorescence level than dCas9 alone. Model fits predicted the parameter value *L*, representing transcriptional leak, for all curves. At all three positions the predicted L value is as small or smaller for dCas9-Mix1 than for steric repression. Error bars represent the s.d. of three biological replicates measured over three separate experiments.

Our approach allowed for the construction of the largest eukaryotic gene circuits, to the best of our knowledge, ever demonstrated (Table 3.1).

Table 3.1: Synthetic circuit size comparison

The best method for quantifying the size of synthetic biological circuits is an open question. Here we took the largest synthetic circuits constructed in recent publications and compared them with the two largest circuits from this paper. We separated the inputs to the circuits from internal components. We also counted the number of connections between the internal components. By our definition, a 'part' is a molecular species that carries information necessary for the internal function of the circuit (as opposed to a helper protein such as Cas9). A 'connection' is a molecular interaction between parts that propagates information within the circuit.

Publication	No. of gates /parts	No. of connections	No. of inputs	Circuit complexity $(\text{gates}^2 + \text{connections}^2)^{1/2}$	Functionally complete parts?	Medium
Cascade circuit	7	6	1	9.22	Yes	*S. cerevisiae*
Nielsen *et al.*[22]	7	6	3	9.22	Yes	*E. coli*
Qian *et al.*[70]	6	5	4	7.81	Yes	*In vitro*
XOR circuit	5	4	2	6.40	Yes	*S. cerevisiae*
Xie *et al.*[11]	5	4	6	6.40	No	Mammalian
Auslander *et al.*[71]	5	4	2	6.40	No	Mammalian
Regot *et al.*[72]	5	3	2	5.83	Yes	Multicellular *S. cerevisiae*
Nissim *et al.*[33]	5	3	1	5.83	No	Mammalian
Stanton *et al.*[19]	4	3	2	5	Yes	*E. coli*
Nielsen *et al.*[18]	3	2	2	3.61	Yes	*E. coli*
Kiani *et al.*[20]	2	2	1	2.83	No	Mammalian

The gate $NOR_{i,j,k}$, with input signals r_i and r_j and output r_k, consists of a gRNA-responsive Pol II promoter ($pGRR_{i,j}$) input stage, driving an output stage, ribozyme-flanked gRNA (RGR_k) (Figure 3.1a). According to NOR logic, r_k is high only when both r_i and r_j are low. A signal, r_i, is defined as a gRNA complexed with a dCas9-Mix1 fusion protein that confers strong transcriptional repression when bound to DNA[35]. The gRNA signals are distinguished by their unique 5' guide sequence. A 20-component library of signals defining r_1–r_{20} was used in this work (Figure B.1). The $pGRR_{i,j}$ promoter contains two, 20 base-pair (bp) target sites that match r_i and r_j respectively. Since we designed 20 signals, there are $20^3 = 8{,}000$ total NOR gates in the set. A $NOR_{i,j,k}$ functions as a $NOT_{j,k}$ if the $pGRR_{i,j}$ contains two identical target sites, if the $pGRR_{i,j}$ contains only one target site from the 20 component library ($pGRR_{i,null}$) or if r_i is simply not used in the circuit. A target sequence of 'null' refers to a pGRR that contains a target sequence that does not match any gRNA used in the containing circuit.

3.4.2 Input stage promoter design

The $pGRR_{i,j}$ promoter is tightly repressed when gRNA-dCas9-Mxi1 is bound to one or both of its two 20 bp target sites. The core region of the $pGRR_{i,j}$, the minimal pCYC1 promoter, was chosen based on its successful use with dCas9 in the past[107]. Because the promoter has relatively low expression levels and we wanted its output to have a strong ON output when not repressed, an upstream activating sequence (UAS) from the strong pGPD promoter[122] was added, forming the base pGRR promoter. The UAS increased the unrepressed expression level of the pGRR output by approximately threefold while maintaining the same OFF state expression level in the presence of r_i and r_j, further separating the digital ON and digital OFF levels (Figure B.3a). A pGRR promoter map highlighting all relevant sequence features is included in Figure B.4. A library of 11 $pGRR_{i,j}$ promoters, with i and j chosen from the 20 guide sequences, showed limited expression variability when driving GFP, with an 18% s.d. from the mean (Figure B.3b) Of the 20 $pGRR_{i,null}$:GFP constructs (i ranging from 1 to 20), 16 were repressed to or near the level of *Saccharomyces cerevisiae* autofluorescence in the presence of the corresponding signal r_i (Figure B.1).

3.4.3 Output stage RNA design

Two different RNA pol II expression methods were used in this work (Figure B.5). The first was an RGR design utilizing a 5' minimal hammerhead ribozyme (mHH) and a 3' hepatitis delta virus ribozyme (HDV), flanking the gRNA[123]. The second was an 'insulated' RGR (iRGR) with the mHH replaced by an avocado sunblotch viroid (ASBV) ribozyme. Both designs are intended to post-transcriptionally remove nuclear export signals, the 5' cap and 3' poly-A tail[124,125]. It has been shown that RNA device folding can be insulated from surrounding sequence context through computational sequence selection[15,126]. Ten guide sequences were chosen for the RGR architecture that were computationally predicted to confer proper folding of the mHH 5' ribozyme. Ten more guide sequences were chosen for the iRGR context whose ASBV 5' ribozyme is predicted to fold properly regardless of guide sequence. We observed similar levels of dCas9-Mxi1-mediated repression with gRNAs expressed from both iRGR and RGR

constructs (Figure B.6). Interestingly, RGR transcripts lacking a 5' ribozyme also showed dCas9-Mxi1-mediated repression. These results are consistent with previous studies that indicate a majority of 5' extended gRNA target sequences are processed to 20 nucleotides[127]. No significant crosstalk was observed when all $r_{1–10}$ (RGR design) and $r_{11–20}$ (iRGR design) were paired with all $pGRR_{1\text{-}20,null}$:GFP among noncognate pairs (Figure 3.2a and Figure B.7). Out of 20 total RGRs ($RGR_{1–10}$ and $iRGR_{11–20}$) when targeted to their cognate $pGRR_{1\text{-}20,null}$:GFP constructs, 16 repressed fluorescence to or near the level of autofluorescence for *S. cerevisiae* (Figure B.1).

3.4.4 Logic circuits

As a demonstration of the complex circuits possible with our NOR gates, six two-input, one-output digital logic circuits were built by integrating up to five NOR gate cassettes into various selectable loci in the yeast genome (Figure 3.3a–f). The output of each circuit was made observable by having the last NOR gate drive the expression of GFP. The circuits were constructed from the 16 guide sequences of the 20-component library that exhibited the strongest repression (Figure B.1). The truth table for each gate was experimentally obtained by constructing four separate strains, one for each pair of possible input values, in which the corresponding gRNA input signals were expressed from constitutive promoters (Table B.2). We observed fluorescence intensity differences in the digital ON and OFF states in various circuits. To distinguish circuit state, value bands for digital ON, OFF and Undefined, fluorescence values were determined with the 16 guide sequences and their cognate pGRR promoters used in circuit construction (Figure B.8). For the state of a circuit to be considered ON or OFF we specified that a majority of cell population fall in the expected fluorescence band. Population fraction tables for all circuits can be found in Table B.3. Circuits containing different NOR gate variants can exhibit a range of behaviours. For example, 15 versions of the XOR, from Figure 3.3e, constructed using different NOR gates exhibited a range of performance (Figure B.9). We hypothesize that circuit performance variations are due to expression differences in the pGRR promoters and repression efficiency variations of the gRNA in the individual NOR gates of the circuit.

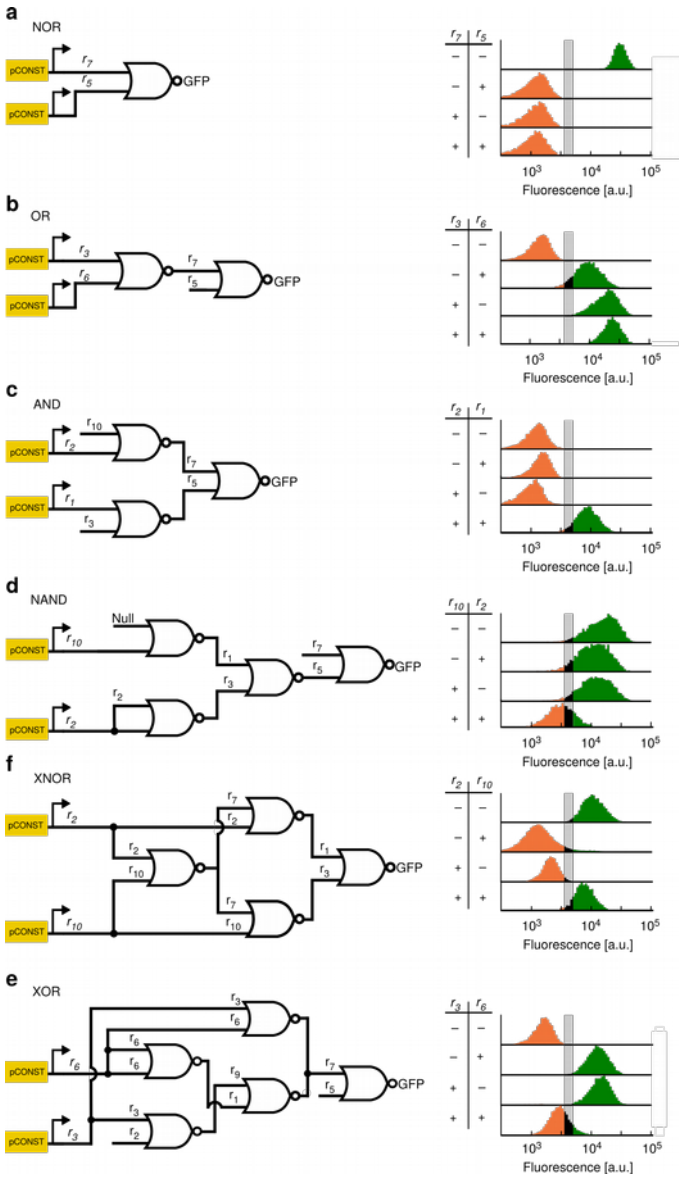


Figure 3.3: NOR gate-based logic circuits

(**a–f**) Six different two-input logic circuits constructed by interconnecting NOR gates. For each of the four input possibilities (- -,- +, + -, and + +), a distinct strain was constructed with the corresponding inputs expressed off of constitutive promoters (for logical +), or not integrated at all (for logical -). Fluorescence values were collected using flow cytometry of cells growing in log phase. The histograms represent population fraction from three different biological replicates measured during a single experiment and were normalized so that area sums to unity. Fluorescence population ratios of the circuits are included in the Table B.3.

3.4.5 Cascades

To test the limits of size and complexity our NOR gate circuits can achieve inverter cascades of depth one through seven were composed with NOT gates (Figure 3.4a). The cascade of depth D was made by the addition of a NOT gate to repress the input stage of the depth D–1 cascade. Each successive addition of a NOT gate inverter resulted in switching the behaviour of the output GFP expression. As seen previously with the two-input logic circuits, there is considerable variability within the ON and OFF states. However, circuits that are expected to exhibit ON or OFF behaviour are clearly distinguishable from one another according to our digital ON and OFF specification. As cascade depth increased the fluorescence levels of the OFF states for all of the odd depth cascades increased. Similarly, except for the cascade of depth 6, as cascade depth increased the fluorescence levels of the ON states decreased. This suggests a gradual degradation of circuit function as the number of layers increased. Similar behaviour was also observed for other repression cascades that were constructed (Figure B.10). Alternative versions of 6 gRNA cascades were constructed and showed variability in their levels of ON (Figure B.11).

To investigate the temporal characteristics of the inverter cascades, we analysed the kinetics of cascades of depth one through four. A β-estradiol-inducible promoter[128] was used to activate transcription of the input gRNA and GFP expression was periodically measured over the course of 30 h of log phase growth (Figure 3.4b). With increasing cascade depth, a clear delay in output response was evident, with the cascades reaching half-maximal expression at 4.1±0.5, 10.8±1.0, 12.0±1.2 and 17.8±1.0 h (residual s.d. deviation) for cascades of depth one through four respectively. The dose response curves of the four cascades were also measured after passaging cells over 5 days (Figure 3.4c). Consistent with the steady-state cascades, the induction of a gRNA targeting the input of the cascade switched the output of the cascade from OFF to ON (even depth cascades) or from ON to OFF (odd depth cascades). Some signal degradation with successive layers was observed (Figure 3.4c), suggesting a limit to the possible depth of the cascades.

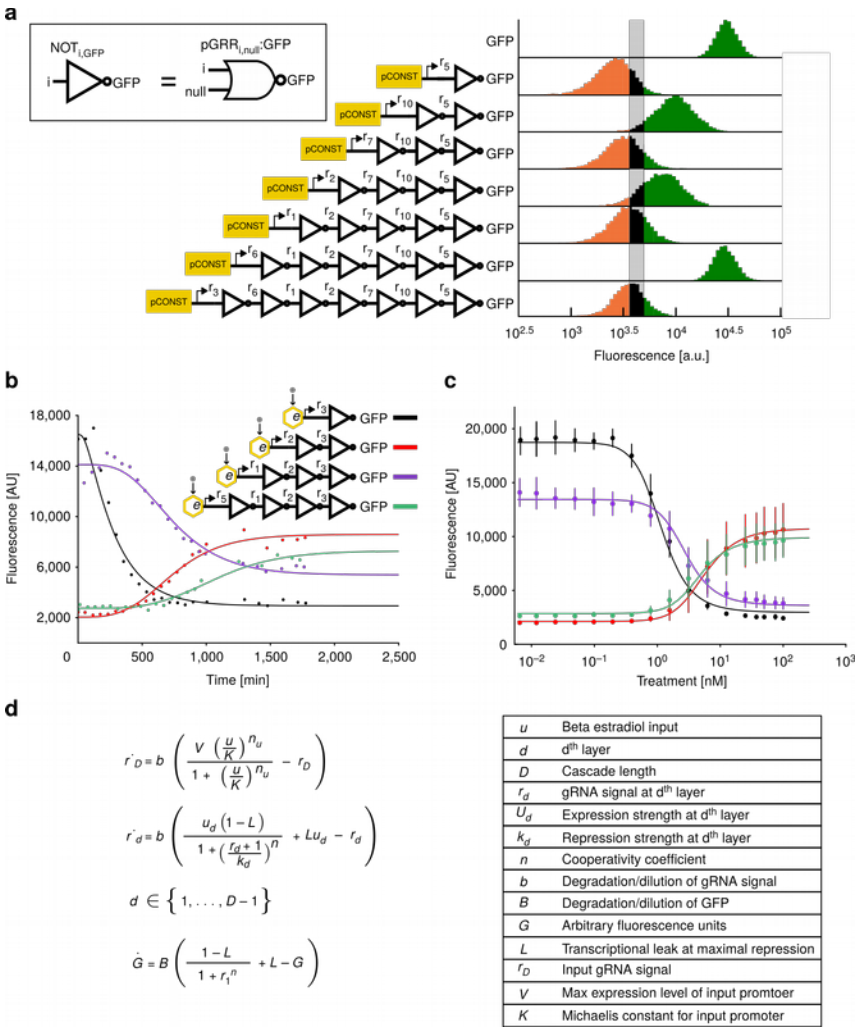


Figure 3.4: Repression cascade characterization

(**a**) Repression cascades of one to seven gRNAs. Cascades were created with sequential genomic integrations of NOT gates. The final output of each cascade is a NOT gate that expresses GFP. Each NOT gate represses the output of a subsequent NOT gate. Cascades with an even number of layers express a high level of GFP, creating a digital ON output, and odd depth cascades express low levels of GFP, creating a digital OFF output. Fluorescence measurements were taken using flow cytometry. The histograms represent population fraction from three different biological replicates measured during a single experiment and were normalized so that area sums to unity. Fluorescence population ratios of the circuits are included in Table B.3. (**b**) Temporal dynamics for cascades of one to four gRNAs. Expression of the input gRNA was induced with β-estradiol. A model of the cascade, in which each layer is treated as a Hill function, was used to fit the data. The plot shows the data from one biological replicate. As the number of layers in the cascade increases, signal degradation and increased time to steady-state is observed. (**c**) The steady-state response function for the four inducible cascades. Error bars represent the s.d. of three biological replicates measured over three separate experiments. (**d**) A representation of the model. The model was used to generate the fits for the steady-state and kinetic inducible cascade experiments.

3.4.6 Mathematical modeling

A kinetic model was constructed to capture the behaviour of our synthetic cascades. The model combines successive Hill functions to represent simple transcription and repression associated with each gRNA-dCas9-Mxi1 signal. The parameters v_d and k_d roughly capture expression and repression strengths of the promoters driving each gRNA-dCas9- Mxi1 signal, r_d. The parameter L represents the transcriptional leak as a percentage of the maximal expression of a given gate when maximally repressed parameters n and b capture the cooperativity of repression. Degradation/dilution of gRNA- dCas9-Mxi1 signals respectively (Figure 3.4d). The steady-state dose response and kinetic time course for inducible cascade data were both fit to the model (Figure 3.4b,c). Due to the different growth conditions of the steady-state and kinetic cascade experiments, two separate model fits were generated for each experiment. As inducible cascades were built in such a way that they shared many of the same pGRR and gRNA components (Figure 3.4b), parameters for the one-, two-, three- and four-layer cascades were shared between the models and fit simultaneously. To address potential model identifiability issues parameter values were constrained based on published biological values (Table B.4). The fitting results were found to correlate well with the experimental data. The measured ~18% s.d. from the mean for the promoter strength values matches well with the ~24% s.d. from the mean of the promoter strength parameters, v_d (Table B.4).

Model fits of the steady-state and time course data predict the transcriptional leak of repression due to dcas9-Mxi1, the value of L, to be effectively zero, $L = 0.6 \pm 0.1\%$ (s.d.), equivalent to the production of roughly one transcript every 5 to 10 cell divisions. The reported value of L was calculated as the average of the predicted transcriptional leak from the model fits from Figure 3.2b. To demonstrate the ability of dCas9-Mxi1 to decrease transcriptional leak compared with steric repression via dCas9, gRNA dose response curves of repression at three pGRR promoter target site positions were performed using dCas9 and dCas9-Mxi1 (Figure 3.2b). At maximal induction, dCas9-Mxi1 represses the promoter to a lower fluorescence level than dCas9 alone at all three positions. Repression via steric hindrance showed promoter positional variations in predicted leak parameter values. The observed positional variation is consistent

with previous results[107]. In all three positions dCas9-Mxi1 was predicted to have the same or lower leak parameter *L*. These data indicate that in the context of our NOR gates, dCas9-Mxi1 confers stronger and more consistent repression than dCas9 alone. Alternative plots comparing dCas9 and dCas9-Mxi1 repression as a function of inducible promoter activation driving gRNA are included in Figure B.12.

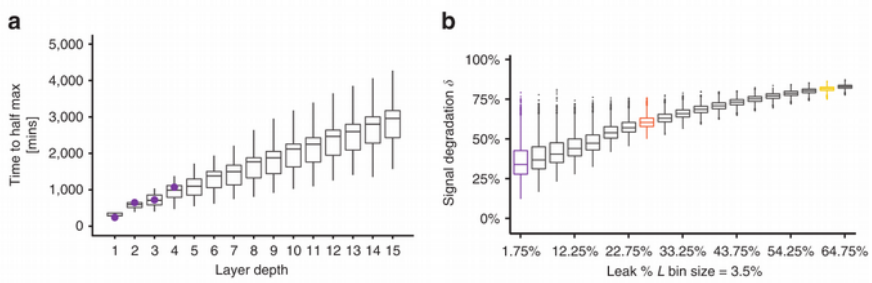


Figure 3.5: Model predictions and analysis of repression cascades

(**a**) Simulations of time to half-maximal response using the model. Increasingly layered cascades show a positive linear relationship between circuit time to half-maximal response and circuit depth, with a slope of 184.9±0.2 (s.e.m.) min layer^{-1}. The first four data points highlighted in purple are experimental data from Figure 3.4. (**b**) Signal degradation, δ, in a cascade increases as transcriptional leak of the gates increase. Boxplots of δ values were plotted with binned values of the leak parameter *L*. At values of $L < 1.75\%$ the spread of performance of the cascades is significantly larger. The bin containing the steady-state experimentally predicted value of dCas9-Mxi1, $L = 0.6 \pm 0.1\%$ (s.d.), is highlighted in purple. The bins highlighted in orange and yellow contain the predicted *L* values for the steric repression measurements in Figure 3.2 of position 1, L=25.0%, and position 3, L=61.3%, respectively.

The temporal responses of the cascades were predicted from simulations using randomly sampled parameters within the range of the model fit. Parameter values for kinetic simulations were resampled from the model fit using the kinetic time course experimental data. Response times were found to rise linearly ($r^2 = 0.83$) with increasing circuit depth. Linear regression analysis estimated the slope of the increase in response time per layer to be equal to 184.9±0.2 (s.e.m.) min layer^{-1} (Figure 3.5a), consistent with our experimental results. Response delay was found to depend primarily on the degradation/dilution rate δ of gRNA-dCas9-Mxi1 (Figure B.13) that controls the overall timescale of the dynamics.

To extrapolate the model to predict the effect of leak on signal degradation for deeper cascades, cascades of various lengths were simulated, with increasing values of *L*, using randomly sampled parameter sets within the range of dose response experimental fits. Dynamic

range of a cascade length D, ρ_D, was calculated for each cascade. Here dynamic range is defined as the log fold change change of the maximal and minimal response of a cascade, $\rho_D = \log\left(\frac{max(G)}{min(G)}\right)$. A log-linear relationship was found between ρ_D and D. This relationship was used to calculate the signal degradation, d, representing the percent loss in dynamic range per each additional layer (Figure 3.5b).

Signal degradation was found to be largely dependent on the transcriptional leak parameter, L (Figure 3.5b and Figure B.12). As leak increases, ρ, on average, increases. At values of L = 480%, the median value of d trends to ~80%. At values of L =1.75%, the spread of performance of the cascades is significantly larger. In this range the performance of the cascade is more sensitive to other parameters in the model. Our estimate of leak from the dose response experiments, L = 0.6±0.1% (s.d.), falls within the sensitive range, indicating the importance of utilizing well-performing NOR gates in large circuits built using our architecture. In addition, these data show the significance of reducing NOR gate leak when constructing larger circuits.

3.5 Discussion

We introduced a class of dCas9-based modular genetic NOR gates that behave digitally, have low variability and show minimal retroactivity or effects on cell growth. These features made these gates relatively easy to combine into Boolean logic circuits that are among the largest ever built in any organism. In particular, we found that most circuits in Figure 3.3 and Figure 3.4 required that only a handful of gate combinations be screened to identify a functional design, and others required only one.

Table 3.1 compares our technology with selected published circuits. We measured circuit complexity with a combination of two metrics: the number of gates and the number of connections among gates, allowing us to locate circuits in a two-dimensional plot (Figure B.14). We can calculate a complexity score using the two metrics, complexity = $(\text{gates}^2 + \text{connections}^2)^{1/2}$. For example, the XOR gate had five gates and four connections, producing a complexity of $(5^2+4^2)^{1/2} = 6.4$, while the cascade has a complexity of $(7^2+1^2)^{1/2} = 9.2$. These complexities compare well with gene circuits developed in *Escherichia coli*, for example. Our NOR gates enabled extremely simple design and construction of large gene circuits. Before genetic circuits can be made much larger, however, many factors that influence the size and complexity of synthetic genetic circuits must be addressed.

First, the gates in any framework must be well behaved. Gates can suffer from retroactivity, where a downstream gate affects the behaviour of upstream gates to which it is not connected by design[49,55,56]. In this case it is quite difficult to design large circuits even with CAD because we may not know the source of the retroactivity, how to model it or how to design with it. In addition, gates can be highly variable, where the outputs levels of one gate do not match the input levels of the next. Electrical engineers call this an impedance mismatch. A recent paper[16] addressed retroactivity by adding insulators to their gates. By meticulously characterizing the performance each gate, and using CAD, they were able to select compatible subsets of parts out of which they constructed circuits as large as those demonstrated here, despite gate

variability. Not all of the circuits predicted to work by the CAD tool functioned correctly, possibly due to residual retroactive effects, requiring the circuits to be screened for function. In contrast, our gates are considerably less variable and do not seem to be confounded by retroactive effects, at least in circuits with complexities <9.0. In such a case and when circuit sizes are small (<20 components) circuits are easy to design by hand since any subset of components from a library is likely to yield a functional circuit. Thus, in our case, the design problem is easy enough that extensive part characterization and CAD tools were not necessary at the circuit level (even though CAD tools such as standard DNA editors and secondary structure predictors for RNA were used at the sequence level).

Second, the host organism presents many unique challenges. Each organism can be thought of as a different computer operating system. Promoters, for example, in *E. coli* are ~60 bp of DNA long, and transcriptional regulation is a fairly well-understood process[129]. In contrast, the size of promoter and regulatory regions vary widely and can range from 250 bp to 10 kb in yeast and other eukaryotes. Transcriptional regulation in eukaryotes is complex, involving a variety of mechanisms including chromatin remodelling[130–134], and understanding it remains a highly active area of research[135]. Therefore, unfortunately, any genetic circuit technology designed for one kingdom of life is unlikely to be easily 'ported' to another, especially those built on transcriptional or translational processes. Thus, directly comparing circuit architectures between organisms, as we did between yeast and *E. coli* in Table 3.1, is difficult. Nevertheless, we believe that because CRISPR-dCas9 functions in mammalian cells[34,35,41,106,107,123] and the human Mxi1 repression domain has been used in synthetic contexts to regulate transcription in human cells[35,112,113] our NOR gates could be ported into mammalian cells, with difficulties of strain engineering likely dominating.

Third, the method by which circuits are constructed and the genetic tractability of the host affects progress toward building large circuits. For example, the circuits we present here are all singly integrated into the yeast genome, because plasmid-based systems exhibit cell-to-cell variation in copy number. That made the process of building and testing strains slow, costly and cumbersome and in fact limited our ability to build circuits much larger than those shown here.

Larger circuits and large libraries of circuit variants will require that we develop, for example, one-pot assembly methods for large DNA constructs[136]. Depending on the technology, such assemblies may be more or less difficult to harness. For example, our circuits currently benefit from the fact that the gates are integrated into disparate genetic locations that decreases the possibility of interference between gates due to chromatin remodelling[133,137] and of yeast's tendency to recombine nearby homologous regions[138].

The success or failure of different approaches to building bigger circuits may depend on how well behaved, insulated, simple and scalable the input low-level devices and gates are. In addition, relaxing the requirement that circuits be digital, so that analogue or mixed analogue/digital circuits can be used when appropriate, will likely open up the design space, further increasing the size of the circuits we can build so that one day they can match the size and performance of natural genetic circuits.

3.6 Methods

3.6.1 Construction of yeast strains

Yeast transformations were carried out using a standard lithium acetate protocol68. Yeast cells were made competent by growing 50 ml cultures in rich media to log growth phase, then spinning down the cells and washing with H20. Next, linearized DNA, salmon sperm donor DNA, 50% polyethylene glycol and 1M LiOAc were combined with 50 ml of competent cells and the mixture was heat shocked at 42C for 15 min. The cells were then spun down, supernatant was removed and they were resuspended in H2O and then plated on selective agar media. Transformations were done into MATa W303-1A and MATalpha W303-1B background strains. Matings of the MATa and MATalpha were performed by coculturing both mating types and plating the culture onto selective agar media.

3.6.2 RNA design

RGR and iRGR sequences were computationally designed to enable the 50 hammerhead ribozymes to fold into their target, functionally active, structures. ViennaRNA (RNAfold 2.1.9) was used to simulate long timescale (thermodynamic equilibrium) at an input temperature of 37C. Kinefold (kinefold_long_static_bianary 20060404) was used to simulate short timescale folding (cotranscriptional folding) with inputs of low and high polymerization rates of 25 and 50 nt/s respectively, helix minimum free energy=6.346 kcal/mol and folded without pseudoknots or entanglements. A total of 12 Kinefold simulations were run for each candidate sequence and agglomerated to generate average folding trace data. Ribozyme target structures needed for both viennaRNA and Kinefold simulation evaluation were determined by folding ribozyme sequences (Minimal HH: 5’-NNNNNNCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGT-CNNNNNN-3’ ASBV1: 5’-GGGACGGGCCATCATCTATCCCTGAAGAGAC GAAGGCTTCGGCCAAGTCGAAACGGAAACGTCGGATAGTCGCCCGTCCC-3’) using RNAfold and Kinefold (melt and anneal of 1 min), respectively. RGR targeting sequences and iRGR insulating sequences were screened in specific 50 promoter contexts (pGAL1min: 5’-

AGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAAC
TATACGGATTCTAGAACTAGTGGATCTACAAA-3', pAHD1: 5'-
CAAGCTATACCAAGCATACAATCAACTATCTCATATACAGGATTCTAGAACTAGTGG
ATCTACAAA-3', pCYC1: 5'-
ACTATACTTCTATAGACACACAAACACAAATACACACACTAATCTAGATATTGGATT
CTAGAACTAGTGGATCTACAAA-3') and in the 3' context of the targeting sequence and the gRNA handle sequence (gRNA handle: 5'-
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTT-3').

Randomly generated 20 bp candidate targeting sequences for RGR, of which the most 5' 6 bp defined the closing stem of the minimal HH ribozyme, were folded in the context of each promoter to confirm that the target structure was present in the MFE structure (viennaRNA) and that the target structure was present at > 90% in the RNA folding trace at both low and high polymerase rates (Kinefold). Targeting sequences that enabled correct folding in the context of each promoter were considered successful. For iRGRs, randomly generated 5' and 3' insulating sequences were designed for each of the three promoter types and were screened for function in the same manner. However, to select for the most robust insulating sequences, each was screened against 75 randomly generated and 10 randomly generated 20 bp guide sequences using viennaRNA and Kinefold, respectively.

3.6.3 Cytometry

Fluorescence intensity was measured with a BD Accuri C6 flow cytometer equipped with a CSampler plate adapter using excitation wavelengths of 488 and 640nm and an emission detection filter at 533nm (FL1 channel). A total of 10,000 events above a 400,000 FSC-H threshold (to exclude debris) were recorded for each sample with and core size of 22mm using the Accuri C6 CFlow Sampler software. Cytometry data were exported as FCS 3.0 files and processed using the flowCore R software package and custom R scripts (Supplementary Software 1) to obtain the mean FL1-A value at each data point.

3.6.4 Data collection for orthogonality matrix

Cytometry readings were taken with cultures inoculated into synthetic complete with cells from freshly struck out on agar. Colonies were picked from plates and grown for 3 h at 30C before reads were taken.

3.6.5 Data collection for logic circuits and static cascades

Cells from saturated culture were diluted 1:100 into fresh media with a Beta Estradiol (be) concentration of 100 nm. Cytometry measurements were taken over an 30 h period. During the time course, cells were periodically diluted to keep them in log growth phase. Experimental data collected for steady state were measured for four strains, each containing four different be-inducible cascades. Each of the four strains was induced with 18 different doses of be ranging from 0 to 100mM in a single batch of 72 cultures. Cells were diluted every 8–15 h to prevent culture saturation. Steady-state fluor- escence readings were taken after 5 days when the cultures were in log phase.

3.6.6 Model description

A deterministic model of our system was described by three ordinary differential equations characterizing transcription, degradation and repression. The gRNA-dCas9-Mxi1 and green fluorescent protein (GFP) molecular constituents were modelled as follows:

$$dr_d/dt = b\left(v_d(1-L)/\left(1+\left(\left(r_{d+1}\right)/k_d\right)^n\right)+L*v_d-r_d\right)$$

$$dr_d/dt = b\left(V*\left(u/K\right)^{n_u}/\left(1+\left(\left(u/K\right)^{n_u}\right)\right)-r_D\right)$$

$$d \in \{1\ldots D-1\}$$

$$dG/dt = B*\left(\left(1-L\right)/\left(1+\left(r_1\right)^n\right)+L-G\right)$$

r_d is the concentration of the dth gRNA-dCas9-Mxi1, d ranges from *1* to *D - 1*, where *D* is the number of layers in the cascade; r_D is the input gRNA driven by the inducible promoter; v_d is the

promoter strength driving each rd in terms of the maximum steady-state concentration of gRNA from the promoter; *G* is the measurable normalized concentration of GFP; *b* is the degradation/dilution rate of all r_d; B is the degradation/dilution for GFP; k_d is the repression strength of r_d to its cognate promoter, in terms of the number of repressors required to suppress a promoter to half strength; to its cognate promoter is modelled with k_d, the number of repressors required to suppress a promoter to half-strength; and n is a Hill coefficient. For the transfer function, *V*, *K*, n_u respectively represent the maximum transcription, Michaelis–Menten constant and Hill coefficient of the inducible promoter; *u* is the input be in mM. Concentration is rescaled as the Michaelis– Menten constant or the number of gRNAs required to suppress a NOT gate to half- maximal. Note that the model makes the assumptions that (1) there is no crosstalk between gRNA components, (2) Mxi1 represses transcription completely with no transcriptional leak and (3) dCas9-Mxi1 bind quickly and irreversibly to gRNA.

3.6.7 Fitting procedure

Parameters were optimized using differential evolution fol- lowed by minimization using the BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm[139]. For the steady-state experiments, optimal parameter fits for the parameters v_0^{ss}-v_3^{ss}, k_0^{ss}-k_3^{ss}, V^{ss}, n^{ss} were generated from three separate experiments. For each of the three experiments, 17 parameter fits were generated using differential evolution/BFGS and means were calculated for a total of 51 steady-state parameter sets. The means from each experiment were used to determine the experimental error (s) for estimating each parameter (Table B.4). For the kinetics experiments, five parameter fits for $v_0^{kinetics}$ - $v_3^{kinetics}$, $k_0^{kinetics}$ -$k_3^{kinetics}$, *b*, *B*, $V^{kinetics}$, $n^{kinetics}$ were generated from a single experiment (Table B.4). As there were only data for a single kinetics experiment, experimental errors for the kinetic parameter values were not calculated. Parameters *K* and n_u were determined in a separate experiment by driving a YFP with the pGALZ$_4$ β-estradiol inducible; this promoter is the same promoter used in the inducible cascades. The kinetics and steady-state parameter sets were resampled in downstream analyses to generate Monte Carlo simulations of longer repression cascades.

3.6.8 Model predictions

Long repression cascades of 1 to 11 ($D \in \{1 \ldots 11\}$) layers were simulated using the system of ordinary differential equations. Parameters for simulated cascades were generated by resampling parameter sets generated during the fitting procedure. For the kinetic model predictions, 10,000 simulated cascades were generated by resampling parameters from 5 parameter sets estimated from the kinetics experiment. The time-to-half max of GFP (G) was calculated for each cascade length D and plotted in Figure 3.5a. For the signal degradation (d) predictions in Figure 3.5b, 100,000 simulated cascades of length $D = 7$ were simulated by resampling parameters from the 51 parameter sets estimated from the 3 steady-state experiments. To compare L versus δ, L was sampled from a uniform distribution between 0 and 1. Signal degradation (δ) was calculated as the percent change in dynamic range per additional layer. The dynamic range at each layer d in a cascade of length D was calculated as:

$$\rho_d = \log\left(\frac{max(r_d)}{min(r_d)}\right)$$

Dynamic range was found to have a log-linear relationship with the length of the cascade, and hence the average slope between d versus $\log(r_d)$ was calculated using linear regression for each of the 100,000 simulations of cascades of length D by:

$$\eta = \frac{D\sum(d\log(\rho_d)) - (\sum d)(\sum \log(\rho_d))}{D\sum(d^2) - \left(\sum(d)\right)^2}$$

With $D = 7$. With η being the change in $\log(r_d)$ with each additional layer, the percent loss in dynamic range per layer or signal degradation δ is calculated as

$$\delta = 1 - 10^{\eta} \quad .$$

Values for L were binned using a bin size of 0.035 and d versus L was plotted to generate Figure 3.5b.

Natural genetic circuits enable cells to make sophisticated digital decisions. Building equally complex synthetic circuits in eukaryotes remains difficult, however, because commonly used components leak transcriptionally, do not arbitrarily interconnect or do not have digital responses. Here, we designed dCas9-Mxi1-based NOR gates in *Saccharomyces cerevisiae* that allow arbitrary connectivity and large genetic circuits. Because we used the chromatin remodeller Mxi1, our gates showed minimal leak and digital responses. We built a combinatorial library of NOR gates that directly convert guide RNA (gRNA) inputs into gRNA outputs, enabling the gates to be 'wired' together. We constructed logic circuits with up to seven gRNAs, including repression cascades with up to seven layers. Modelling predicted the NOR gates have effectively zero transcriptional leak explaining the limited signal degradation in the circuits. Our approach enabled the largest, eukaryotic gene circuits to date and will form the basis for large, synthetic, cellular decision-making systems.

Chapter 4. Engineering predictable CRISPRi transcriptional networks with ribozyme- and aptazyme-regulated gRNA expression

4.1 Abstract

CRISPR-Cas transcriptional networks that regulate the expression of gRNAs and genes have enormous potential for engineering programmed biological responses. Here, we developed an experimental and modeling framework for characterizing Pol II gRNA expression architectures through the quantitative analysis of 44 one-, two- and three-layer CRISPRi repression cascades in yeast. gRNA are expressed using a Ribozyme-gRNA-Ribozyme (RGR) architecture. By measuring the responses of CRISPRi repression cascades as sensitivities to RGR-inputs, design elements impacting gRNA expression and function from RGR were identified. These insights led to the creation of the Ribozyme-gRNA-Aptazyme (RGA) architecture. The response dynamics of this architecture were readily modeled and as much as a 10-fold increases in ligand-dependent system-output were observed. We further show that limited dCas9 expression significantly alters system response, likely caused by gRNA competition. These effects are mitigated in systems with excess dCas9, a condition where system behaviors are accurately predicted from individual component characteristics. Notably, there was fair agreement between the model-derived predictions and experimental outcomes for a large, 7 gRNA, XOR logical evaluation circuit. Though studying a CRISPRi engineering framework, this work addresses key barriers in creating complex CRISPR-Cas information processing systems.

4.2 Introduction

A robust and modularized platform that can be used to build predictable gene regulatory networks will have significant implications for tissue engineering[4], cell-based analytics[5], and metabolic engineering[6]. Such platforms have been created using CRISPR-Cas9 networks based on nuclease inactive Cas9 (dCas9), derived from *Streptococcus pyogenes*[29]. These networks are composed of multiple guide RNAs (gRNAs) that, upon binding to dCas9, regulate their own transcription and the transcription of specific genes¹⁴⁰. They have been engineered in bacteria[40], yeast[42] and mammalian[41] cells. In yeast, these networks have been used for logical evaluations requiring the programmed expression of as many as 7 gRNAs in the same cell[42] and multi-gene metabolic control[141]. Despite these proofs-of-concept, much is unknown about how to engineer and characterize the dynamics of CRISPR-Cas9 transcriptional regulation. There are many validated genetic platforms to express gRNA *in vivo*, however, of the few that have been modularized little work has been done to understand how engineerable they are as a platform. Even with robust gRNA expression, simultaneously expressed gRNA may influence each others dynamics through dCas9 competition. As a consequence, the construction of multi-gRNA CRISPR-Cas transcriptional networks can be unpredictable and often requires trial-and-error screening which limits their realizable complexities[42]. Here, we leverage our previously developed CRISPR interference (CRISPRi) platform for building transcriptional networks[42] to study these phenomena in yeast.

CRISPRi represses gene expression through dCas9 mediated steric interference of polymerase binding[34,106]. Our platform uses a dCas9-Mxi1 fusion[35] that makes dCas9 a stronger repressor and less sensitive to promoter target location[42]. For dCas9 to function it must first bind to a gRNA. This short non-coding RNA is composed of a structured handle sequence and a 20-base target sequence that addresses the gRNA-dCas9 complex to its complementary DNA sequence[29]. Though the sequence of the gRNA handle[142,143] and gRNA target[144–146] have been studied and found to have significant impact on CRISPR activity, much less is understood about how *in vivo* gRNA transcription impacts CRISPR activity.

gRNA transcription for our CRISPRi networks is driven by an engineered set of gRNA responsive Pol II promoters (pGRR). We avoid using non-coding RNA Pol III promoters to express gRNA[41,147–150] as their sensitivity to sequence modifications[120,151] makes them ill-suited to build multi-gRNA systems. Though Pol II are readily engineerable[121], the RNA they express undergoes post-transcriptional modifications that append a 5' cap and a 3' poly-A tail. These modifications are thought to mediate nuclear export, rendering gRNA unable to bind its target in the nucleus[152]. Strategies that utilize trans-acting enzymes to mediate post-transcriptional removal of these signals have been successful, but these are not easily portable between species[153]. Consequently, our system removes the signals by encoding cis-cleaving ribozymes (Rbz) that flank the gRNA, creating an Rbz-gRNA-Rbz (RGR) architecture[123].

RGRs have been used as a successful means of Pol II-mediated gRNA expression in mammalian[108], amphibian[154], fungal[155] and yeast[42] hosts. Despite their host portability and prevalence, little work has been done to characterize the versatility of the architecture and to understand how variations to RGR design, such as Rbz structures and post-cleavage 'scars', impact gRNA expression and function. In fact, there are even questions about whether 5' *cis* Rbz cleavage of RGR transcripts is necessary[42,156]. Despite successes in appending RNA ligand-sensing aptamers to other gRNA expression systems to create inducibly active gRNA[78,157], no work has been done to extend such functionality to RGRs. A better understanding of the RGR architecture and proof-of-concept engineered extensional applications will further validate its use as a gRNA expression platform.

The expression of multiple gRNA, which is necessary to create CRISPRi networks, has been shown to unevenly decrease individual gRNA activities[142,158,159]. The implication is that gRNAs with higher Cas9 binding affinities out-compete gRNAs with lower affinities. The effects of gRNA competition has been speculated to impact the behaviour of transcriptional network functions compared to what would be expected on the basis of the activities of individual gRNAs[44]. If true, the capacity for a gRNA to direct transcriptional regulation may depend not just on gRNA expression[160], but also on the levels of dCas9 and other gRNAs in the

cell[158]. To our knowledge, the magnitude of these effects and the degree to which they influence the ability to predict CRISPRi transcriptional networks behaviors has not been addressed.

Here we develop an experimental and modeling framework and us it to study the robustness of RGR architectures, their effect on gRNA expression and effects of gRNA competition for dCas9. This was accomplished through the construction and analysis of 44 CRISPRi transcriptional networks with inducible RGR inputs and measurable fluorescent outputs. We find that RGR inputs without 5' Rbzs require 1.8-3.6 fold more transcriptional input to achieve the same output as their RGR counterparts and that RGR inputs without 3' Rbzs result in non-functional gRNA expression. We examine the versatility of the RGR architecture by substituting the original 5' and 3' Rbz with five different hammerhead Rbz variants with diverse sequences and secondary structure topologies. We find that functional RGRs can be engineered from these Rbz and show that post-cleavage scar sequences can cause up to 10-fold reductions in gRNA activity. We then leverage this understanding to create a novel Rbz-gRNA-Aptazyme (RGA) architecture for engineering ligand-dependent gRNA expression. This RGA architecture is straightforward to implement and was shown to generate a quantitatively-predictable 10-fold increase in CRISPRi network outputs. We study the effect gRNA competition for dCas9 has on CRISPRi transcriptional network behavior by varying the level of dCas9 protein expression in otherwise identical networks. Multi-gRNA networks with decreased dCas9 expression significantly deviate from networks with high dCas9 expression requiring, at the extremes, ~1.6x and ~0.3x the transcriptional input. We then compare model predicted network responses to experimental responses to quantify dCas9 competitive effects. Predictions of systems with high levels of dCas9 agreed well with system responses indicating only minor gRNA competition, however predictions of systems with low dCas9 saw significant deviations (up to 60% loss in explanatory power) indicating significant gRNA competitive effects. Finally, we investigate the applicability of using characterized network behaviors to estimate the aggregate performance of a 7-RGR two-input, one-output Boolean XOR gate in yeast[42]. Our modeling approach was able to fairly predict output responses and identify input-states that are most likely to fail. In sum, we have identified RGR design elements needed for efficient gRNA activity, created a new RGA

architecture for aptazyme-regulated gRNA expression, identified gRNA competition for dCas9 can dramatically effect CRISPRi networks and shown that these networks can be predicted. Our work has broad implications for engineering multi-gRNA transcriptional networks and addresses key barriers to engineering even larger CRISPR-Cas transcriptional regulatory systems.

4.3 Methods

4.3.1 Plasmid construction

All plasmids constructed were for stable integration into the yeast genome. All sequences were based on pGADT7-Rec 2micron plasmid (Clontech). Recurring genetic elements including promoters, yeast integration makers, ribozymes, gRNA, and open reading frames are listed in Supplemental Data. All DNA genetic elements were isolated from existing plasmids with PCR amplification or were directly synthesized (IDT geneblocks). The Gibson method was used for plasmid assembly[161]. All cloning plasmids contained AMP resistance and were ultimately transformed into DH5α E. coli and plated on LB agar plates containing 50 mg/ML ampicillin. Multiple colonies were from each plate were picked and Sanger sequenced (Gene Wiz) to verify assembly. Correctly identified colonies were grown overnight at 37C in LB under 50 mg/mL ampicillin selective pressure. Plasmid DNA was extracted and purified using QIAGEN miniprep kit.

4.3.2 Creation of stable yeast strains

Ultimately all strains tested were diploid W303 Saccharomyces cerevisiae strains. All components were first integrated into the haploid background strains MATa W303-1A and MATalpha W303-1B. Stable yeast transformations were done using a lithium acetate protocol[162]. Plasmids were first linearized with a Pmei (NEB) digestion. Linearised fragments are mixed with necessary reagents and transformed at 42C. Cells were plated on SDO with the appropriate selective pressure. Following this transformation colony PCR were used to confirm that the gene

element was successfully integrated. Stable diploid yeast strains were generated by co-culturing MATa and MATalpha strains and subsequently plating the culture onto SDO with two selection markers that were unique to MATa and MATalpha strains.

4.3.3 Yeast growth condition

For each experiment all strains tested were taken from plates streaks from glycerol stocks that were no older than ten days. All experiments were carried out in 96 deep well plates. For each tested strain, three colonies were picked to inoculate three wells of the plate containing 1 mL of SC media. This plate was then incubated overnight overnight at 30C shaking at 450 RPM. The following day wells containing cells were back diluted 1:1000 in 1mL fresh SC containing the required inducer, and were incubated under the same conditions for 16 hours. To best approximate steady state growth, these cells were then diluted 1:50 in 1mL fresh media with appropriate inducers and allowed to grow for an additional 6 hours.

4.3.4 Testing RG, GR, iRGR, RGiR devices

In this work many strains are directly compared to another strain. All strains that were tested were done at the same time as their RGR counterpart. This allowed for us to account for any variation in day-to-day variability in growth conditions. For all assays three biological replicates were completed. Yeast strains were tested with either six or eight concentrations of beta-estradiol.

4.3.5 Testing RGA one-layer cascades

Three biological replicates were completed for each condition. Two beta-estradiol series (six levels of induction) were completed at 0 and 5 mM theophylline. In addition, two theophylline series (six levels of induction), were tested at three levels of beta-estradiol induction. Finally a single biological replicate was run for a positive control (diploid strain containing with no RGR expression) and a negative control (diploid strain with no fluorescent reporter) at the same theophylline series.

4.3.6 Flow cytometry

The fluorescence of all yeast strains were measured using flow cytometry. We used a BD Accuri C6 flow cytometer equipped with a CSampler plate adapter. sfGFP fluorescence was measured using excitation wavelengths of 488 and 640 nm and an emission detection filter at 533 nm (FL1 channel). Briefly 100 uL of cells are transferred from 96 deep well plate to a 96 well assay plate. At least 20,000 cells were measured above a 400,000 AU FSC-H threshold (to screen out cellular debris). These data were then gated for the healthy yeast cells (SSC-A, FSC-A) and singlet cells (FSC-H, FSC-A) and the resulting median FL-1 values were calculated using custom python scripts.

4.3.7 Calibrating and normalizing experiments

Calibration curves mapping beta-estradiol input-concentrations to transcriptional strength were generated using a diploid calibration strain where pGalZ4[163] was driving the expression of eYFP. Each beta-estradiol stock solution generated in the course of experiments was calibrated using this strain with three biological replicates. Each replicate was exposed to six to twelve concentrations of beta-estradiol. Median fluorescent values determined by flow-cytometry were background subtracted using a no fluorescence negative control strain. The no induction controls were subtracted to a level of 20-25 arbitrary units to account for known pGalZ4 transcriptional leak. The resulting data was then fit using a Hill equation. The resulting Hill equations were normalized to a value of 1 and this level of fluorescence was fixed to be 1 TU.

The maximum expression of each promoter used in this study was converted to TU by normalizing their fluorescent expression in a control strain to the maximally induced pGalz4 Promoter. A control strain is a diploid strain expressing yeGFP from the relevant promoter, expressing dCas9 from a pGPD promoter, and with no expressed gRNA targeted to the relevant promoter. Experimentally tested strains are grown concurrently with background controls and promoter controls. Their induced response is normalized to this range which is then converted to the maximal TU the control strain was determined to have. The above beta-estradiol calibration curve is then used to convert input beta-estradiol concentrations to TU.

4.3.8 Fitting system response with a Hill-equation

The input-output response of experimentally tested strains was modeled using a Hill-equation. Once the experimental beta-estradiol input and fluorescent output have been converted to TU, data are fit with the Hill equation [f(L, vMin, vMax, n, Ka) = (vMax - vMin)/((Ka/L)^n + 1) + vMin] where L is input TU, vmin is the minimum expression, vmax is the maximum expression, n is the Hill coefficient, and ka is the input amount resulting in half-saturation of the response. To avoid under-fitting low levels of system output, output was log10 scaled and data was fit to a log10 transformed Hill-equation. Specifically these data were fit using Python (2.7.6) scripts which call the SciPy (0.18.1) optimize.curve_fit() function. The vmin parameter was bound to be no less than 10-2 TU.

4.3.9 Quantifying relative sensitivity and sensitivity ratio

We developed sensitivity metrics to quantitatively compare the input-output response of CRISPRi systems over their shared linear response range. A comparison is made between a reference data-set and experimental data-set. After hill-equations are fit to both data-sets the linear response of each equation are coarsely determined by isolating output region which is 50% of the maximum response, vmax, and 150% the minimum response, vmin. For each data-point that falls in this region, the ratio of the raw-input level is taken to the input-level that the reference hill-equation to achieve the same output as the raw-output. Consequently the average sensitivity of the reference data is ~1. The resulting population of relative sensitivities for the experimental and reference device were then compared using a two-tailed Welch's t-test. If pval < 0.05 we consider the means of the population to be significantly different and we compute the sensitivity ratio by taking the ratio of the experimental to the reference mean.

4.3.10 Simulating CRISPRi systems

CRISPRi transcriptional systems were simulated with the modeled hill-equation parameters of components in one-layer contexts. All simulations completed using Python (2.7.6) scripts. To simulate two-layer cascades the hill-equation of the input component is propagated

through the internal hill-equation. The simulated two-layer response is generated by fitting resulting input-output response. XOR gate systems were simulated by first creating the XOR gate architecture of connected NOR gate nodes. Each node contains hill equations parameters which define its input-output relationship. To simulate a desired input-state the input of level of the respective input nodes is set based on input state (ON value = 1 TU, OFF value = 0.001 TU) and then these nodes outputs are propagated through all downstream nodes until the response of the ultimate node is determined. To simulate aggregate XOR gate performance, 20,000 random XOR gates were created through random sampling of a normal distribution of hill-equation parameter bounds (Supplemental Table S4) and each was solved for each of the four XOR input-states. Finally to simulate the effects of using inputs with differences in relative sensitivity,1,000 random samplings of XOR systems were solved with the ka parameter for each node scaled by a constant value.

4.3.11 Computational design of insulated ribozyme devices

Novel RGRs were computationally designed to bias the Rbz to fold into their putative catalytically active secondary-structures using minimum free energy[72] and kinetic co-transcriptional RNA folding simulations[79,81].

For iRGR design devices were tested in the context of pGalZ4 promoter with the upstream sequence of the ribozyme being described as (5'-
CCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCGCGCCG
CACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAA
AAATTGGCAGTAACCTGGCCCCACAAACCTTCAAATTAACGAATCAAATTAACAAC
CATAGGATGATAATGCGATTAGTTTTTTAGCCTTATTTCTGGGGTAATTAATCAGCG
AAGCGATGATTTTTGATCTATTAACAGATATAT

AAATGGAAAAGCTGCATAACCACTTTAACTAATACTTTCAACATTTTCAGTTT
GTATTACTTCTTATTCAAATGTCATAAAAGTATCAACAAAAAATTGTTAATATACCT
CTATACTTTAACGTCAAGGAGAAAAAACTATACGGATTCTAGAACTAGTGGATCTAC
AAA-3'). Following this is a 5' insulating sequence that is allowed to be 1-10 bp long, followed

by the ribozyme then followed by another 1-10 bp insulating sequence and then a 20 bp target sequence which is forced to have a GC content between 35 and 65 percent. This is followed by a gRNA handle sequence (5'gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgctttt-3') and finally an HDV ribozyme (5'-ggccggcatggtcccagcctcctcgctggcgccggctgggcaacatgcttcggcatggcgaatgggac-3'). 400 initial insulating candidate sequences were generated for each ribozyme. An insulating candidate consisted of randomly created, in length and sequence, upstream and downstream insulating sequence. An insulating sequence candidate made it into the initial pool if they enabled the ribozyme to fold into its target secondary structure in the previously specified sequence context in the presence of 75 randomly generated target sequences, as determined by thermodynamic minimum free energy (MFE) folding simulations (ViennaRNA, RNAfold 2.1.9: RNAfold -d2 -T37) [72]. The target secondary structure of a ribozyme was determined by its MFE structure when folded in isolation (RNAfold -d2 -T37). Each candidate insulating sequences that passed thermodynamic screening was then screened using the kinetic co-transcriptional RNA simulations (kinefold_long_static_bianary 20060404, helix minimum free energy=6.346 kcal mol-1 and folded without pseudoknots or entanglements) [81]. For each insulating sequence 20 target sequences are folded 10 times at a random polymerization rate from 10 and 30 nt/s. Sequences are folded in the sequence context starting from the pGalZ promoter sequence and going N bases past the 3' base of the ribozyme, where N = (polymerization rate)*1 sec. Each set of 10 simulations are aggregated into folding trace data which enables the quantification of specific substructures over time. Candidate sequence are then scored by the time average fraction of the population that folded into the target ribozyme structures (target ribozyme structure determined from 100, 1 minute renaturation simulations kinefold_long_static_bianary 20060404, helix minimum free energy=6.346 kcal mol-1 and folded without pseudoknots or entanglements), determined by dividing the numerically integrated trace by its simulation time. By this metric, the top 80 insulation candidates are then screened against 80 more random target sequences sequences. Finally the best average performing set of insulating sequences is selected. The same method was used for RGiR design but the context was now considered to be 5' - target

sequence – gRNA handle – 5’ insulating sequence – ribozyme – 3’ insulating sequence – terminator (5’ – tgataccgtcgacctcgagtcaattagttatgtcacgcttacattcacgccctccccccacatccgctctaaccgaaaaggaaggagttagacaacctgaagtctaggtccctatttatttttttatagttatgttagtattaagaacgttatttatatttcaaatttttctttttttttctgtacagacgcgtgtacgcatgtaacattatactgaaaaccttgcttgagaaggttttgggacgctcgaaggctttaatttg – 3’).

4.3.12 *In vitro* ribozyme cleavage rates

50 µL transcription reactions were set up with 10X NEB T7 Reaction Buffer (supplied with NEB M0251) supplemented with 6 mM MgCl2 for a final concentration of 12 mM MgCl2, 120 Units RiboLock RNase Inhibitor (ThermoFisher EO0381), 250 Units NEB T7 RNA Polymerase (NEB M0251), 3 mM each of ATP, UTP, CTP, and GTP (NEB N0450), and between 25 and 100 ng of linear DNA bearing the ribozyme sequence in context. Transcriptions were run at 37° C on a thermocycler; 10 µL from each transcription reaction was removed and added to 20 µL of a transcription stop solution of 80% formamide and 10 mM EDTA at 4° C. We measured cleavage at at 30 minutes.

Quenched transcription reaction time points were stored at -80° C immediately after collection, then heated to 90° C for 2 min and set on ice prior to analysis by PAGE. Low range ssRNA ladder (NEB N0364) was run as a reference for PAGE; ladder was prepared by mixing ladder with formamide loading buffer (80% formamide, 10 mM EDTA, 1 mg/mL xylene cyanol, 1 mg/mL bromophenol blue) at the ratio of 1:19 and heating at 90° C for 2 min followed by storage on ice. Transcription products were run on a denaturing Ureagel 10% polyacrylamide gel for about 90 minutes at 7 W; polyacrylamide gels were stained with SYBR Gold (ThermoFisher S11494) in 0.5X TBE for 10 minutes and imaged using a Bio-Rad Pharos FX molecular imager. Gel images were analyzed using ImageJ (Version 1.50b); the relative amounts of cleavage products were calculated by measuring the intensity of bands at positions corresponding to the length of the expected uncleaved and cleaved RNA products.

4.4 Results

4.4.1 Quantifying system sensitivity to gRNA inputs expressed from pol II promoters

To investigate how RGR variations impact CRISPRi transcriptional network activities, we first created a framework for their characterization (Figure 4.1). The immediate goal was to identify how functional RGR elements and gRNA identity impact gRNA function when expressed from Pol II promoters. To study this, input-output response dynamics of a transcriptional networks with a single RGR-input, an RGR expressed from an inducible promoter, are analyzed. Dynamics are experimentally generated by titrating the expression of the RGR-input and measuring the system response. Any RGR-input that causes a change in the expressed gRNA efficacy will likely necessitate compensatory gRNA transcription making networks appear more or less sensitive to their inputs. Consequently, we made quantitative comparisons of 'sensitivities' between the input-output response of networks which only differ in their RGR-input. Unless otherwise stated, comparisons are made relative to the canonical RGR architecture that is defined by its 5' and 3' Rbz sequences and structures[123]. In this framework, RGR design variations that improve gRNA expression and function make the network more responsive, or sensitive, to RGR inputs.

In the experimental yeast systems stably integrated RGR-inputs were expressed from a β-estradiol inducible pGalZ4 promoter[128] which then repressed transcription of either an RGR or a reporter promoter (Figure 4.1A). The β-estradiol input and fluorescent-output of systems were converted to an arbitrary unit (TU) which was normalized to pGalZ4 at maximal induction to be a value of 1 (Figure 4.1B). On this scale, the strongest constitutive yeast promoters (pGPD/pTDH3) are on the order of 10 TU[164]. The input-output response data were then modeled with a Hill-equation that estimates the sensitivity of the system (ka), the responsiveness of the system (n) and the output dynamic-range of the system (vmax, vmin) (Figure 4.1C and D).

Experimental data and Hill-equation fits were used to quantify the 'relative sensitivity' between "reference" and "experimental" system responses. For all data that fell within the shared

linear output response of the reference and experimental Hill-equations, we calculated a ratio of the experimental input-level to the input-level a reference Hill-equation required to achieve the same output (Figure 4.1D). If the populations of relative sensitivities between the reference and experimental responses were statistically different we take a ratio of their means to calculate "sensitivity ratio" (SR). The SR quantifies how much more transcription the experimental system requires to achieve the same system output as the reference system.

We first studied the effects of removing either the 5' ribozyme (5'-GR) or 3' ribozyme (RG-3') from the canonical RGR (Figure 4.2A). gRNA target sequences W8 and W10, previously characterized to be functional and orthogonal [42], were used as inputs to investigate if defects were intrinsic to target sequences. These inputs were tested in both a one-layer repression cascade with the input RGR directly repressing the reporter promoter and in a two-layer repression cascade where the input RGR represses the expression of an internal RGR which otherwise represses the reporter promoter (Figure 4.2B).

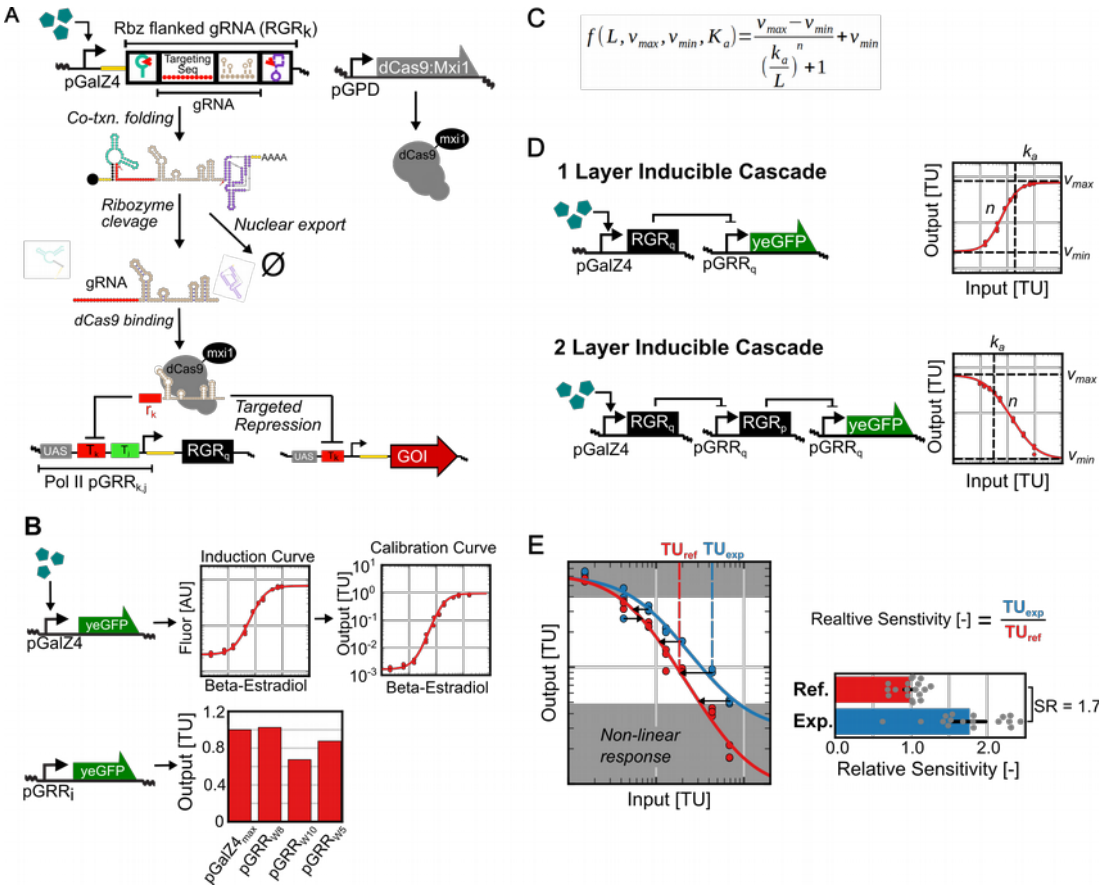


Figure 4.1: Quantifying the relative activity of gRNA-directed transcriptional inhibition.

(A) Biochemical overview of dCas9-Mxi1 repression. An RGR is expressed from a beta-estradiol inducible promoter. Ribozymes mediate the removal of nuclear export signals from gRNA which represses engineered promoters. (B) The inducible promoter pGalZ4 was calibrated using a fluorescent reporter with a Hill function. The maximal expression of pGalZ4 was normalized to be 1 TU. All reporter pGRR promoters used in this study were normalized to pGalZ4max to determine the expression strength in TU. (C) The specific form of the Hill function used for all fits. (D) The system architecture and resulting system response with Hill function parameters indicated for one and two layer repression cascades. (E) Method for calculating sensitivity ratio. An experimental dataset (blue) is compared to a reference dataset (red). Both data sets are fit with a Hill function. Only data which falls in the shared linear Hill function response are considered (not shaded). A sensitivity-ratio is calculated by taking the ratio of the amount of input expression for the experimental device to the amount of input the reference device's Hill function would require to achieve the same level of output for each experimental datapoint. All of the data are then collected and averaged to compare how much more input on average is required to achieve the same level of output.

The absence of a 5' Rbz (GR) decreased gRNA expression efficiency and the absence of a 3' Rbz (RG) abolished gRNA activity, in both one- and two-layer repression cascades. The input-output response of both one- and two- layer systems showed that inputs GRW10 and GRW8 required more input expression to achieve the same systems repression compared to their RGR counterparts. More extremely, input RGW8 showed no dynamic-output response (Figure 4.2C). We quantified the relative sensitivity for the GR and RG devices relative to their RGR counterparts (Figure 4.1D). GRW8 required more relative expression from a one- to two-layer cascade (two-tailed t-test; n=15,11; pval = 0.006) than their RGR counterparts with SRs of 1.8 and 3.5, respectively. On the other hand, GRW10 showed a consistent SR (two-tailed t-test; n=10,18; pval=0.5) for one- and two- layer cascades with SRs of 3.5 and 3.3, respectively. Though both GR-gRNA inputs showed that a lack of 5' Rbz processing reduces expressed gRNA efficacy, their distinct SRs and unequal changes in SR with increasing system size imply that the extent of expression defects are a function of gRNA identify and that these defects can become worse with increasing system size. Because the RGW8 did not exhibit a linear output in the one- and two- layer cascades, the SRs are estimated to be at least 45 and 150, respectively. This demonstrated that 3' Rbz processing is essential for functional gRNA expression from RGRs.

In summary, we found that quantifying the sensitivity of CRISPRi repression cascades to titrated gRNA inputs allows us to evaluate how RGR design variations affect gRNA expression efficiency. The activities of gRNAs expressed from RGRs without 5' ribozymes can require 3.5x more expression and that the 3' ribozyme is essential for functional gRNA expression. gRNA identity influences the extent of decreased gRNA activity from GR devices and finally, expression defects may be exacerbated in larger systems.

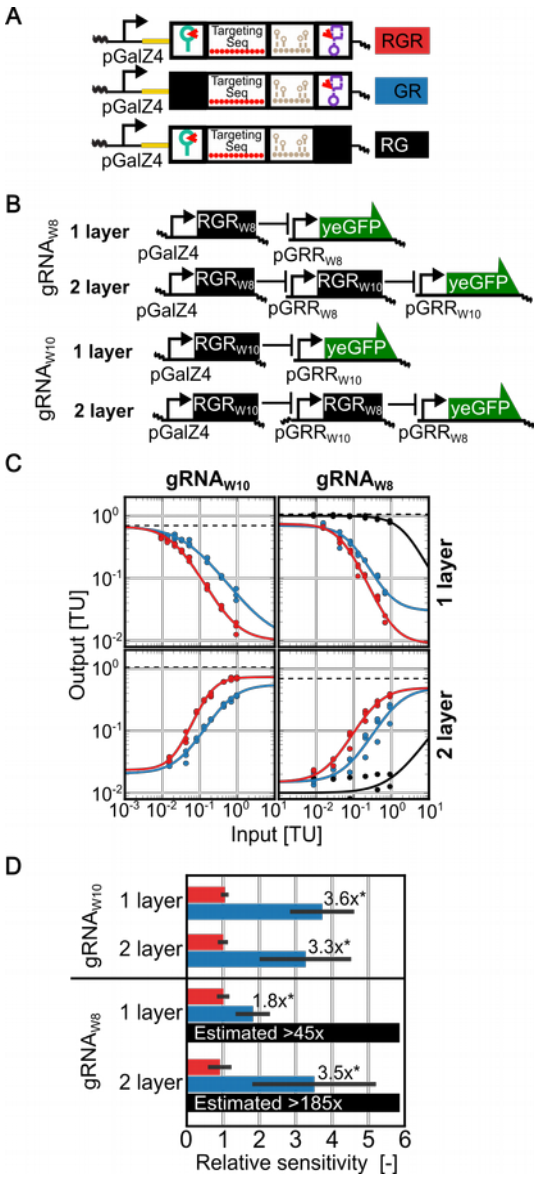


Figure 4.2: Analyzing one- and two- layer cascade performance.

(A) Description of the three color coded input architectures used for the experimental one-layer systems used to investigate the utility of 5' and 3' ribozymes. (B) System architecture for the gRNAW10 and gRNAW8 one- and two- layer cascades that used in this study. (C) Biological triplicate titration response data with hill equation fit for one- and two- layer systems is shown. (D) Fold differences in input sensitivity between inputs and their respective RGR control are indicated. The RG device data was outside of the range of the calculation so the estimated decrease in input sensitivity is estimated using ka values. * indicates pval<0.0001 calculated using two tailed Welch's t-test. Error bars indicate ±STD.

4.4.2 RGRs can be engineered from diverse ribozymes

We next turned to the question of how functional gRNA expression depends on the sequences, structures and activities of the RGR Rbzs. To test this we designed two classes of novel RGRs which substitute the canonical RGR Rbzs with one of five well-characterized hammerhead Rbzs, each with a distinct sequence and secondary structure topology (Figure 4.3, Figure C.1, Table C.1). These designed RGRs contain either an insulated 5' Rbz (iRGR), that unlike the canonical RGR Rbzs are sequence independent of the target sequence[42], or an insulated 3' Rbz (RGiR). Unlike the canonical RGR, iRGR and RGiR architectures produce gRNA with 5' and 3' cleavage scars, respectively (Figure 4.3A). Given that cis sequence interactions between gRNA target sequences and the gRNA handle[159], and that 5' and 3' gRNA sequence extensions have been engineered to reduce gRNA activity[78,165], we expected these residual, post-cleavage scar sequences could negatively impact gRNA performance. Here we designed and tested eleven iRGR and RGiR designs (Figure C.2, Table C.1).

The insulated RGR were computationally designed to bias the Rbz to fold into their putative catalytically active secondary-structures (see methods). Briefly, short insulating sequences (1-10 bp) that flank the Rbz were selected that bias the Rbz to fold into its active secondary structure within the RGR sequence context. All insulated RGR devices, but iRGR ASBV-1, were found to have cleavage activity equivalent to, or greater than, the canonical RGR Rbzs (Figure 4.3B) with an in vitro co-transcriptional cleavage assay (see methods). Insulated RGR were tested as inputs driving the expression of gRNAW8 in a one- and two-layer cascade (Figure 4.2B). The titrated responses of each iRGR and RGiR (Figure C.3) was used to generate SRs relative to their respective RGR control (Figure 4.3B).

The majority of designed RGRs performed worse than their canonical RGR counterpart. All five of the iRGR devices had decreased sensitivity, with an average SR of 4.5. Despite these devices having high 5' cleavage activity, all tested iRGR performed equally or worse than the no 5' Rbz control (GR). In contrast, two of the six RGiR devices performed comparably to the their RGR counterpart (pval > 0.01) and were more sensitive than the iRGR having an average SR of

2.2. Unlike RGRs lacking a 3' Rbz all RGiR devices result in a measurable change in system output. Both iRGR and RGiR devices produce gRNA with GC rich scars but the iRGR scar is significantly longer (58 compared to 13 base pairs, on average) resulting in gRNA with more stable secondary structure. We found that with increased post-cleavage gRNA structure there is an increase in the average SR (Spearman-rank correlation -0.58, pval=0.02) (Figure C.4). Implying that the activity of highly structured gRNA inhibits device activity.

To summarize, residual scar sequences and structures associated with a gRNA after cis-Rbz cleavage can have significant impacts on expressed gRNA activity. Nonetheless, we observe that functional RGRs can be assembled from sequence and structurally diverse hammerhead Rbzs. We find that 5' Rbz cleavage is not sufficient for highly efficient gRNA expression. Finally, we show that the 3' HDV Rbz can be readily replaced by a shorter hammerhead Rbz (68 compared to 52 bp) with little to no reduction in gRNA activity.

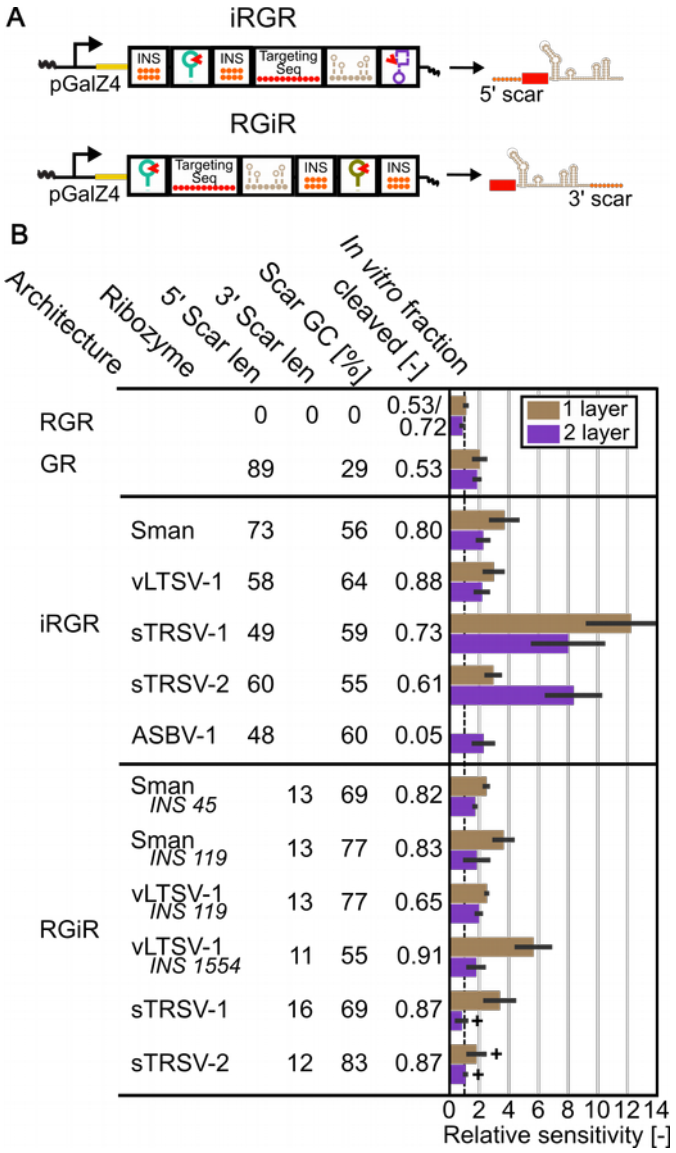


Figure 4.3: Expressing gRNA with 5' ribozyme and 3' ribozyme insulated iRGR and RGiR.

(A) Architecture of the iRGR and RGiR, where 5' and 3' ribozymes are flanked with short insulating sequences designed to enable proper ribozyme folding. Both architectures result in vestigial ribozyme scar made up of a section of the ribozyme and the insulating sequence (B) Basic scar metrics and sensitivity ratios for each RGiR and iRGR device. For each device a sensitivity ratio ratio is calculated relative to the RGR in their respective layer. Error bars indicate ±STD. + indicates that the device population and RGR control are indistinguishable (pval>0.01).

4.4.3 Aptazyme-regulated gRNA expression with RGAs

Our results showed that eliminating 3' Rbz cleavage suppressed functional gRNA expression from an RGR and that functional variations of the canonical RGR can be made by substituting the 3' HDV Rbz with hammerhead Rbz. Drawing on these insights, we hypothesized that ligand-responsive control of gRNA expression could be achieved by substituting the 3' RGR Rbz with an aptamer-regulated Rbz (aptazyme). The resulting Rbz-gRNA-Aptazyme (RGA) architecture would then provide a mechanism for engineering aptazyme-regulated gRNA expression through ligand-dependent aptazyme cleavage (Figure 4.4A).

As a proof-of-concept, we assembled RGAs using the theophylline aptazymes Theo(A)CAAGUGAA and Theo(A)AAAA (Table C.1). Both theophylline aptazymes were previously evolved in yeast to catalytically cleave themselves at rates inversely proportional to ligand concentration[13]. Consequently, adding theophylline reduces the amount of aptazyme cleavage which in-turn should reduce functional gRNA expression. RGAs were tested in an inducible one-layer repression cascade expressing gRNAW8 (Figure 4.4A). Like the RGiR devices, RGA devices produce gRNA with a 3' scar and were observed to be less effective than their RGR counterpart, with an average SR of 2.1 (Figure 4.4B; Figure C.3). To characterize the relative decrease in functional gRNA expression with increased theophylline, both RGAs were characterized at three levels of induction and six theophylline concentrations. Increasing concentrations of theophylline caused up to a 9.6 fold decrease in RGA mediated repression, relative to a no theophylline control (Figure 4.4C and G). Thus, we validate that the RGR architecture can be extended to conditionally express functional gRNA by replacing the 3' Rbz with an aptazyme.

We reasoned that if aptazyme inhibition resulted in a proportional decrease in functional gRNA expressed from the RGA we would then observe a consistent decrease in the sensitivity of the ligand induced systems. To test this we calculated relative sensitivity values for each theophylline concentration relative to the no theophylline response. ANOVA analysis of each concentration of theophylline showed that, regardless of RGA induction, there was an

indistinguishable increase in relative sensitivity (Figure 4.4E and I). Because of this consistent behavior, we were able to model the response of RGA one-layer cascades as a function of both RGA expression level and theophylline concentration. This was done by multiplying the no-theophylline Hill-equation parameter ka by the calculated SR for each theophylline concentration (Figure 4.4D and H; Table C.3). This model of RGA behavior accurately described the experimentally observed data (r2=0.99 for both RGA devices) (Figure 4.4F and J). To summarize, we created two theophylline inducible RGA devices and showed that with increasing ligand concentration there is a predictable decrease functional gRNA expression. This RGA architecture provides a novel architecture for implementing small molecule control over gRNA expression in Eukaryotes.

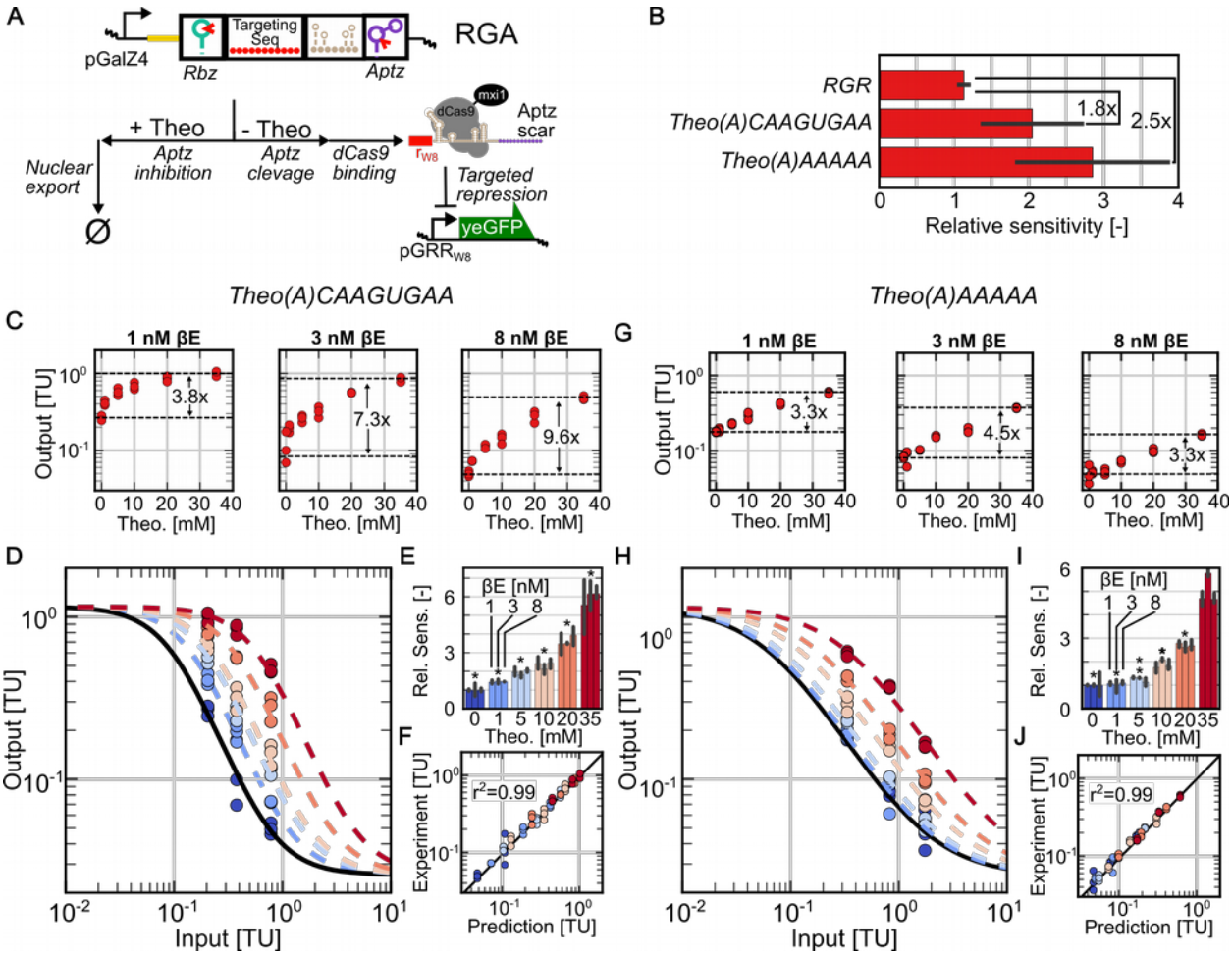


Figure 4.4: gRNA expression can be dynamically regulated with the inclusion of an aptazyme.

(A) Architecture of the ribozyme-gRNA-aptazyme (RGA). In the absence of theophylline, the 3' theophylline aptazyme cleaves, resulting in the functional expression of gRNA. In the presence of theophylline, aptazyme cleavage is inhibited, resulting in gRNA export from the nucleus. (B) Ratio of performance take relative to an RGR. Error bars indicate ±STD, * indicates a pvalue < 0.001 calculated using two tailed Welch's t-test. Panels C, D, E, and F represent RGA Theo(A)CAAGUGAA and panels G, H, I and J represent RGA Theo(A)AAAA (C) and (G) Increasing theophylline inhibits RGA mediated gRNA repression at multiple levels of transcription (increasing bE induction) for and (D) and (H) Theophylline responsiveness of RGAs in an inducible one-layer context. The characteristic Hill function in black is derived from RGA with no theophylline. Each dashed line is the characteristic transfer function with its ka parameter modified to reflect the experimentally observed decrease in sensitivity induced by theophylline. (E) and (I) Sensitivity ratio of data presented in panel D at different levels of induction (bE) and theophylline taken relative to the characteristic transfer function. Asterisks indicates a one-way ANOVA pvalue > 0.05. (F) and (J) Comparison of the between experimentally observed output to the prediction based on the observed increase in sensitivity ratio.

4.4.4 Decreased expression of dCas9 exacerbates gRNA competition

We next study the effects gRNA competition for dCas9 has on RGR-input performance. Instead of directly testing this phenomena by with systems which express many gRNA, gRNA competition is simulated through decreasing the expression of dCas9 in one-, two-, and three-layer repression cascades (Figure 4.2B and Figure 4.5A). More specifically, systems with RGR and GR inputs for both gRNAW10 and gRNAW8 that express either high or low levels of dCas9 were compared (20 total systems). dCas9 transcription was driven by either the strong pGPD promoter (promoter used for all previous systems) or the ~20x less transcriptionally active pGRRnull (Figure 4.5B). In all repression cascades we expected that decreased dCas9 expression would require a proportional increase in gRNA expression to achieve the same level of action, causing the systems to become relatively less sensitive[158]. The effects of decreased dCas9 expression were quantified by computing the relative sensitivity of system responses (SFigure C.5) to their respective pGPD dCas9 counterpart system response, which were then used to calculate SRs (Figure 4.5C).

Decreasing dCas9 expression significantly affected the sensitivities of six out of ten systems. In one-layer repression cascades, both RGRW10 and RGRW8 showed decreased sensitivity with decreased dCas9 expression (SRs of 1.2 and 1.3 respectively). Interestingly, in these systems both of the GR inputs showed similar sensitivities. We speculated that in a one-layer context the decreased performance of GR outweigh the consequences of decreased dCas9 expression, making them appear unaffected. In two-layer repression cascades, only systems with RGRW8 and GRW8 inputs were affected by decreased dCas9 expression, resulting in increased sensitivity (SR of 0.53 and 0.29, respectively). This increase in system sensitivity implies that the input gRNAW8 is able to out compete the internal gRNAW10 for dCas9. If true, when dCas9 expression is decreased gRNAW8 would segregate the majority of available dCas9 from gRNAW10. When the network was inverted, with gRNAW10 as the input, system sensitivity was unaffected by decreased dCas9 expression which is consistent with the idea that gRNAW8 segregates dCas9. We next extended the gRNAW8 input two-layer cascades by appending RGRW5 after the internal RGRW10 layer to create a three-layer cascade (Figure 4.5A). In both

RGRW8 and RGW8 input systems sensitivities decreased with decreased dCas9 (SR of 1.5 and 1.4 respectively), indicating that any competitive advantage gRNAW8 had when repressing gRNAW10 is diminished relative to the new competition it experienced with the internal gRNAW5. Importantly, whereas the input-GRW8 was unaffected by decreased dCas9 expression in a one-layer system, it is affected in the three-layer cascade. This demonstrates that effects of dCas9 limitations, and potentially gRNA competition, are exacerbated in larger systems. Overall, we found that systems with decreased dCas9 expression experienced significant and, in some cases, unintuitive shifts in their sensitivities to RGR-inputs. These effects are likely caused by gRNA competition and are consistent with the ideas that the behaviour of multi-gRNA systems, especially under dCas9 limitations, may be difficult to predict from individually characterized behaviors.

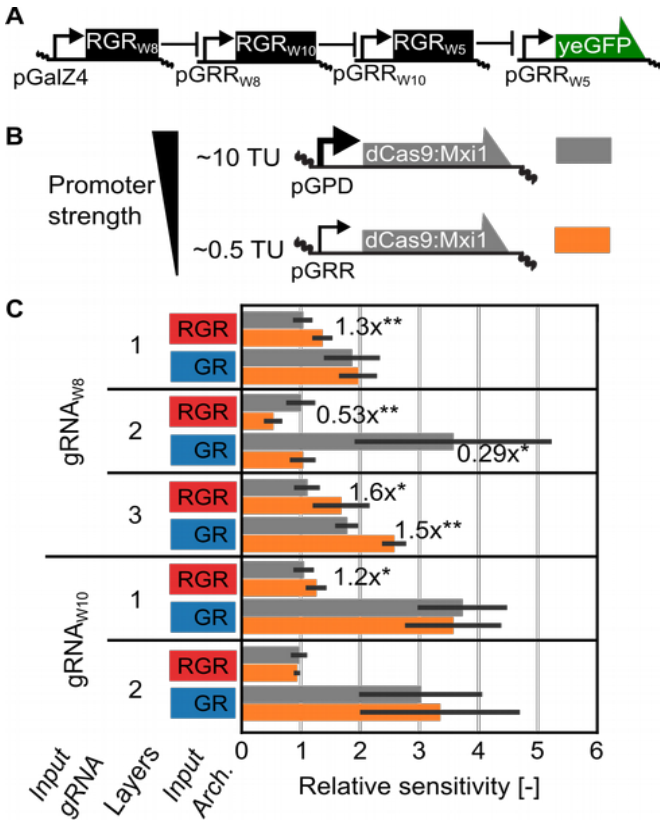


Figure 4.5: Variations in dCas9 expression both increase and decrease input sensitivity depending on context.

(A) The three-layer system that used in this study. (B) dCas9:Mxi1 is driven from one of two promoters, pGPD is ~20x as strong as the pGRR promoter. Relative strengths were experimentally determined with a fluorescent assay where each promoter expresses eYFP in yeast cells (C) Sensitivity ratio is presented for each gRNA input architecture in each layer taken relative to their respective pGPD, RGR strain. * indicates pval<0.005, ** indicates pval<0.0001 calculated using two tailed Welch's t-test, and error bars indicate ±STD.

4.4.5 Predicting the functions of multi-gRNA systems

As seen in the previous section, limited dCas9 expression caused significant perturbations to system responses. It is not known the extent to which these effects limit the ability to predict system responses from RGR promoter interactions that were characterized in isolation (one-layer cascades). Here, we tested these effects by comparing the experimental system response to a system response predicted from characterized interactions. As others have shown, it is possible

to predict the behavior of systems composed of interconnected parts by propagating the modeled responses of all parts through one another[16]. This approach was used to predict the full input-output dynamic range of two-layer cascades and to predict the average output performance of a five component Boolean XOR gate. First, we use characterized one-layer cascade responses to predict two-layer systems expressing high and low levels of dCas9 (Figure 4.6A). The quality of these predictions were quantified by taking the difference in explained variance between the prediction and a direct fit of the experimental data.

Experimental behavior was accurately described by predictions for four of the eight tested systems (<5% decrease in explained variance) (Figure 4.6B). On average, predictions were more accurate for systems with high dCas9 expression than low dCas9 expression (4.5% compared to 27.8% average decrease in explained variance). This further shows that systems with limited dCas9 can result in unpredictable system behavior. Additionally, predictions were less accurate for systems with gRNAW8 as their input than gRNAW10 as their input (27.8% vs. 3.6% average decrease in explained variance). gRNAW8 inputs were predicted to be less sensitive than what was experimentally observed, supporting our previous hypothesis that gRNAW8 is able to out compete the internal gRNAW10 for dCas9. Interestingly, this competitive advantage was seen in high dCas9 systems for RGRW8 but not GRW8 (11% vs. 0% decrease in explained variance). This promoted the idea that even in systems with high dCas9, effective gRNA expression, if relatively competitive with other gRNA in the system, can result in unpredictable system behavior. Lastly, whereas three of the four predictions that had large deviations from experimental behavior were caused by shifts in sensitivity, at low dCas9 the predictions of input GRW10 deviates only by overestimating the vmin and vmax response of the system (10% decrease in explained variance). This demonstrated that effects of gRNA competition can solely influence the accessible dynamic output-range of a system. Together, this suggests that gRNA competition affects system behavior and that it is exacerbated when dCas9 is limited. Next, we apply these techniques to predict the expected behavior of a multi-component Boolean XOR gate.

Previously these CRISPRi components were used to create digital logic gates that produce a binary, ON or OFF, system response in presence or absence of input gRNA (Gander 2017). This framework abstracted two-input pGRR promoters driving the expression of an RGR to be a 2-input, 1-output NOR gate, that is only ON if neither input is present (Figure 4.6C). Five NOR gates were connected by re-coding the target sequence and target sites of the NOR gates, to create a 2-input 1-output Boolean XOR gate (Figure 4.6D). For a 2-input logic gate to be considered functional it must produced the correct output response for each of the four possible input-states. Fifteen distinct versions of the XOR gate were built and only two were functional. Predictive models of CRISPRi transcriptional networks that accurately estimate success rates and identify components that are critical for success of large systems, like the XOR gate, would inform their construction and maximize successful designs. Here, we attempted to predict the average success and performance of XOR gates if they were built from random NOR gates. If the aggregate behavior of the simulations agreed with our previously characterized experimental outcomes, we would expect this approach to be useful to inform future designs. Aggregate predictions of XOR gate function were generated by simulating 20,000 randomly created XOR gates for each of the four input-states. For each random XOR gate, all component NOR gates were assigned random parameters that were bound by the Hill-equation parameters derived from RGR inputs in high expressing dCas9 one-layer cascades (Table C.4).

The aggregate modeled XOR gate performance fairly predicted both the observed input-output levels and identified the high failure-rate input-states of the experimental XOR gates. For each experimental and predicted XOR gate input state we aggregated the raw output response (Figure 6E) and recorded if it fell within it's target ON or OFF response range (Gander 2017) (Figure 6F). On average, input-state (0,0) was predicted have higher system output than the experimental observations. The output of input-state (1,1) was predicted to be significantly less than experimental observations which ultimately caused the success rate of of input-state (1,1) overestimated (68% predicted compared to 27% experimental success). This prediction error was the main cause of overestimating overall XOR gate success (53% predicted compared to 13% experimental success). To investigate if XOR gate success was sensitive to perturbations of

parameters of these NOR gates, simulated XOR gates were partitioned between successful and unsuccessful and the resulting distributions of parameters were compared (Figure C.6A). Parameters distributions of ka, vmax and n for NOR gates 4 and 5 were the only that appeared to significantly deviate from each other. These parameters also had the strongest Spearman rank correlation coefficient when computed relative to XOR gate success (Figure C.6B). This indicates the that behavior of NOR gates 4 and 5 are essential for successful XOR gates. As these 5 gRNA XOR gates likely experience some shift in system sensitivity through gRNA competition which may break otherwise successful designs, we next globally scaled the sensitivity parameter, ka, to see its effect on generating successful XOR gates.

The characterized sensitivity of our RGR-inputs from one-layer devices was found to be nearly optimal for the construction of successful XOR gates, with significant increases or decreases in sensitivity resulting dominantly failed systems. For this analysis all ka values for 1,000 randomly generated XOR gates were scaled by ten factors and the fraction of functional XOR gates were calculated (Figure 6G). This showed that if the effective sensitivity of all components in this system were decreased (larger ka) by a factor equal to our worst performing iRGR devices (>8x) or if it were possible to globally increase sensitivity (<0.5x) the majority of all XOR gates would fail. It is clear from this analysis systematic shifts in sensitivity, which can be caused by gRNA competition, will have significant effects on large network performance. Thus, we demonstrated that computational models can be used to predict CRISPRi transcriptional network dynamics, that these predictions become worse with increased gRNA competition, and that modeling can be used to inform large network design.

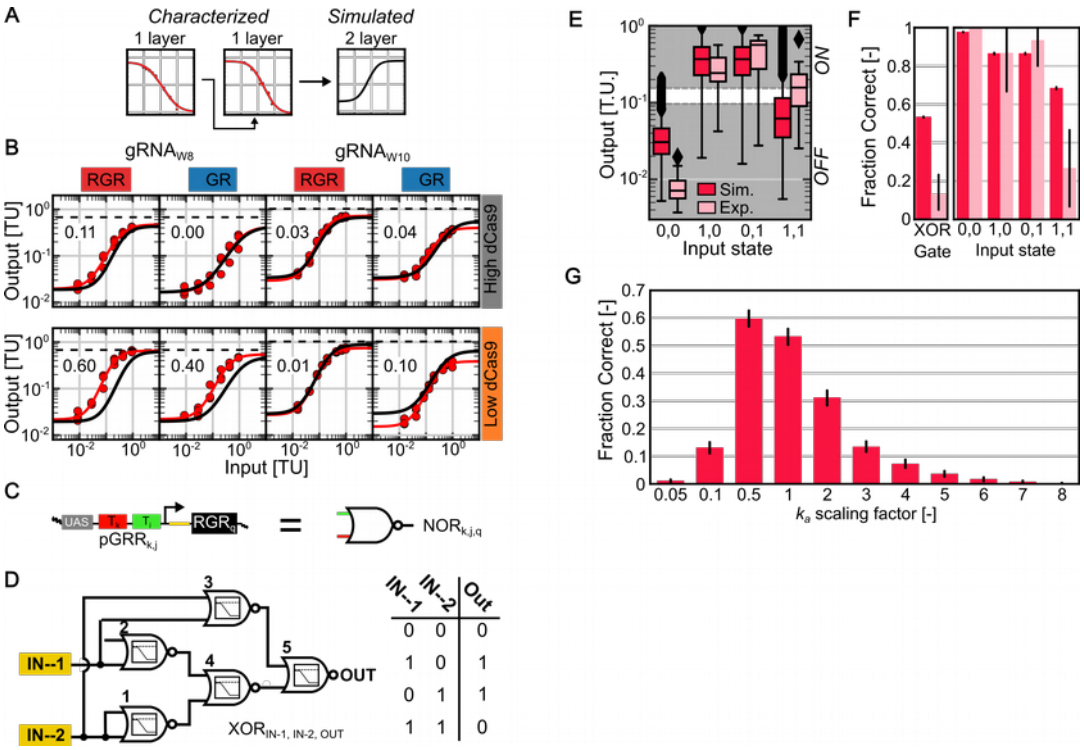


Figure 4.6: Predicting the performance of large CRISPRi networks from individually characterized gRNA expression components.

(A) Two-layer system performance is predicted by propagating the respective component transfer Hill functions, determined through experimental characterization, through one another. (B) Comparison of the simulated two-layer systems (black) to the experimentally characterized two-layer systems (red). The architectures are detailed in Figure 2B. gRNA W8 and W10 are expressed with RGR and GR as inputs in high (pGPD expressed, grey) and low (pGRR expressed, orange) dCas9 systems. Numbers inside are the difference in log-r2 of the red fit and black fit of the red experimental data. (C) A NORk,j,q gate is defined by a pGRRk,j input-input stage and a RGRq output-stage. Scaling. (D) The experimentally designed architecture for XOR logic gate is composed of five interconnected NOR gates. The truth table indicates how high input (1) and low input (0) for each of the inputs should affect the output level. (E) The experimentally measured output for experimentally constructed XOR gates (n=15) alongside the output of simulated XOR gates (n=20,000), the box represent the interquartile range and whiskers are 1.5x the interquartile range less or more Q1 and Q3 respectively. Output values considered ON on and OFF in the upper and lower grey shaded region, respectively. (F) The fraction of successful XOR gates and individual input states for both experimental (n=15) and simulated (n=20,000) gates. Input-states from (E) were considered correct if their associated output fell within their target ON/OFF response. A XOR gate was considered to be overall correct if all input-states were correct. Error bars are the 95% confidence interval estimated from 1000 bootstrapping iterations. (G) The fraction of simulated XOR gates that were successful with global scaling of ka for each NOR gate. Each bar represents simulations where ka values were scaled by the indicated scaling factor (n=1,000). Error bars are the 95% confidence interval estimated from 1000 bootstrapping iterations.

4.5 Discussion

In this study, we sought to better understand the engineerability of the RGR architecture, to understand how CRISPRi system behavior is influenced by gRNA competition and to predict the behavior of large systems. To accomplish this, we developed a quantitative framework to characterize and model the input-output responses of 44 unique CRISPRi transcriptional networks in yeast. This involved normalizing inducer-inputs and fluorescent-outputs to a common transcriptional unit based on normalized fluorescence. This assumes that steady-state protein fluorescence levels are proportional to the rate of transcription. Though this oversimplifies the complex relationship between mRNA abundance and protein expression[166,167], our ability to precisely predict system performance (Figure 4.6B) makes us believe that this assumption is valid. Normalized data was fit with with a four parameter Hill-equations[16] that enabled the easy comparisons system responses. This simplified method was used over lower level kinetic models of system performance[14,42] that can describe kinetic mechanisms, as these models have many degrees of freedom, require parameter estimation and were not needed for our quantitative comparisons. Our analysis based on the relative sensitivities between input-RGRs allowed us to quantify how much more or less input-expression is required for them to function, a method well suited for systems which have similar responsiveness. We believe that in the case of systems with different responsiveness, statistical comparisons were sufficient to delineate possible differences. Ultimately our developed framework and method of analysis allowed us to critically analyze and compare dynamic system responses.

Beyond the creation of CRISPRi transcriptional networks Pol II based methods of gRNA expression have widespread applications for tissue specific transcription[156,168], programmed levels of transcription[164] and inducible transcription[169]. Our initial investigations of the popular RGR expression platform found that 3’ Rbz processing is essential and 5’ Rbz processing results in improved performance. However, as each RGR contains 87% redundant sequence (185/211 bases, 123 bases from Rbzs) multiple uses of it risks host-mediated recombination when building large systems[138]. To enhance sequence diversity and to study the engineerability of the

architecture we substituted the canonical Rbzs for the computationally insulated versions (iRGR, RGiR). iRGR devices decouple the necessity sequence upstream of the 5' Rbz's catalytic core for each target sequence[123] making it more suitable for gRNA library expressions[170–172]. Of the eleven novel device, two of the RGiR performed as well as their RGR counterpart. The decrease in performance for both the iRGR and RGiR was found to correlate structure resulting from Rbz cleavage scars (Figure C.4) a trend that has been previously observed[146173159]. Therefore it conceivable that ribozymes which minimize their cleavage scars or through design of the scars such that they do not favorably interact with gRNA handle could be used to create better performing iRGR or RGiR devices. These insights on RGR mediated gRNA expression will inform the creation of compact gRNA expression systems[156] and multi-gRNA expression from a single transcript[174].

In understanding that 3' Rbz cleavage is essential to RGR function we correctly reasoned that we could create a small molecule inducible variation that employs a 3' ligand sensitive aptazyme (RGA). We successfully used two theophylline aptazymes (Theo(A)-CAAGUGAA and Theo(A)-AAAAA) that were originally selected and characterized in the context of the 3' UTR of coding mRNA[13]. Though, Theo(A)-CAAGUGAA was characterized to have a smaller dynamic range than Theo(A)-AAAAA in its native context, its cleavage was characterized to be more inhibited by theophylline. This ultimately resulted in a dynamic range that was ~2x as large as Theo(A)-AAAAA in the context of the RGA. Both of these aptazymes were not fully inhibited in their native context, implying that aptazymes with lower rates of background cleavage could achieve upwards of 45 fold decreases in gRNA expression that were observed for no 3' Rbz control. Elsewhere, ligand-dependent control of gRNA function has been achieved through mechanisms employing either aptamers[157], or aptazymes[78], that regulate gRNA activity through conditionally forming structure with the targeting sequence, thought to antagonize gRNA interactions with dCas9 or dCas9:gRNA interactions with DNA. Though these achieve 5-10 fold changes of gRNA mediated activity, these mechanisms require redesigning sections of the aptamer sequence for each target sequence, limiting their applicability for library based CRISPR studies[170–172,175]. That being said, the RGA architecture is compatible with aptamers

which could be used to make composite small molecule sensing or enhanced response to a single molecule. Additional, it should be possible to construct RGAs with other aptazymes[13] and should function in any organism where RGRs have been used successfully used[42,108,154,155].

This work has shown that gRNA competition for dCas9 can cause CRISPRi transcriptional networks to behave differently than what would be expected based on components characterized in isolation. These system effects, where the presence of components reactively interact with eachother in ways that were not explicitly programmed, are commonly referred to as retroactive effects[54–56,176]. These retroactive effects agree with what was seen when multiple gRNA were used to affect gene expression in parallel[158] and were expected due to the observations that gRNA rates of association with Cas9 deceased up to 30-fold due to non-specific RNA inhibition[142] and measured affinity for gRNA has been shown to decrease three-fold in the presence of other gRNA[159]. These effects caused the sensitivities of multi-gRNA systems to require, on the extremes, 1.5 and 0.29 fold the input expression to achieve the same level of system response and caused explanatory predictions of system performance to decrease by as much 60%. Importantly these effects were mostly mitigated in systems with increased levels of dCas9 expression. This study was able to quantify these effects, where others did not[40–42], because of the complete characterization components in isolation and in the context of larger networks. The effects of gRNA competition could be circumvented through the careful selection of gRNA that have similar affinity for Cas9, however this would require quantification of gRNA libraries for affinity[159] or the development or predictive models. As others have suggested these effects may also be mitigated by including orthogonal gRNA and Cas9 variants[177]. On the other hand, kinetic models which accommodate gRNA affinity for dCas9 could be used to predict, and potentially exploit, gRNA competition effects to build more robust networks. Regardless, the effects of gRNA competition for dCas9 can have substantial impacts on any multi-gRNA systems and more work should be done to understand it.

Though dCas9 limitations negatively affected our ability to predict system performance, systems with excess dCas9 were accurately modeled. We were even able to fairly approximate the aggregate performance of an experimental XOR gate[42], assuming that it was built from

random components that behaved similarly to the devices we characterized in a one-layer context. Predictions ultimately overestimated the success of the experimental XOR gates. Analysis of component parameters identified NOR gates 4 and 5 as having the most significant impact on XOR gate success. These components happened to be consistent for each of the fifteen experimental XOR gates that were considered. If these components were non-optimal, experimental systems would have been biased to fail, potentially explaining our predictions deviation. Investigation of global changes to sensitivity of the system, a possible consequence of gRNA competition, found that the majority of systems failed with increased or decreased sensitivity. Changing sensitivity caused the dynamic input-range to shift, which is governed by ka and n of the Hill-equations, to be incompatible with the dynamic output-range of the NOR gates causing a compression of the dynamic range of aggregate systems[16]. This analysis showed that the native sensitivity of our RGR-inputs is nearly perfectly tuned to minimize input-output mismatch caused failures of systems. Taken together we have demonstrated that modeling of CRISPRi transcriptional systems can inform system designs in ways that will reduce rates of failure.

In sum, we have identified RGR design elements needed for efficient gRNA activity, created a new RGA architecture for aptazyme-regulated gRNA expression, and shown that gRNA-directed transcriptional network functions can be predicted. Taken together, our work has broad implications for engineering multi-gRNA transcriptional networks and addresses key barriers to engineering even larger CRISPR-Cas information processing systems." By addressing key barriers to generating reliable gRNA expression and function and accurately predicting the activities of multi-gRNA systems, this work sets the stage for engineering even larger CRISPR-Cas transcriptional networks.

Chapter 5. Conclusions

This thesis has covered significant progress made in understanding gRNA expression and CRISPRi transcriptional networks. Chapter 2 provides an in depth look at computational RNA folding and design. Here, the computational package pyrfold for python is presented. This allows users to easily designed structured RNA using specified and unspecified nucleotides, simulate the folding of those components using minimum free energy and kinetic co-transcriptional folding algorithms and then process the resulting data to make meaningful conclusions. This work was leveraged in all subsequent chapters.

Chapter 3 introduces CRISPRi transcriptional networks in yeast and then uses them to construct Boolean logic genetic programs. This involved the creation a modular 2-input and 1-output NOR gate which uses gRNA as both its input and output. This NOR gate was used to create five additional Boolean classifiers and layered seven times to create the longest repression cascade to date. Computational analysis of simulated system performance identified that transcriptional leak, a consequence of incomplete repression, was the critical design feature. This work highlighted the versatility of CRISPRi transcriptional control platform in yeast and demonstrated that large systems could be built using minimally characterized components.

Chapter 4 used the previously introduced CRISPRi platform experimentally and analytically study the dynamics of the transcriptional networks in yeast. This work saw the creation of eleven novel gRNA expression architectures from Pol II promoters and the creation of a novel aptazyme regulated gRNA expression platform, the RGA. This ligand sensitive architecture enables complex biosensing functionality to be baked into CRISPRi transcriptional networks at the level of gRNA. It was reported that gRNA competition for a dCas9 can result in significant perturbations to network behavior and that excess dCas9 mitigates these effects. Finally, predictive computational models accurately predicted experiential system behavior and identified possible failure mechanisms.

Beyond what was presented in the main Chapters of thesis there are a collection of preliminary data which highlight potential future research topics (APPENDIX D). Data shows that it is possible to change the response dynamics of gRNA in the context of a system by including many copies of 'decoy' target sites. It is possible that this could be used debug or optimize transcriptional networks after they are constructed. The models generated for ligand inducible RGA (Chapter 4) can be used to select constitutive promoters to enable to RGA to act an inducible system input. This platform has applications for any inducible gRNA expression system. Finally, it may be possible to tune the repression dynamics of pGRR promoters (Chapter 3 and 4) by including additional target sites. A more diverse collection of response dynamics could expand the robustness of large CRISPRi networks.

In total, this work has created, studied, and refined a versatile architecture for the creation of gene regulatory networks in yeast. The insights gained from this have significant implications for all any system which is expressing multiple gRNA and any system which expresses gRNA from a Pol II promoter.

REFERENCES

1. Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: From microbes to mammals. *Cell* **144,** 910–925 (2011).

2. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316,** 1497–502 (2007).

3. Ravasi, T. *et al.* An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. *Cell* **140,** 744–752 (2010).

4. Lee, E., Tabor, J. & Mikos, A. Leveraging synthetic biology for tissue engineering applications. *Japanese Soc. Inflamm. Regen.* **34,** 15–22 (2014).

5. Slomovic, S., Pardee, K. & Collins, J. J. Synthetic biology devices for in vitro and in vivo diagnostics. *Proc. Natl. Acad. Sci.* **112,** 14429–14435 (2015).

6. Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **14,** 135–149 (2016).

7. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).

8. Gaber, R. *et al.* Designable DNA-binding domains enable construction of logic circuits in mammalian cells. *Nat. Chem. Biol.* **10,** 203–208 (2014).

9. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science (80-.).* **353,** aad8559 (2016).

10. Chappell, J., Takahashi, M. K. & Lucks, J. B. Creating small transcription activating RNAs. *Nat. Chem. Biol.* **11,** 214–220 (2015).

11. Rodrigo, G., Landrain, T. E. & Jaramillo, A. De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 15271–6 (2012).

12. Hallberg, Z. F., Su, Y., Kitto, R. Z. & Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem.* **86,** annurev-biochem-060815-014628 (2017).

13. Townshend, B., Kennedy, A. B., Xiang, J. S. & Smolke, C. D. High-throughput cellular RNA device engineering. *Nat. Methods* 1–10 (2015). doi:10.1038/nmeth.3486

14. Wang, Y. H., McKeague, M., Hsu, T. M. & Smolke, C. D. Design and Construction of Generalizable RNA-Protein Hybrid Controllers by Level-Matched Genetic Signal Amplification. *Cell Syst.* **3,** 549–562.e7 (2016).

15. Carothers, J. M., Goler, J. a, Juminaga, D. & Keasling, J. D. Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression. *Science (80-.).* **334,** 1716–1719 (2011).

16. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352,** aac7341 (2016).

17. Bradley, R. W., Buck, M. & Wang, B. Recognizing and engineering digital-like logic gates and switches in gene regulatory networks. *Curr. Opin. Microbiol.* **33,** 74–82 (2016).

18. Daniel, R., Rubens, J. R., Sarpeshkar, R. & Lu, T. K. Synthetic analog computation in living cells. *Nature* **advance on,** 619–623 (2013).

19. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **7,** 856–866 (2012).

20. Kittleson, J. T., Wu, G. C. & Anderson, J. C. Successes and failures in modular genetic engineering. *Curr. Opin. Chem. Biol.* **16,** 329–336 (2012).

21. Cech, T. R. & Steitz, J. A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **157,** 77–94 (2014).

22. Mortimer, S. a, Kidwell, M. A. & Doudna, J. a. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15,** 469–79 (2014).

23. Batey, R., Rambo, R. & Doudna, J. Tertiary Motifs in RNA Structure and Folding. *Angew. Chem. Int. Ed. Engl.* **38,** 2326–2343 (1999).

24. Lutz, B., Faber, M., Verma, A., Klumpp, S. & Schug, A. Differences between cotranscriptional and free riboswitch folding. *Nucleic Acids Res.* 1–10 (2013). doi:10.1093/nar/gkt1213

25. Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E. & Altuvia, S. A pH-responsive riboregulator. *Genes Dev.* **23,** 2650–62 (2009).

26. Roßmanith, J. & Narberhaus, F. Exploring the modular nature of riboswitches and RNA thermometers. *Nucleic Acids Res.* gkw232 (2016). doi:10.1093/nar/gkw232

27. Lander, E. S. The Heroes of CRISPR. *Cell* **164,** 18–28 (2016).

28. Chen, J. S. & Doudna, J. A. The chemistry of Cas9 and its CRISPR colleagues. *Nat. Rev. Chem.* **1,** 78 (2017).

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

30. Komor, A. C. *et al.* CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168,** 20–36 (2017).

31. Greene, A. C. CRISPR-Based Antibacterials: Transforming Bacterial Defense into Offense. *Trends Biotechnol.* **36,** 127–130 (2018).

32. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509,** 487–491 (2014).

33. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (80-.).* **343,** 84–87 (2014).

34. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).

35. Gilbert, L. a. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–451 (2013).

36. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10,** 973–6 (2013).

37. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33,** 510–517 (2015).

38. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155,** 1479–1491 (2013).

39. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32,** 569–76 (2014).

40. Nielsen, A. A. K. & Voigt, C. A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Mol. Syst. Biol.* **10,** 763 (2014).

41. Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11,** 723–726 (2014).

42. Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M. & Klavins, E. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat. Commun.* **8,** 15459 (2017).

43. Ma, K. C., Perli, S. D. & Lu, T. K. Foundations and Emerging Paradigms for Computing in Living Cells. *J. Mol. Biol.* (2016). doi:10.1016/j.jmb.2016.02.018

44. Brophy, J. a N. & Voigt, C. a. Principles of genetic circuit design. *Nat. Methods* **11,** 508–20 (2014).

45. L, R. & Y, R. Noise Tolerance Analysis for Reliable Analog and Digital Computation in Living Cells. *J. Bioeng. Biomed. Sci.* **6,** (2016).

46. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30,** 354–359 (2012).

47. Weinberg, B. H. *et al.* Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35,** 453–462 (2017).

48. Zong, Y. *et al.* Insulated transcriptional elements enable precise design of genetic circuits. *Nat. Commun.* **8,** 1–12 (2017).

49. Mishra, D., Rivera, P. M., Lin, A., Del Vecchio, D. & Weiss, R. A load driver device for engineering modularity in biological networks. *Nat. Biotechnol.* (2014). doi:10.1038/nbt.3044

50. Rubens, J. R., Selvaggio, G. & Lu, T. K. Synthetic mixed-signal computation in living cells. *Nat. Commun.* **7,** 1–10 (2016).

51. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science (80-.).* **324,** 1199–1202 (2009).

52. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems. *Biotechnol. J.* **7,** 856–66 (2012).

53. Ang, J., Harris, E., Hussey, B. J., Kil, R. & McMillen, D. R. Tuning response curves for synthetic biology. *ACS Synth. Biol.* **2,** 547–67 (2013).

54. Qian, Y., Huang, H.-H., Jiménez, J. I. & Del Vecchio, D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synth. Biol.* acssynbio.6b00361 (2017). doi:10.1021/acssynbio.6b00361

55. Del Vecchio, D., Ninfa, A. J. & Sontag, E. D. Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.* **4,** 161 (2008).

56. Jayanthi, S., Nilgiriwala, K. & Vecchio, D. Del. Retroactivity Controls the Temporal Dynamics of Gene Transcription. *ACS Synth. Biol.* 1–28 (2013).

57. Sleight, S. C. & Sauro, H. M. Visualization of evolutionary stability dynamics and competitive fitness of Escherichia coli engineered with randomized multigene circuits. *ACS Synth. Biol.* **2,** 519–528 (2013).

58. Elowitz, M. & Lim, W. Build life to understand it. *Nature* **468,** 889–890 (2010).

59. Wilson, D. & Szostak, J. In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.* 611–647 (1999).

60. Salis, H. M., Mirsky, E. a & Voigt, C. a. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–50 (2009).

61. Huang, F. W. D. & Reidys, C. M. Shapes of topological RNA structures. *Math. Biosci.* **270,** 57–65 (2015).

62. Smolke, C. D. & Keasling, J. D. Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. *Biotechnol. Bioeng.* **80,** 762–76 (2002).

63. Isambert, H. & Siggia, E. D. Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6515–20 (2000).

64. Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T. & Lucks, J. B. Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.* 1–10 (2016). doi:10.1038/nsmb.3316

65. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31,** 147–157 (1982).

66. De la Peña, M., Gago, S. & Flores, R. Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO J.* **22,** 5561–5570 (2003).

67. Riccitelli, N. J. & Lupták, A. Computational discovery of folded RNA domains in genomes and in vitro selected libraries. *Methods* **52,** 133–40 (2010).

68. Lai, D. & Meyer, I. M. e-RNA: a collection of web servers for comparative RNA structure prediction and visualisation. *Nucleic Acids Res.* 1–4 (2014). doi:10.1093/nar/gku292

69. Zhu, J. Y. a, Steif, A., Proctor, J. R. & Meyer, I. M. Transient RNA structure features are evolutionarily conserved and can be computationally predicted. *Nucleic Acids Res.* **41,** 6273–85 (2013).

70. Wuchty, S., Fontana, W., Hofacker, I. L. & Schuster, P. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49,** 145–65 (1999).

71. Zuker, M., Mathews, D. & Turner, D. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. *RNA Biochem. Biotechnol.* (1999).

72. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6,** 26 (2011).

73. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32,** 170–173 (2011).

74. Isambert, H. The jerky and knotty dynamics of RNA. *Methods* **49,** 189–96 (2009).

75. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* **41,** 2541–2551 (2013).

76. Ogawa, A. Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. *RNA* **17,** 478–88 (2011).

77. Ferry, Q. R. V., Lyutova, R. & Fulga, T. A. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat. Commun.* **8,** 14633 (2017).

78. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8,** 1–8 (2017).

79. Thimmaiah, T., Voje, W. & Carothers, J. M. Computational Design of RNA Parts, Devices, and Transcripts with Kinetic Folding Algorithms Implemented on Multiprocessor Clusters. *Methods Mol. Biol.* **1244,** 137–65 (2015).

80. Taneda, A. Multi-objective genetic algorithm for pseudoknotted RNA sequence design. *Front. Genet.* **3,** 36 (2012).

81. Xayaphoummine, A., Bucher, T. & Isambert, H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* **33,** W605-10 (2005).

82. Bernot, G., Comet, J. P., Richard, A. & Guespin, J. Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *J. Theor. Biol.* **229,** 339–347 (2004).

83. Hinkelmann, F. & Laubenbacher, R. Boolean models of bistable biological systems. *Discret. Contin. Dyn. Syst. - Ser. S* **4,** 1443–1456 (2011).

84. Yuh, C. H., Bolouri, H. & Davidson, E. H. Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control. *Development* **128,** 617–629 (2001).

85. Hobert, O. Common logic of transcription factor and microRNA action. *Trends Biochem. Sci.* **29,** 462–468 (2004).

86. Shmulevich, I., Dougherty, E. R. & Zhang, W. From boolean to probabilistic boolean networks as models of genetic regulatory networks. *Proc. IEEE* **90,** 1778–1792 (2002).

87. Flames, N. & Hobert, O. Gene regulatory logic of dopamine neuron differentiation. *Nature* **458,** 885–889 (2009).

88. Watterson, S., Marshall, S. & Ghazal, P. Logic models of pathway biology. *Drug Discov. Today* **13,** 447–456 (2008).

89. Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci.* **100,** 5136–5141 (2003).

90. Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol.* **15,** 289–294 (2014).

91. Roybal, K. T. *et al.* Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors. *Cell* **167,** 419–432.e16 (2016).

92. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333,** 1307–11 (2011).

93. Guye, P. *et al.* Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6. *Nat. Commun.* **7,** 1–12 (2016).

94. Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D. & Arkin, A. P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 8617–22 (2011).

95. Tamsir, A., Tabor, J. J. & Voigt, C. a. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469,** 212–5 (2011).

96. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491,** 249–253 (2012).

97. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31,** 448–452 (2013).

98. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10,** 99–105 (2014).

99. Didovyk, A., Borek, B., Hasty, J. & Tsimring, L. Orthogonal Modular Gene Repression in Escherichia coli Using Engineered CRISPR/Cas9. *ACS Synth. Biol.* **5,** 81–88 (2016).

100. Rinaudo, K. *et al.* A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat. Biotechnol.* **25,** 795–801 (2007).

101. Klug, A. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q. Rev. Biophys.* **43,** 1–21 (2010).

102. Zhang, F. *et al.* Efficient construction of sequence-specific for mTAL effectorsodulating mammalian transcription. *Online* **29,** 149–154 (2011).

103. Joung, J. K. & Sander, J. D. TALENs: A widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **14,** 49–55 (2013).

104. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science (80-.).* **326,** 1509–1512 (2009).

105. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers (vol 5, pg 374, 2008). *Nat. Methods* **5,** 575 (2008).

106. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8,** 2180–2196 (2013).

107. Farzadfard, F., Perli, S. D. & Lu, T. K. Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. *ACS Synth. Biol.* **2,** 604–613 (2013).

108. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Mol. Cell* **54,** 698–710 (2014).

109. Lapique, N. & Benenson, Y. Digital switching in a biosensor circuit via programmable timing of gene availability. *Nat. Chem. Biol.* **10,** 1020–1027 (2014).

110. Bowyer, J. *et al.* Mechanistic Modeling of a Rewritable Recombinase Addressable Data Module. *IEEE Trans. Biomed. Circuits Syst.* **10,** 1161–1170 (2016).

111. Nelson, V., Nagle, H. T., Irwin, J. D. & Carroll, B. *Digital Logic Circuit Analysis and Design*. (Prentice Hall, 1995).

112. Schreiber-Agus, N. *et al.* An amino-terminal domain of Mxi1 mediates anti-myc oncogenic activity and interacts with a homolog of the Yeast Transcriptional Repressor SIN3. *Cell* **80,** 777–786 (1995).

113. Lee, T. C. & Ziff, E. B. Mxi1 Is a Repressor of the c- myc Promoter and Reverses Activation by USF Mxi1 Is a Repressor of the c- myc Promoter and Reverses Activation by USF. *J. Biol. Chem.* **274,** 595–606 (1999).

114. Shankar, S. & Pillai, M. R. Translating cancer research by synthetic biology. *Mol. Biosyst.* **7,** 1802–10 (2011).

115. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–30 (2010).

116. Richter, F. *et al.* Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res.* **44,** 10003–10014 (2016).

117. Zetsche, B. *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35,** 31–34 (2017).

118. Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35,** 789–792 (2017).

119. Tang, X. *et al.* A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. plants* **3,** 17018 (2017).

120. Huang, Y. & Maraia, R. J. Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human. *Nucleic Acids Res.* **29,** 2675–90 (2001).

121. Blazeck, J., Garg, R., Reed, B. & Alper, H. S. Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnol. Bioeng.* **109,** 2884–2895 (2012).

122. Sun, J. *et al.* Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **109,** 2082–2092 (2012).

123. Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* **56,** 343–9 (2014).

124. Lewis, J. D. & Izaurflde, E. The Role of the Cap Structure in RNA Processing and Nuclear Export. *Eur. J. Biochem.* **247,** 461–469 (1997).

125. Dunn, E. F., Hammell, C. M., Hodge, C. A. & Cole, C. N. Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev.* **19,** 90–103 (2005).

126. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* 1–10 (2013). doi:10.1093/nar/gkt163

127. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell* **154,** 1380–1389 (2013).

128. McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41,** 1–10 (2013).

129. Blazeck, J. & Alper, H. S. Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8,** 46–58 (2013).

130. Trifonov, E. N. Cracking the chromatin code: Precise rule of nucleosome positioning. *Phys. Life Rev.* **8,** 39–50 (2011).

131. Rando, O. J. & Winston, F. Chromatin and transcription in yeast. *Genetics* **190,** 351–387 (2012).

132. Keung, A. J., Joung, J. K., Khalil, A. S. & Collins, J. J. Chromatin regulation at the frontier of synthetic biology. *Nat. Rev. Genet.* **16,** 159–171 (2015).

133. Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J. & Khalil, A. S. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. *Cell* **158,** 110–120 (2014).

134. Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: Advances through genomics. *Nat. Rev. Genet.* **10,** 161–172 (2009).

135. Rajkumar, A. S., Dénervaud, N. & Maerkl, S. J. Mapping the fine structure of a eukaryotic promoter input-output function. *Nat. Genet.* **45,** 1207–1215 (2013).

136. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. *Bioeng. Bugs* **3,** 38–43 (2012).

137. Naumova, N., Smith, E. M., Zhan, Y. & Dekker, J. Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods* **58,** 192–203 (2012).

138. Shinohara, A. & Ogawa, T. Homologous recombination and the roles of double-strand breaks. *Trends in Biochemical Sciences* **20,** 387–391 (1995).

139. Storn, R. & Price, K. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *J. Glob. Optim.* **11,** 341–359 (1997).

140. Didovyk, A., Borek, B. B., Tsimring, L. & Hasty, J. Transcriptional regulation with CRISPR-Cas9: principles, advances, and applications. *Curr. Opin. Biotechnol.* **40,** 177–184 (2016).

141. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* **160,** 339–350 (2015).

142. Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K. & Severinov, K. Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3′-terminal segment of guide RNA. *Nucleic Acids Res.* **44,** 2837–2845 (2016).

143. Briner, A. E. *et al.* Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Mol. Cell* **56,** 333–339 (2014).

144. Chuai, G. hui, Wang, Q. L. & Liu, Q. In Silico Meets In Vivo: Towards Computational CRISPR-Based sgRNA Design. *Trends in Biotechnology* **35,** 12–21 (2017).

145. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (80-.).* **351,** 84–88 (2015).

146. Wong, N., Liu, W. & Wang, X. WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **16,** 1–8 (2015).

147. Jakočiūnas, T. *et al.* Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. *Metab. Eng.* **28,** 213–222 (2015).

148. Schwartz, C. M., Hussain, M. S., Blenner, M. & Wheeldon, I. Synthetic RNA polymerase III promoters facilitate high efficiency CRISPR-Cas9 mediated genome editing in *Yarrowia lipolytica*. *ACS Synth. Biol.* acssynbio.5b00162 (2015). doi:10.1021/acssynbio.5b00162

149. McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science (80-.).* **353,** 237–241 (2016).

150. Yin, L. *et al.* Multiplex conditional mutagenesis using transgenic expression of Cas9 and sgRNAs. *Genetics* **200,** 431–441 (2015).

151. Ma, H. *et al.* Pol III promoters to express small RNAs: Delineation of transcription initiation. *Mol. Ther. - Nucleic Acids* **3,** e161 (2014).

152. Schmid, M. & Jensen, T. H. Nuclear quality control of RNA polymerase II transcripts. *Wiley Interdiscip. Rev. RNA* **1,** 474–485 (2010).

153. Port, F. & Bullock, S. L. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat. Methods* **13,** 852–854 (2016).

154. Fei, J. *et al.* Tissue- and time-directed electroporation of CAS9 protein–gRNA complexes in vivo yields efficient multigene knockout for studying gene function in regeneration. *npj Regen. Med.* **1,** 16002 (2016).

155. Weber, J. *et al.* Functional reconstitution of a fungal natural product gene cluster by advanced genome editing. *ACS Synth. Biol.* **6,** 62–68 (2017).

156. Yoshioka, S., Fujii, W., Ogawa, T., Sugiura, K. & Naito, K. Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. *Sci. Rep.* **5,** 18341 (2016).

157. Liu, Y. *et al.* Directing cellular information flow via CRISPR signal conductors. *Nat. Methods* (2016). doi:10.1038/nmeth.3994

158. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517,** 583–8 (2015).

159. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.* **7,** 11750 (2016).

160. Yuen, G. *et al.* CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. *Nucleic Acids Res.* **45,** 12039–12053 (2017).

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

162. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350,** 87–96 (2002).

163. McIsaac, R. S. *et al.* Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae. *Mol. Biol. Cell* **22,** 4447–4459 (2011).

164. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4,** 975–986 (2015).

165. Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K. & Arkin, A. P. Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Res.* **40,** 5775–86 (2012).

166. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165,** 535–550 (2016).

167. Lahtvee, P. J. *et al.* Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Syst.* **4,** 495–504.e5 (2017).

168. Shen, Z. *et al.* Conditional knockouts generated by engineered CRISPR-Cas9 endonuclease reveal the roles of coronin in C. elegans neural development. *Dev. Cell* **30,** 625–36 (2014).

169. Besada-Lombana, P. B., McTaggart, T. L. & Da Silva, N. A. Molecular tools for pathway engineering in Saccharomyces cerevisiae. *Curr. Opin. Biotechnol.* **53,** 39–49 (2018).

170. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14,** 297–301 (2017).

171. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32,** 1262–1267 (2014).

172. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–198 (2014).

173. Jensen, K. T. *et al.* Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett.* **591,** 1892–1901 (2017).

174. Xu, L., Zhao, L., Gao, Y., Xu, J. & Han, R. Empower multiplex cell and tissue-specific CRISPR-mediated gene manipulation with self-cleaving ribozymes and tRNA. *Nucleic Acids Res.* **45,** e28 (2017).

175. Liu, X. *et al.* High□throughput CRISPRi phenotyping identifies new essential genes in Streptococcus pneumoniae. *Mol. Syst. Biol.* **13,** 931 (2017).

176. Kulkarni, V., Kharisov, E., Hovakimyan, N. & Kim, J. Load capacity improvements in nucleic acid based systems using partially open feedback control. *ACS Synth. Biol.* **3,** 617–26 (2014).

177. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–1123 (2013).

178. Futcher, a B. & Cox, B. S. Copy number and the stability of 2-micron circle-based artificial plasmids of Saccharomyces cerevisiae. *J. Bacteriol.* **157,** 283–90 (1984).

179. de Silva, C. & Walter, N. G. Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based on the hammerhead ribozyme. *Rna* **15,** 76–84 (2009).

180. Redden, H., Morse, N. & Alper, H. S. The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Research* **15,** 1–10 (2015).

APPENDIX A Code for Chapter 2

All relevant code can be found at https://github.com/wmvoje

APPENDIX B Supplement for Chapter 3

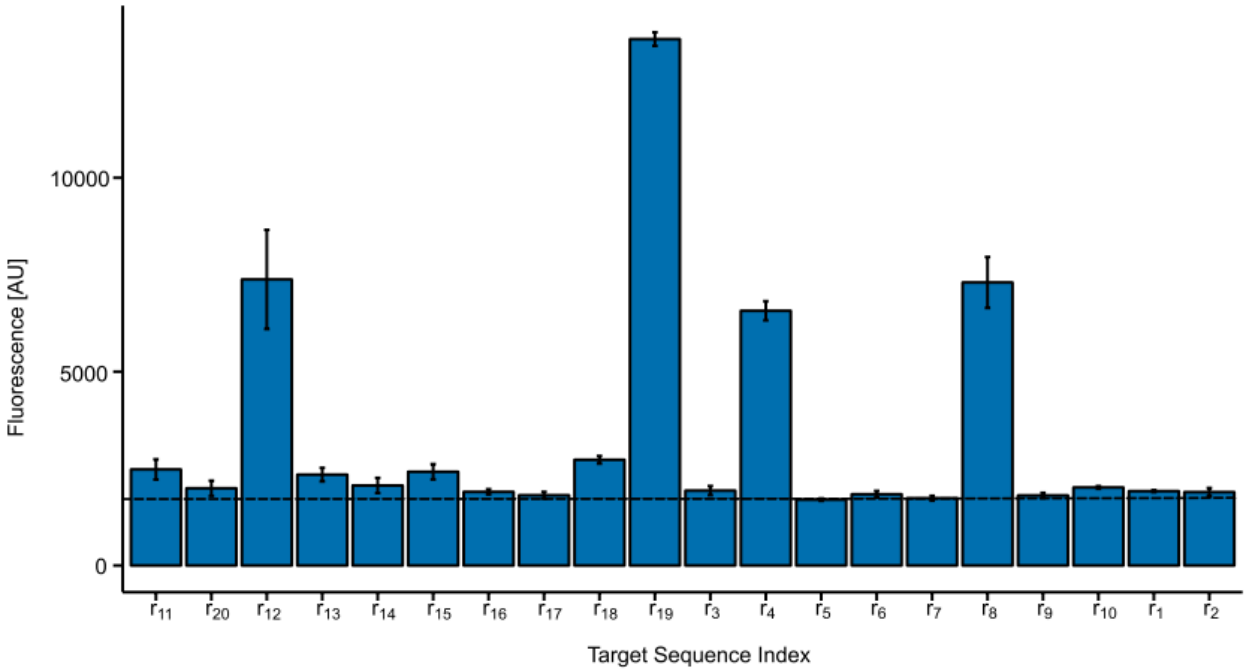


Figure B.1: Diagonal of orthogonality matrix repression variation

Bar chart representation of the diagonal of the orthogonality matrix from Figure 3.2. Sixteen of the twenty guide sequences, when matched with their cognate promoter, show GFP repression near or at the level of autofluorescence for diploid *S. cerevisiae*. Autofluorescence, 1718.63 AU, is indicated by the black dashed line. Four of the guide sequences exhibit significantly worse repression. The sixteen sequences that exhibit strong repression also exhibit variation in level of repression, indicating different levels of efficacy for each guide sequence. Error bars are standard deviation of fluorescence measurements from three biological replicates collected during one experimental run.

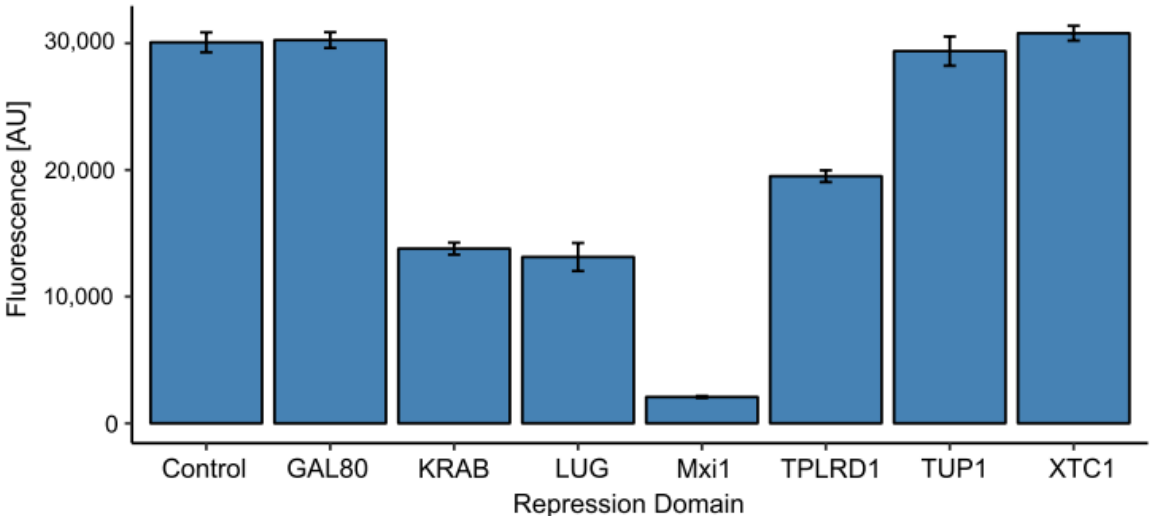


Figure B.2: Repression Domain Comparison

Protein fusions of dCas9 and a panel of repression domains that have been shown to function in yeast were compared for repression level. Repression domain fusions were expressed constitutively, along with a pGRR promoter driving GFP and a constitutively expressed cognate RGR. The control strain contains a pGRR promoter driving GFP alone. The repression domains tested are GAL80, KRAB, LUG, Mxi1, TPLRD1, TUP1 and XTC1. Mxi1 shows the greatest amount of repression. Error bars are standard deviation of fluorescence measurements from three biological replicates collected during one experimental run.

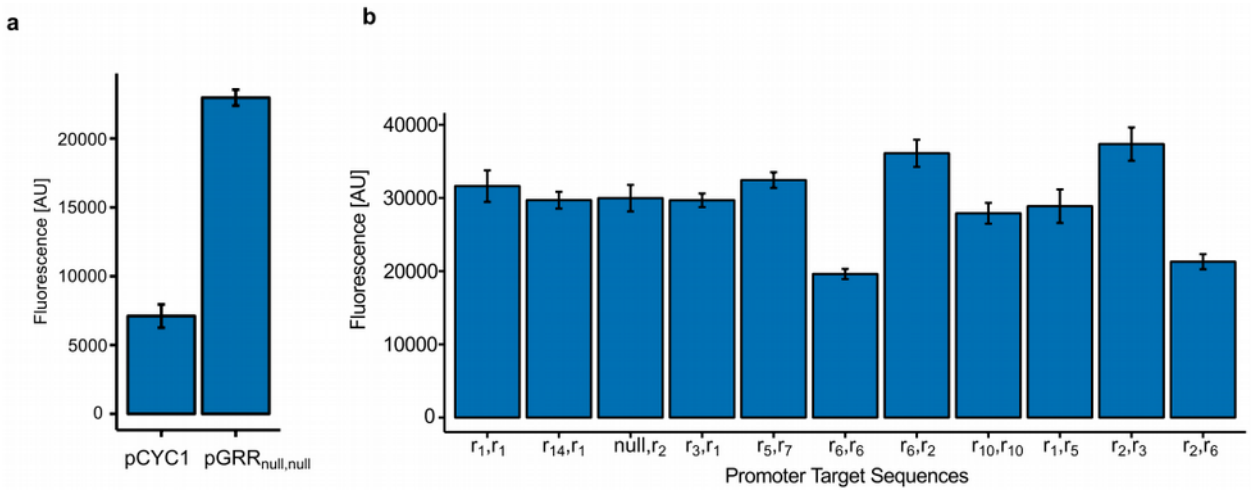
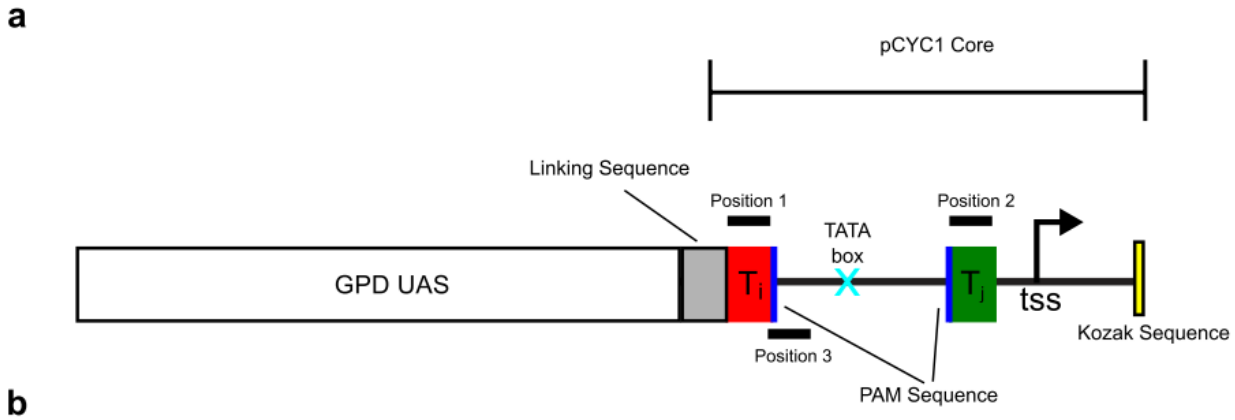


Figure B.3: pGRR promoter variability

a Addition of the pGPD upstream activating sequence (UAS), to the pCYC1 minimal promoter, increases the expression of GFP 3.23 fold. Error bars represent the standard deviation of three biological replicates collected during one experimental run. **b** A subset of 11 $pGRR_{i,j}$ promoters driving GFP has a mean fluorescence of 29511.78 [AU], a standard deviation of 5357.249 [AU] and a range of 17751.67 [AU]. Error bars represent the standard deviation of three biological replicates collected during one experimental run.



b

agtttatcattatcaatactcgccatttcaaagaatacgtaaataattaatagtagtgattttcctaactttatttagtcaaaaaattagccttttaa
ttctgctgtaacccgtacatgcccaaaatagggggcgggttacacagaatatataacatcgtaggtgtctgggtgaacagtttattcctggc
atccactaaatataatggagcccgcttttaagctggcatccagaaaaaaaaagaatcccagcaccaaaatattgttttcttcaccaacca
tcagttcataggtccattctcttagcgcaactacagagaacaggggcacaaacaggcaaaaaacgggcacaacctcaatggagtgat
gcaacctgcctggagtaaatgatgacacaaggcaattgacccacgcatgtatctatctcattttcttacaccttctattaccttctgctctctctg
atttggaaaaagctgaaaaaaaaggttgaaaccagttccctgaaattattcccctacttgactaataagtaaattcctgcagcccgggtac
tgtatNNNNNNNNNNNNNNNNNNNtggcatgcatgtgctctgtatgtatataaaactcttgttttcttctttttctctaaatattctttc
ctNNNNNNNNNNNNNNNNNNNNgcataaatt ctatacttctatagacacacaaacacaaatacacacactaatctagat
attggattctagaactagtggatctacaaa

Figure B.4: pGRR promoter sequence structure

a A sequence structure diagram that includes annotations of relevant sequence elements of the $pGRR_{i,j}$ promoter is shown. The positions 1, 2 and 3 targeted in Figure 3.2b are marked on the diagram. **b** The sequence of a typical pGRR promoter is shown along with highlighted portions that correspond to annotations from Figure B.3a. Colors of the annotations match the diagram in Supplementary figure 3a. The red and green highlighted text corresponds to target sites i and j respectively are listed as 20 Ns to indicate the programmability of these sequence elements. The non-highlighted sequence upstream of the grey highlighted portion of sequence is the GPD UAS.

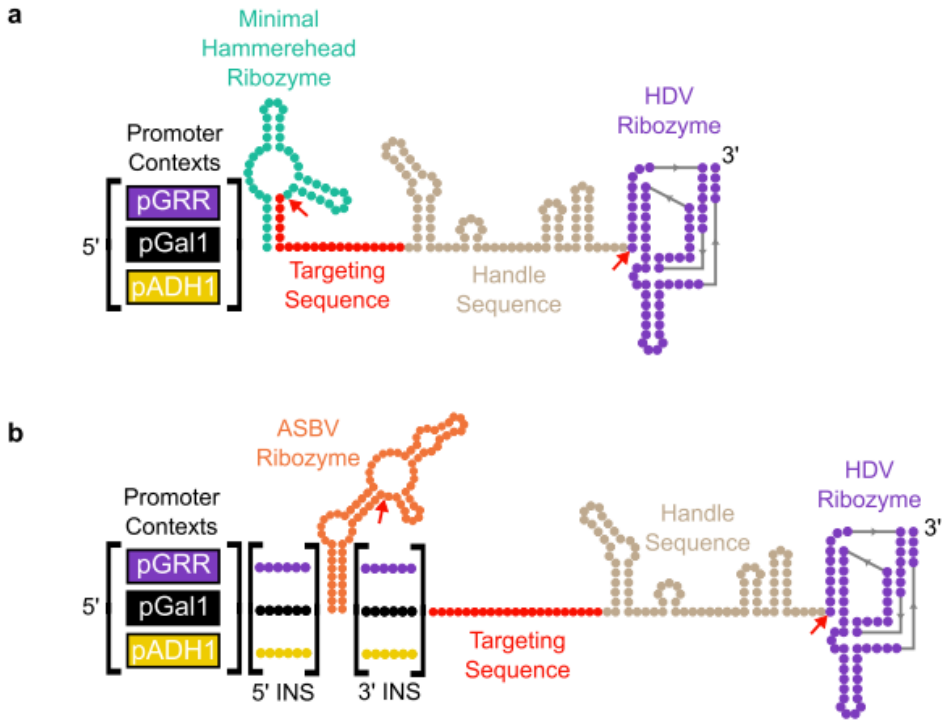


Figure B.5: Schematic of pol II gRNA expression systems

a The RGR architecture. All RGR constructs have guide sequences that were computationally predicted to confer proper folding of the minimal hammerhead ribozyme in all three promoter sequence contexts used in the work. Cleavage sites are indicated by red arrows. **b** The insulated RGR (iRGR) architecture. The iRGR has unique 5' and 3' insulating sequences, designed for three promoter sequence contexts, flanking the ASBV ribozyme. In the presence of the insulating sequences, proper ASBV folding is predicted for the majority of guide sequences. Cleavage sites are indicated by red arrows.

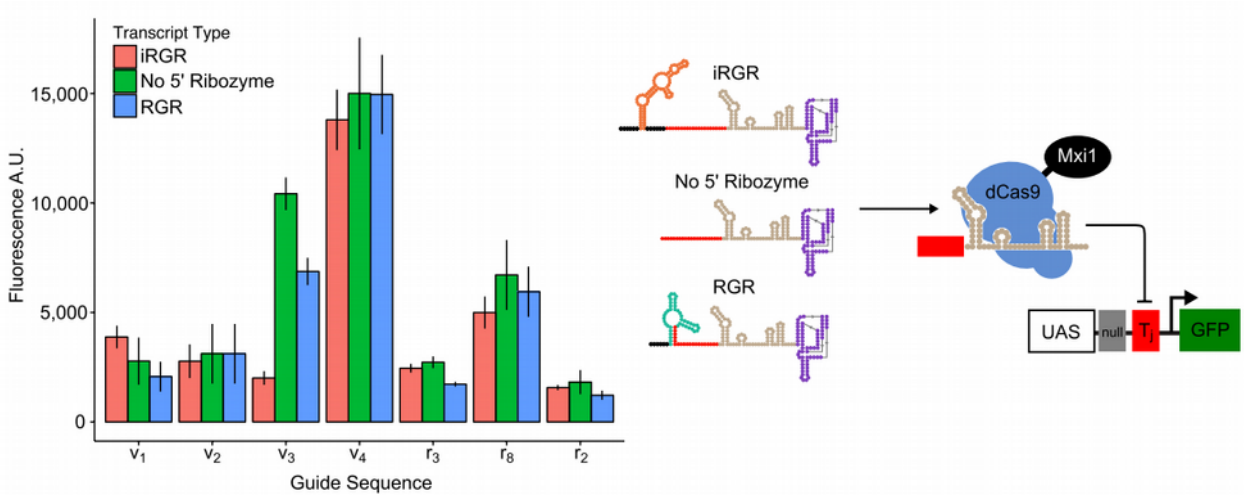


Figure B.6: Comparison of pol II gRNA expression designs

Seven gRNAs were expressed via three different designs, the RGR, the iRGR and an altered RGR design lacking the 5' ribozyme. Guide sequences r_2, r_3, and r_8 were drawn from the gRNAs used in the main body of this paper, while guide sequences v_1-v_4 were randomly generated guide sequences not contained within the original 20 component library. Fluorescence levels of repressed cognate pGRR promoters were measured via flow cytometry and error bars indicate standard deviation from 6 biological replicates, except for r_3 RGR, r_2 iRGR and r_8 iRGR which represent 5 biological replicates. Data was collected across two different experimental runs. For all three transcript types, across all seven guide sequences except for v_3, we observed comparable gRNA mediated repression of pGRR promoters. These data suggest that for many of guide sequences, the 5' ribozyme is not a contributing factor in the behavior of the gRNAs in our system.

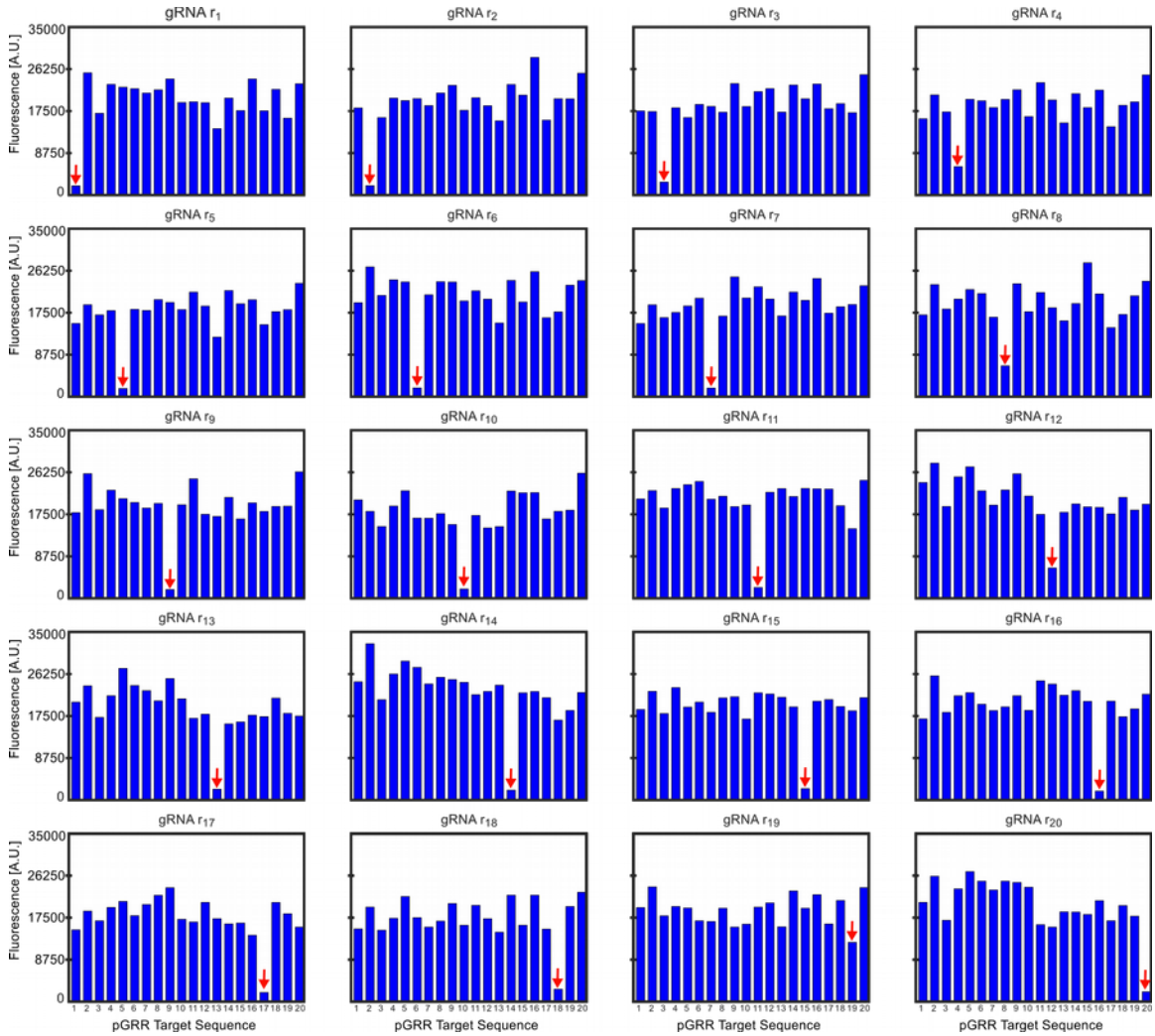


Figure B.7: Bar chart of fluorescence values of orthogonality matrix

Fluorescence values for all 400 strains in the orthogonality matrix. The strains are segmented by the 20 gRNA target sequences. Promoter target sequence index are in the same order for each subplot. Red arrows indicate a cognate pair of gRNA and pGRR promoter.

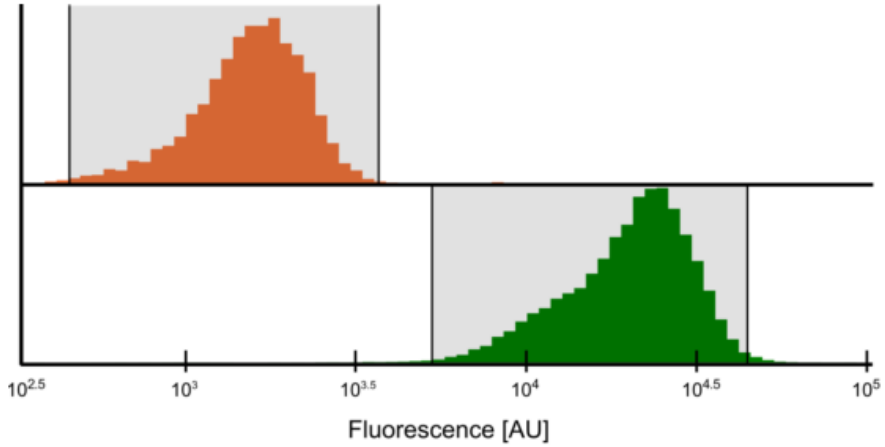


Figure B.8: ON OFF and Undefined fluorescence intervals

The ON and OFF histograms used to specify the fluorescence intervals for the circuits are shown with their middle 99% mass ranges. The histograms were generated by measuring fluorescence of a set of strains with pGRR promoters driving GFP with and without their cognate gRNAs. The OFF interval was defined by the 16 strains of the diagonal of the orthogonality matrix that were used in circuit construction (Figure 3.2 and Figure B.1). The ON interval was defined by the 16 unrepressed cognate promoters of the gRNAs used in the circuit construction. The histograms represent the sum of three biological replicates of the two strain sets. The intervals were defined by the middle 99% range of the histograms. The OFF set histogram has an upper middle 99% mass value of 3650 [AU]. The ON set histogram has a lower middle 99% mass value of 5039 [AU]. We specified the undefined interval range from 3650 [AU] to 5039 [AU]. OFF interval ranges from 0 [AU] to 3650 [AU]. The ON interval ranges from 5039 [AU] to ∞ [AU]. For a circuit to be considered in the ON or OFF state a majority of cell population must be in the correct interval.

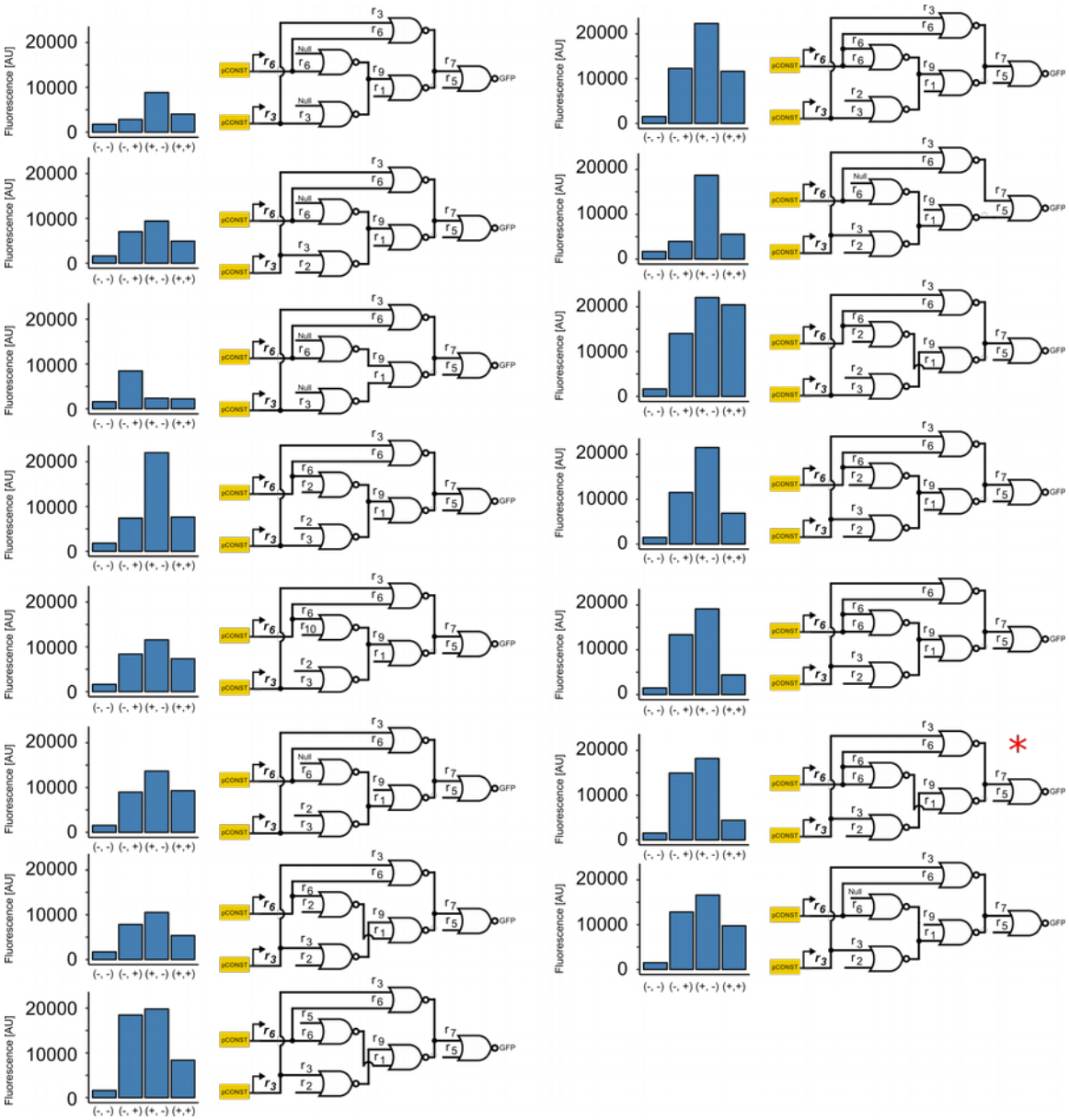


Figure B.9: XOR circuit performance variation

Fluorescence values for 15 different XOR circuit architectures containing different NOR gates exhibit variation in their output states. This illustrates the impact the performance of the component NOR gates have on overall circuit performance. The red asterisk indicates the XOR architecture that appears in Figure 3.3e. Fluorescence values represent one measurement from cells in log phase.

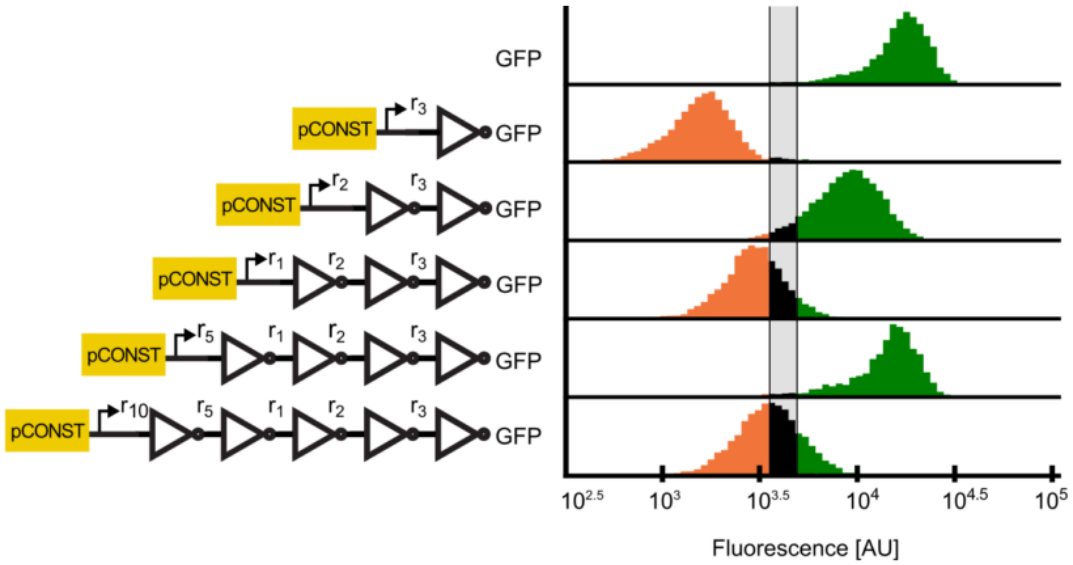


Figure B.10: Additional repression cascades

A 5-layer repression cascade. The histograms represent population fraction from three different biological replicates measured during a single experiment. Generally, additional layers added to the cascade decrease ON state fluorescence values and OFF state fluorescence values increase.

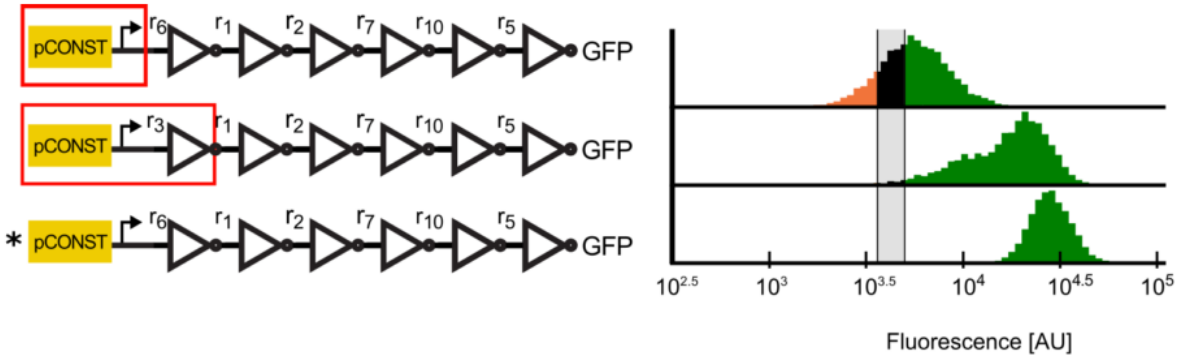


Figure B.11: Six layer cascade comparison

The fluorescence levels of two different 6 gRNA cascades are compared to the 6 gRNA cascade from Figure 3.4, indicated by a *. The histograms represent population fraction from three different biological replicates measured during a single experiment. Differences in the composition of the cascades are highlighted in red. The pCONST promoter in the top most cascade is replaced with pGRR-r_3,r_{19}. The middle cascade has the pCONST promoter replaced with pGRR-r_9. The new pCONST is expressing gRNA-r_3 instead of gRNA-r6. In addition, the promoter expressing gRNA-r1 has been replaced with pGRR-r_3,r_3. Different combinations of promoters and gRNAs yield different levels of fluorescence in the ON state for these cascades. We hypothesize this is due to variations in the parameters associated with each gate, such as promoter strength and gRNA repression strength.

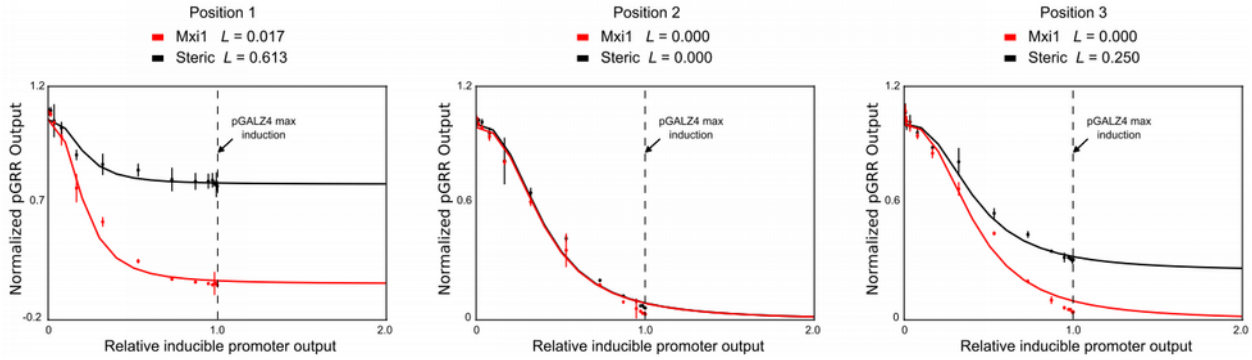


Figure B.12: Alternative dCas9-Mxi1 vs. dCas9 repression comparison dose response curves

Alternative plots of the dCas9-Mxi1 vs. dCas9 response curves from Figure 3.2b are shown as a function of relative pGALZ4 inducible promoter output. A model fit of the pGALZ4 inducible promoter dose response function was used to scale the data. The maximum level of pGALZ4 induction is indicated on the plot. The model is extrapolated past the maximum induction level of pGALZ4 to observe the full behavior of the response curves. As the induction level of the inducible promoter goes to infinity the curves settle into an asymptote equivalent to the model predicted transcriptional leak parameter *L*.

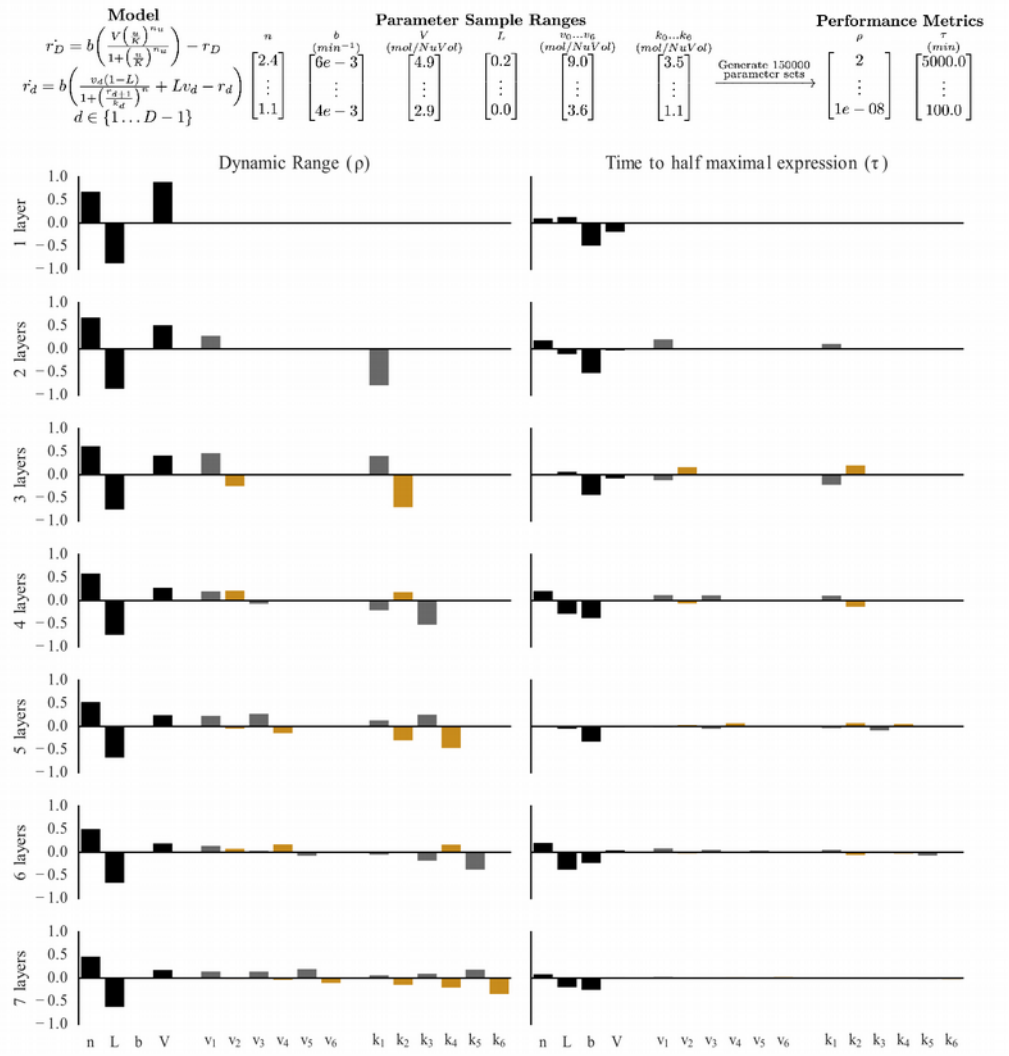


Figure B.13: Model parameter sensitivity

a 150,000 parameter sets were resampled from a uniform distribution over the intervals shown and applied to our repression cascade model (see methods). **b** Partial rank correlation coefficient (PRCC) was used to determine the contribution of each parameter has on either dynamic range or the time-to-half max. PRCCs were calculated using R (R Foundation for Statistical Computing, Vienna, Austria). Parameters associated with odd and even layers are colored grey and orange respectively. At all layers in the time-to-half maximal plot, *b* is very correlated with the output. In the dynamic range plot, n and L is strongly positively correlated at all layers with the output.

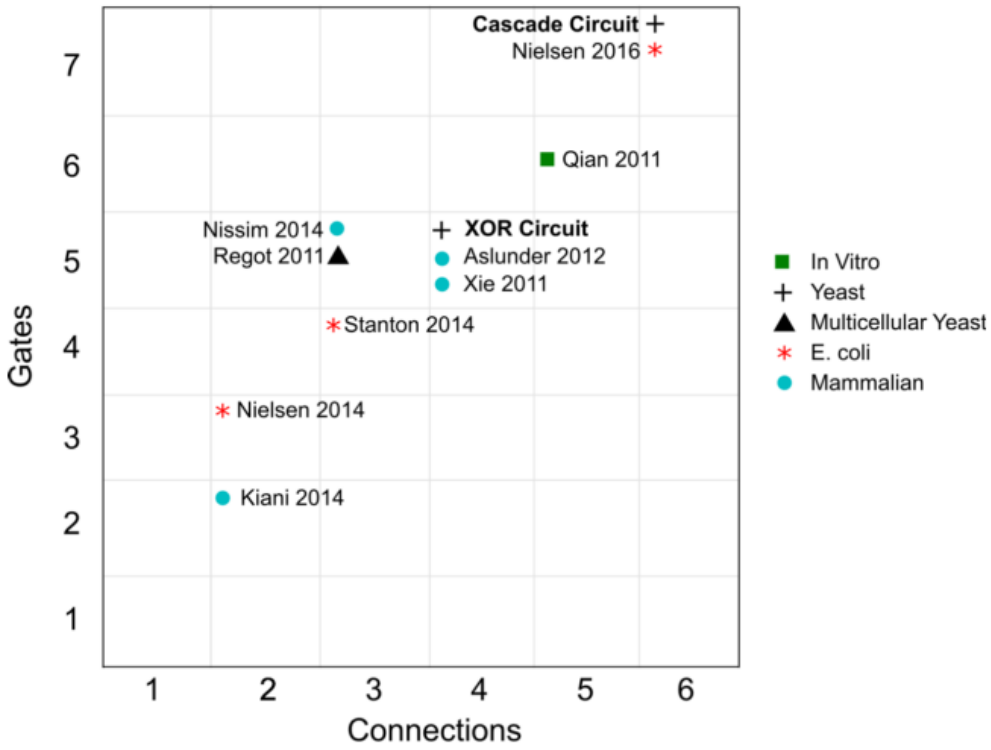


Figure B.14: Synthetic circuit size comparison

The best method for quantifying the size of synthetic biological circuits is an open question. Here we took the largest synthetic circuits constructed in nine recent publications and compared them to the two largest circuits from this paper. We separated the inputs to the circuits from internal components. We also counted the number of connections between the internal components. By our definition, a "part" is a molecular species that carries information necessary for the internal function of the circuit (as opposed to a helper protein such as dCas9). A "connection" is a molecular interaction between parts that propagates information within the circuit.

Table B.1: Guide sequence table This table lists all the guide sequences used in this work.

gRNA index	Sequence
r_1	GGAACGTGATTGAATAACTT
r_2	ACCAACGCAAAAAGATTTAG
r_3	CATTGCCATACACCTTGAGG
r_4	GAAAATCACAACTCTACTGA
r_5	GAAGTCAGTTGACAGAGTCG
r_6	GTGGTAACTTGCTCCATGTC
r_7	CTTTACGTATAGGTTTAGAG
r_8	CGCATTTCCTATTCAAACTT
r_9	GCAACCCACAAATATCCAGT
r_{10}	GTGACATAAACATTCGACTC
r_{11}	GGGCAAAGAGACGCTTGTCG
r_{12}	GAAGTCATCGCTTCTTGTCG
r_{13}	GAGTTGACAAAGTATAACTT
r_{14}	GAAGTTTCAGAATCTCGACG
r_{15}	GGCTAGGATCCATCTGACTT
r_{16}	GCAACCATAGACTCTCCAGG
r_{17}	ACCACAACTGAGTCGAACCT
r_{18}	GGGTAGCAACACTCGTACTT
r_{19}	GTAAAAGATAACTCTGTTGC
r_{20}	TCTACCCGAGACTCAAACGG
v_1	GTACATACAGTAGGATCCTA
v_2	TTTGGCACTACCGACACGAA
v_3	TGGTCAAAAGTGCGGCTTTC
v_4	CTTTCACAATCTTGACCTGC

Table B.2: pCONST promoter table

This table lists all constitutive promoter inputs for circuits built in this work.

Strain	pConst
Orthogonality Matrix Strains	pADH1:RGR$_i$
Figure 3 NOR	pADH1:RGR-r$_7$,pADH1:RGR-r$_5$
OR	pADH1:iRGR-r$_3$,pGRR:RGR-r$_6$
AND	pADH1:RGR-r$_2$,pGRR-r$_5$:RGR-r$_1$
NAND	pADH1:RGR-r$_2$,pGRR-nullnull:RGR-r$_{10}$
XOR	pADH1:iRGR-r$_3$,pGRR:RGR-r$_6$
XNOR	pAHD1:RGR-r$_2$,pGRR-nullnull:RGR-r$_{10}$
StaticCascade 1 Layer	pGRR-r$_{10}$:RGR-r$_5$
StaticCascade 2 Layer	pGRR-r$_7$:RGR-r$_{10}$
StaticCascade 3 Layer	pGRR-r$_2$:RGR-r$_7$
StaticCascade 4 Layer	pGRR-r1:RGR-r$_2$
StaticCascade 5 Layer	pGRR-r$_6$:RGR-r$_1$
StaticCascade 6 Layer	pGRR-r$_3$:RGR-r$_6$
StaticCascade 7 Layer	pGRR-r$_9$:RGR-r$_3$

Table B.3: Interval population fractions of logic circuits

This table displays the population fractions of the defined fluorescence intervals for all the static logic circuits appearing in the figure 3 and 4 and in Supplementary figures 10 and 11. A circuit's state was deemed acceptable if it met the specifications of having a majority of the cell population in the expected fluorescence interval.

NOR	(-,-)	(-,+)	(+,-)	(+,+)
ON population fraction	0.9910	0.0025	0.0028	0.0043
Undefined population fraction	0.0006	0.0025	0.0025	0.0012
OFF population fraction	0.0084	0.9951	0.9948	0.9944

OR	(-,-)	(-,+)	(+,-)	(+,+)
ON population fraction	0.0020	0.8799	0.9992	0.9923
Undefined population fraction	0.0027	0.0854	0.0004	0.0061
OFF population fraction	0.9953	0.0347	0.0004	0.0015

AND	(-,-)	(-,+)	(+,-)	(+,+)
ON population fraction	0.0177	0.0030	0.0044	0.9116
Undefined population fraction	0.0034	0.0045	0.0041	0.0642
OFF population fraction	0.9789	0.9924	0.9914	0.0242

NAND	(-,-)	(-,+)	(+,-)	(+,+)
ON population fraction	0.9560	0.8846	0.9179	0.2085
Undefined population fraction	0.0252	0.0614	0.0457	0.2047
OFF population fraction	0.0188	0.0540	0.0364	0.5868

XNOR	(-,-)	(-,+)	(+,-)	(+,+)
ON population fraction	0.9753	0.0468	0.0111	0.8768
Undefined population fraction	0.0184	0.0327	0.0308	0.0941
OFF population fraction	0.0063	0.9205	0.9581	0.0290

XOR	(-,-)	(-,+)	(+,-)	(+,+)
ON population fraction	0.0127	0.9786	0.9600	0.1337
Undefined population fraction	0.0029	0.0093	0.0060	0.1762
OFF population fraction	0.9844	0.0121	0.0341	0.6900

7 Layer Cascade	Layer 0	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 7
ON population fraction	0.9973	0.0448	0.9352	0.1262	0.7941	0.1978	0.9898	0.1680
Undefined population fraction	0.0004	0.1617	0.0500	0.2479	0.1421	0.2875	0.0032	0.3210
OFF population fraction	0.0022	0.7934	0.0148	0.6259	0.0638	0.5147	0.0069	0.5110

5 Layer Alternate Cascade	Layer 0	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
ON population fraction	0.9719	0.0219	0.9059	0.0732	0.9601	0.1844
Undefined population fraction	0.0141	0.0187	0.0653	0.2074	0.0222	0.2942
OFF population fraction	0.0139	0.9595	0.0288	0.7194	0.0177	0.5215

Alternate 6 Layer Cascades	Original	Alternate 1	Alternate 2
ON population fraction	0.9898	0.9969	0.6214
Undefined population fraction	0.0032	0.0019	0.2382
OFF population fraction	0.0069	0.0013	0.1403

Table B.4: Parameter fit values

Table describing parameter estimates for model from differential evolution for the steady-state and kinetics experiments. Using an estimated nuclear volume of ~6 μm^3 for diploid yeast and the published dissociation constant of ~1.2nM for Cas9 binding to its cognate site, bounds for parameters during optimization were selected based on estimates of transcription rates and protein degradation rates in yeast. Standard deviations of the steady-state parameters were determined from three independent experiments. Kinetic parameters were determined from a single experiment and do not have an estimate for experimental error.

parameter	mean	std	units	fitting bounds	description
V^{ss}	16.854	1.073	Molecule NucVol^{-1}	(0.434, 130.078)	Maximum transcription from inducible promoter
K	2.880	NA	nM	(2.880, 2.880)	Michaelis-Menten constant for βe inducible promoter
n_u	1.239	NA	dimensionless	(1.239, 1.239)	hill-coefficient for inducible promoter
v_0^{ss}	1.000	NA	AU	NA	Max fluorescence of reporter normalized to 1.0
v_1^{ss}	31.114	2.436	Molecule NucVol^{-1}	(0.434, 43.359)	Maximum transcription from pGRR promoter
v_2^{ss}	20.876	2.469	Molecule NucVol^{-1}	(0.434, 43.359)	Maximum transcription from pGRR promoter
v_3^{ss}	21.183	4.107	molecule/NucVol^{-1}	(0.434, 43.359)	Maximum transcription from pGRR promoter
k_0^{ss}	1.000	NA	Molecule NucVol^{-1}	NA	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
k_1^{ss}	6.129	0.992	Molecule NucVol^{-1}	(0.434, 43.359)	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
k_2^{ss}	12.229	4.065	Molecule NucVol^{-1}	(0.434, 43.359)	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
k_3^{ss}	11.782	3.442	Molecule NucVol^{-1}	(0.434, 43.359)	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
n^{ss}	1.882	0.107	dimensionless	(2.384, 2.385)	hill-coefficient
$V^{kinetics}$	12.923	0.019	Molecule NucVol^{-1}	(0.434, 130.078)	Maximum transcription from inducible promoter
$v_0^{kinetics}$	1.000	NA	AU	NA	Max fluorescence of reporter normalized to 1.0
$v_1^{kinetics}$	23.631	0.266	Molecule NucVol^{-1}	(0.434, 43.359)	Maximum transcription from pGRR promoter
$v_2^{kinetics}$	19.367	0.172	Molecule NucVol^{-1}	(0.434, 43.359)	Maximum transcription from pGRR promoter
$v_3^{kinetics}$	19.054	0.173	Molecule NucVol^{-1}	(0.434, 43.359)	Maximum transcription from pGRR promoter
$k_0^{kinetics}$	1.000	NA	Molecule NucVol^{-1}	NA	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
$k_1^{kinetics}$	6.771	0.072	Molecule NucVol^{-1}	(0.434, 43.359)	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
$k_2^{kinetics}$	12.988	0.061	Molecule NucVol^{-1}	(0.434, 43.359)	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
$k_3^{kinetics}$	14.411	0.121	Molecule NucVol^{-1}	(0.434, 43.359)	dissociation constant of gRNA-dCas9-Mxi1 to its cognate promoter
B	0.005	0.000	min^{-1}	(0.003, 0.011)	degradation/dilution of GFP reporter
b	0.006	0.000	min^{-1}	(0.003, 0.011)	degradation/dilution of gRNA-dCas9-Mxi1
$n^{kinetics}$	2.231	0.006	dimensionless	(0.500, 3.000)	hill-coefficient
Reported transcription rate	0.03 – 0.5	-	min^{-1}	-	transcription rate of most S. cerevisae promoters
Reported protein degradation rate	0.00019 – 0.058	-	min^{-1}	-	degradation rate of most S. cerevisae proteins
Kd of Cas9	1.200	-	nM	-	dissociation constant of gRNA-dCas9 to target measured *in vitro*

APPENDIX C Supplement for Chapter 4

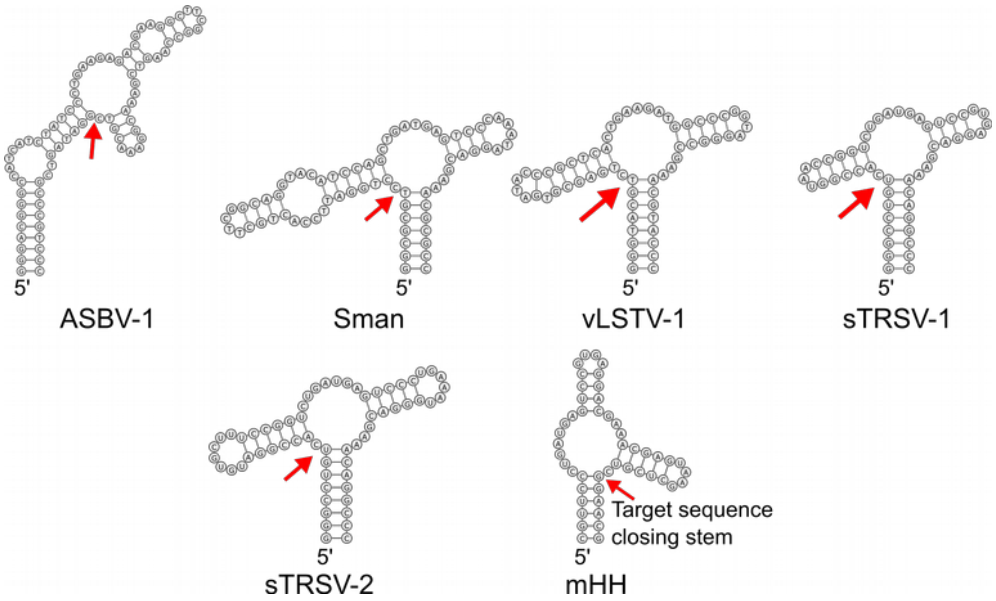


Figure C.1: Putative catalytically active secondary structures of Rbzs

Red arrow indicates where cis-mediated cleavage occurs.

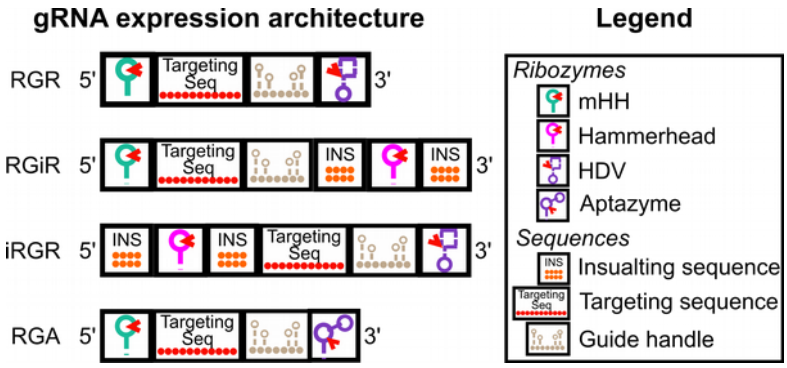


Figure C.2: Graphical description of gRNA expression architectures.

Though not indicated graphically the mHH ribozyme of the RGR, RGiR, and the RGA is partially composed of the target sequence.

Table C.1: gRNA expression architecture sequences.

This corresponds to the graphical description of devices displayed in Figure S2. The 3' half of the ribozyme post cleavage is highlighted in yellow.

Name	Sequence 5' -> 3'
Ribozyme - sTRSV-1	GGGCCTGTCACCGGTAACCGGTCTGATGAGTCCGTGAGGACGAAACAGGCCC
Ribozyme - sTRSV-2	GGGCCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCCTGAAATGGGACGAAACAGGCCC
Ribozyme - vLTSV-1	GGGTACGTCTGAGCGTGATACCCGCTCACTGAAGATGGCCCGGTAGGGCCGAAACGTACCC
Ribozyme - vLTSV-1-Sman	GGCGCGTCTGAGCGTGATACCCGCTCACTGAAGATGGCCCGGTAGGGCCGAAACGCGCC
Ribozyme - Sman	GGCGCGTCCTGGATTCCACTGCTTCGGCAGGTACATCCAGCTGATGAGTCCCAAATAGGACGAAACGCGCC
Ribozyme - ASBV-1	GGGACGGGCCATCATCTATCCCTGAAGAGACGAAGGCTTCGGCCAAGTCGAAACGGAAACGTCGGATAGTCGCCCGTCCC
Ribozyme - mHH	[NNNNNN]'CTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC[NNNNNN]
Ribozyme + Target Sequence - mHH W5	CGTTCCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCGGAACGTGATTGAATAACTT
Ribozyme + Target Sequence - mHH W8	GTTGGTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCACCAACGCAAAAAGATTTAG
Ribozyme + Target Sequence - mHH W10	TGGCAATGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCATTGCCATACACCTTGAGG
Ribozyme - HDV	GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC
Aptazyme - Theo(A)AAAAA	AAACAAACAAAGCTGTCACCGGAATACCAGCATCGTCTTGATGCCCTTGGAAGTCCGGTCTGATGAGTCCAAAAAGGACGAAACAGCAAAAAGAAAAATAAAAA
Aptazyme - Theo(A)CAAGUGAA	AAACAAACAAAGCTGTCACCGGAATACCAGCATCGTCTTGATGCCCTTGGAAGTCCGGTCTGATGAGTCCCAAGTGAAGGACGAAACAGCAAAAAGAAAAATAAAAA
Guide handle	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
Target Sequence - W8	ACCAACGCAAAAAGAUUUAG
Target Sequence - W5	GGAACGTGATTGAATAACTT
Target Sequence - W10	CATTGCCATACACCTTGAGG

Table C.2: Insulating sequences for iRGR and RGiR devices

Device Name	Insulating sequences	
	5'	3'
iRGR Sman	CTCTTGACC	GGGTTCTCT
iRGR sTRSV-1	ATGATAGAGG	CTTCA
iRGR sTRSV-2	ACTT	AGTTCA
iRGR vLSTV-1	ACCCTGGGG	TCCCG
iRGR ASBV-1	GCTTAGT	ATCTCCC
RGiR Sman 119	CTACCC	CCACAAAA
RGiR vLSTV-1-Sman 119	CTACCC	CCACAAAA
RGiR sTRSV-1 1167	AGCCTGAG	TACGAGCAAG
RGiR sTRSV-2 1080	GCGC	GTTCT
RGiR vLSTV-1 1554	GTT	AACAACAA
RGiR Sman 45	CTTTGC	TCAAATCGT

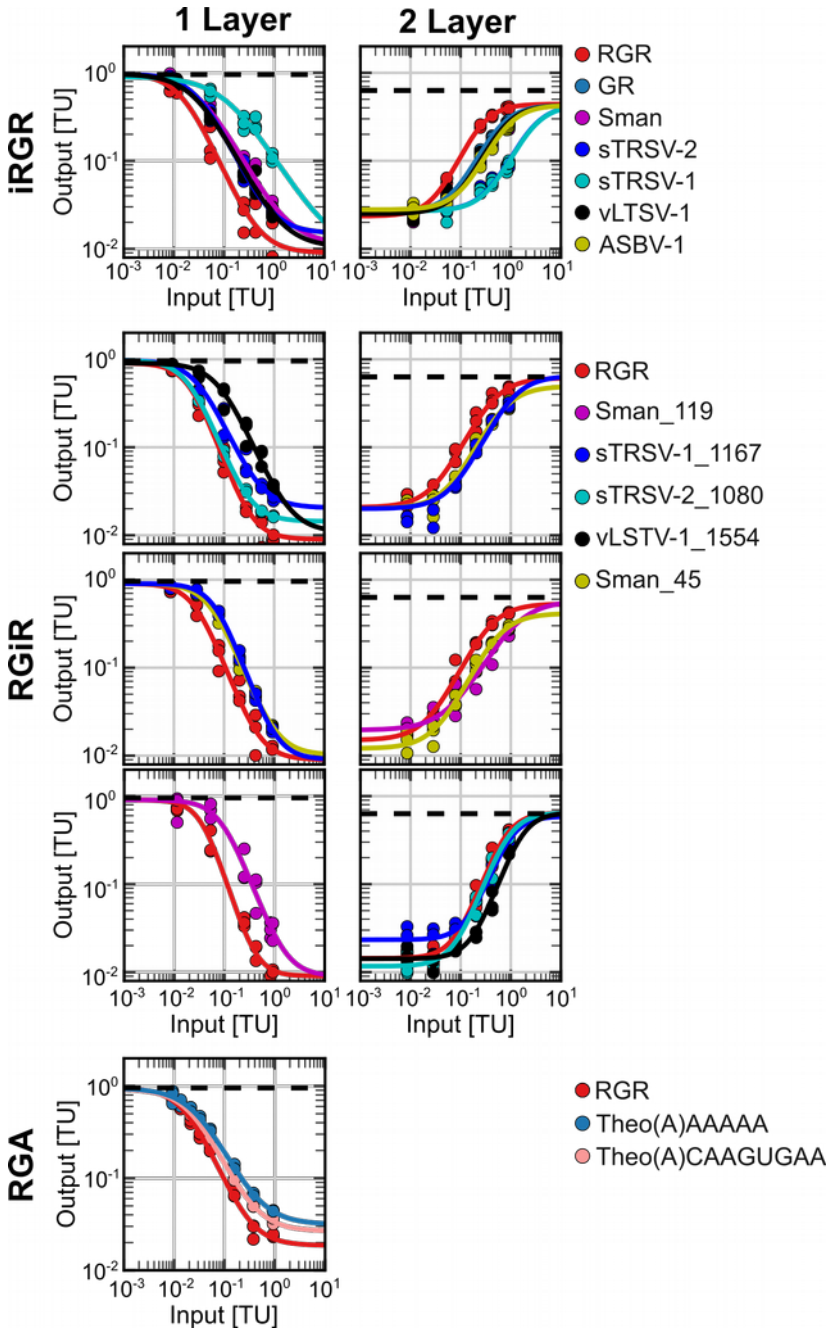


Figure C.3: iRGR, RGiR and RGA titration response data for one- and two-layer systems

Biological triplicate titration response data with Hill-equation fit are shown. RGA titration curves were collected with no theophylline induction.

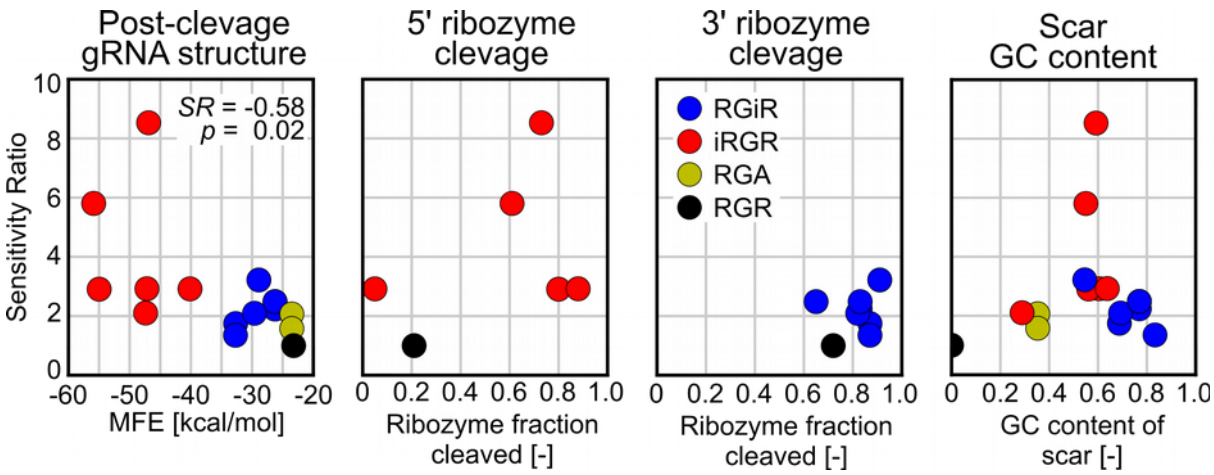


Figure C.4: Comparisons of RGR, iRGR, RGiR and RGA performance

Post-cleavage gRNA structure considered the remaining ribozyme sequence post cleavage and the remaining half of the designed insulating sequence and the gRNAW8 sequence - this was folded using viennaRNA RNAfold at 30C. Indicated SR values are the measured Spearman correlation.

Table C.3: Model parameters and quality of fit for the modeled response of RGA devices.

For each RGA device Hill equations parameters are presented for each level of theophylline induction, note that only the ka value is changing. The root mean squared error (RMSD), logarithmic root mean squared error (log RMSD), and the coefficient of determination (r2) are included to describe the quality of the model fit of the resulting data. Finally p value of the one way ANOVA calculation is presented for each cluster of three points at the indicated level of theophylline induction (Figure 4E and 4I).

RGA	Theophylline [mM]	k_a [TU^{-1}]	v_{min} [TU]	v_{max} [TU]	n [-]	RMSD	log RMSD	r^2	ANOVA p_{val}
Theo(A)CAAGUGAA	0	0.010	0.025	1.13	-1.88	0.03	0.23	0.91	0.99
	1	0.015	0.025	1.13	-1.88	0.02	0.07	0.98	0.82
	5	0.020	0.025	1.13	-1.88	0.04	0.11	0.97	0.27
	10	0.023	0.025	1.13	-1.88	0.05	0.14	0.95	0.37
	20	0.037	0.025	1.13	-1.88	0.04	0.12	0.96	0.39
	35	0.059	0.025	1.13	-1.88	0.05	0.07	0.94	0.67
Theo(A)AAAAA	0	0.007	0.025	1.13	-1.20	0.01	0.15	0.97	1.00
	1	0.007	0.025	1.13	-1.20	0.01	0.13	0.96	0.90
	5	0.009	0.025	1.13	-1.20	0.01	0.07	0.99	0.10
	10	0.013	0.025	1.13	-1.20	0.02	0.09	0.95	0.07
	20	0.019	0.025	1.13	-1.20	0.01	0.05	0.99	0.59
	35	0.036	0.025	1.13	-1.20	0.03	0.08	0.97	0.002

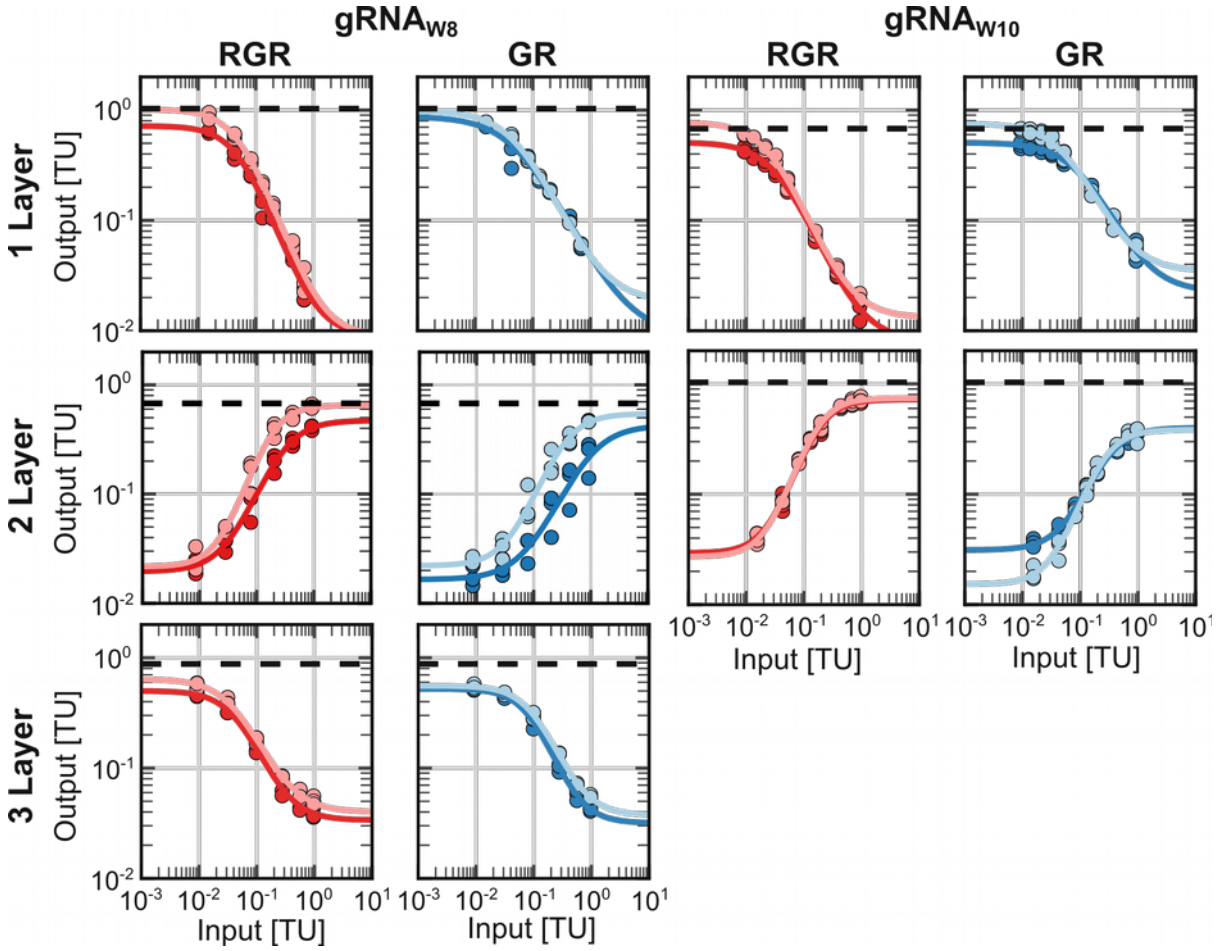


Figure C.5: Effects of limited dCas9 on system performance.

gRNA W8 and W10 inputs are expressed from RGR (Red) or GR (Blue) devices. The darker shade of data are in high dCas9 expression, pGPD:dCas9-Mxi1, the lighter shade are data with low dCas9 expression, pGRR:dCas9:Mxi1. Data are shown for 1, 2, and 3 layer devices. Biological triplicate titration response data with hill equation fit are shown.

Table C.4: Hill equation parameter ranges for XOR gate simulations.

	Low	High
V_{min} [TU]	0.005	0.024
V_{max} [TU]	0.500	1.300
k_a [TU^{-1}]	0.010	0.100
n [-]	1.033	1.800

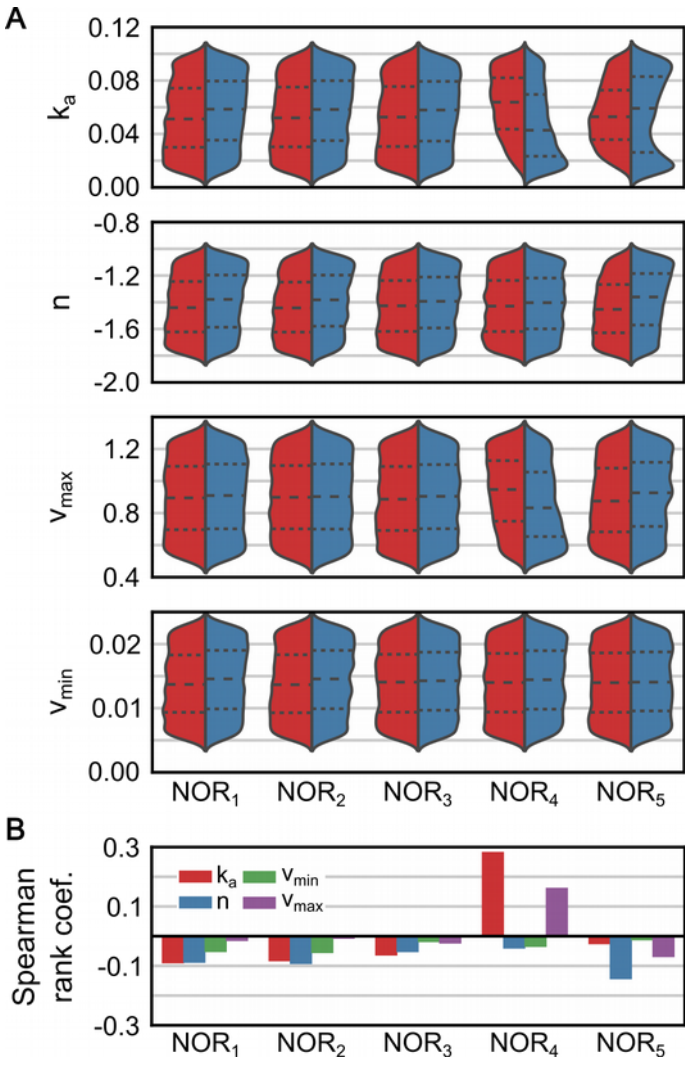


Figure C.6: Parameter sampling analysis of XOR gate simulations.

(A) Distribution of parameter values which resulted in successful (red) and unsuccessful (blue) simulated XOR gates for each NOR gate. NOR gate numbers reference Figure 6D. Data are presented as a violin plot with long and short dashed lines represent the median and upper and lower quartile of the distributions. Parameters of NOR gates which significantly affect XOR gate performance should display significant differences part performance in show significant differences in distributions. nred = 10,670, nblue = 9330 (B) Spearman rank coefficients for the parameters of each NOR gate (n=20,000). Spearman rank was calculated for each NOR gate parameter relative to the successful/unsuccessful outcome of the XOR gate. All presented values had pval < 0.001.

APPENDIX D Future directions and preliminary data

During my graduate career I did more research than could make it into publications. These research thrusts were primarily concerned with the CRISPRi transcriptional network. Most of these ideas are my own and those which where influenced by others are appropriately attributed.

D.1 Plasmid based shunt sites change CRISPRi dynamics (att. Justin Vrana)

Preliminary experiments were done which introduced a 2micron plasmid containing 'decoy' target sites which dilute the input-RGR (Chapter 4) response. The inclusion of "decoy" targets which serve no functional purpose other than to occupy transcription factors has been shown to decrease background expression (decrease vmin)[53], decrease sensitivity to input the (increase ka), and in some cases sharpen response function (increase n) [45]. To test the applicability of this in our system we designed a "shunt plasmid" which contains decoy target sites to titrate input gRNA. The shunt is realized with a 2μ plasmid containing seven decoy target sites spaced between 45 nucleotides of random sequence (Figure 3.2A). 10 to 30 copies of 2μ plasmid are maintained in yeast resulting in 70 to 210 copies of our decoy sites[178]. This number of decoy sites compared to the singly integrated target site cause a significant shift in the input gRNA responsiveness in both one and two-layer cascades. Low levels of RGRW8 and GRW8 induction, in one and two-layer cascades, show deceased levels of functional input in the presence of the shunt – demonstrating that the shunt plasmid effectively decreases background expression. Similarly there is a clear increase in the amount of input required to achieve the same level of output (increased ka) confirming that the shunt site titrates functional gRNA from their target (Figure 3.2B). Because this shunt is a 2μ plasmid this modification is very "portable" meaning shunts can easily be designed and integrated into completed systems to tune specific component response.

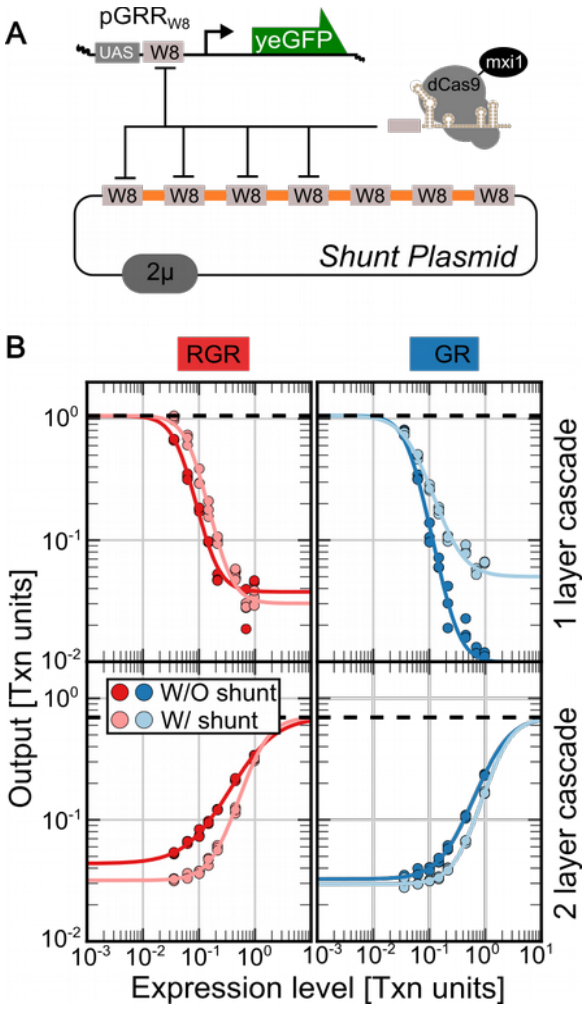


Figure D.1: Shunt plasmid impact on system response to RGR-input

(A) Cartoon of shunt plasmid decoy sites titrating dCas9 activity. (B) One and two layer dose responses of $gRNA_{W8}$ cascades from Figure 2B with or without shunt plasmids present.

D.2 Incorporating gRNA competition into computational models

Computational prediction of CRISPRi networks will be expanded to account for experimentally observed system-level impacts and will be used for the forward design of gRNA repression networks. A computational model able to accurately consider system level retroactivity paired with well characterized components would dramatically improve our ability

to build functioning circuits de novo[16]. The current computational framework (Chapter 4) enables easy specification of a network topology that once defined allows for specification of parameters for individual gates (instantiated as transfer functions). This network of interconnected transfer functions can easily be solved by passing the output values of upstream nodes to downstream nodes. However, this model does not consider any retroactive effects that arise in larger systems. This model can be expanded in the following ways: parameters such as the point of half-maximal induction (ka) or the hill-coefficient (n) could be made into a function that considers the total concentration of dCas9. Partitioning of dCas9 could be simulated by scaling the available dCas9 by the expression of all other gRNA species in the system weighted by an estimated relative affinity parameter. More fine grained models which consider individual biochemical species are able to accurately predict these interactions[49,56]. So another option would be to expand and modularize the model of the previous work[42] to easily describe arbitrary genetic networks. This model considered the system on more of bio-molecular level, but the parameters describing the system did not consider dCas9 availability and reduced gRNA activity to a single parameter. This model could be expanded to intermediate gRNA that binds to the available dCas9 with a specific affinity. Successful implementation of these concepts will dramatically decrease the false-positive prediction rate of the modeled systems. Even without marked improvement the current form of the model still would help to decrease the construction of failed constructs.

D.3 Tuning constitutive promoters for RGA expression

RGA expression will be used to implement switch behavior for an RGA input, additional aptazymes will be integrated to create additional aptazyme devices, and an RGA will be made into an inducible gRNA expression platform through constitutive promoter tuning. In a genetic circuit the RGA functions as a “normally-closed switch,” which only allow gRNA signal to pass if there is no input theophylline. This switch behavior will be tested in the context of a genetic circuit. We will build and characterize additional RGA devices. In collaboration with David Sparkman-Yager in my lab I will help characterize these newly developed riboswitch aptazymes in an RGA context. These computationally designed apatazymes are specified to have low

background cleavage in their uninduced state, an issue which plagues the dynamic range of most aptazyme devices[179]. Low uninduced cleavage is especially useful for RGA applications as this would result in effectively no functional gRNA activity. I will also consider additional characterized theophylline aptazymes[13]. Finally I will create theophylline inducible CRISPR transcription factor. Optimal levels of constitutive expression, which result in the maximal dynamic range of effective gRNA expression, will be identified using the RGA characteristic equation (Figure D.2). Promoters with desired strength will be selected from a characterized part library[180] and will be characterized in yeast.

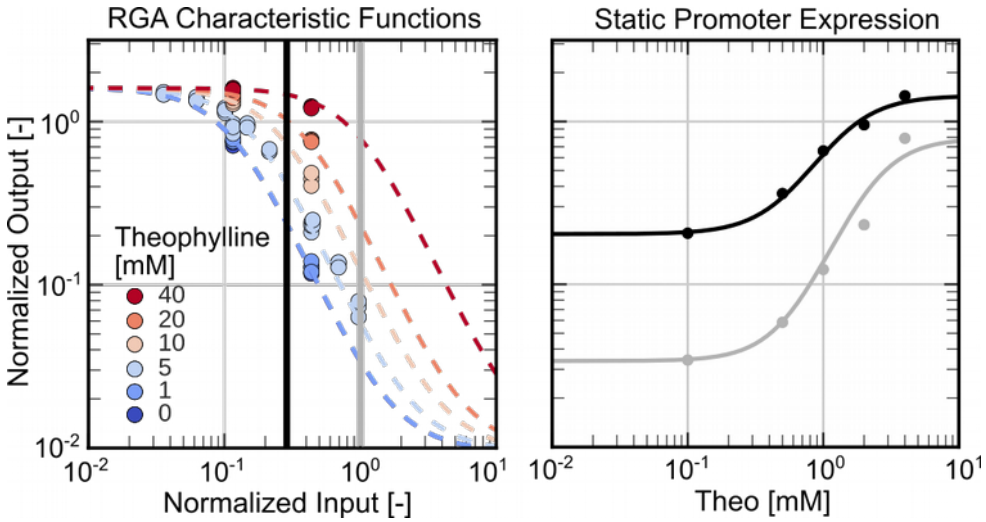


Figure D.2: RGA expression from constitutive promoters

On the left is the characteristic equation of an RGA and on the right is the predicted theophylline response when driven from a constitutive promoter. The black and gray vertical lines (left) indicated the constitutive promoter strength that would result in the black and gray response curves (right)

D.4 Promoter design to tune gRNA promoter response

pGRRs containing more inputs will be created to investigate how they effect response of components. Preliminary data shows that a pGRRw8,w8 promoter exhibits a much sharper

response to the same RGRw8 input (Figure D.3). Presently, it is unclear whether this increase in response is due the location of second target site or because of the target sites cooperative behavior. The remaining single input pGRR will be tested to confirm this. To test that this behavior is not only a function of the W8 target sequence we will validate this with additional target sequences. Finally we will explore how more than two target sites effect response dynamics.

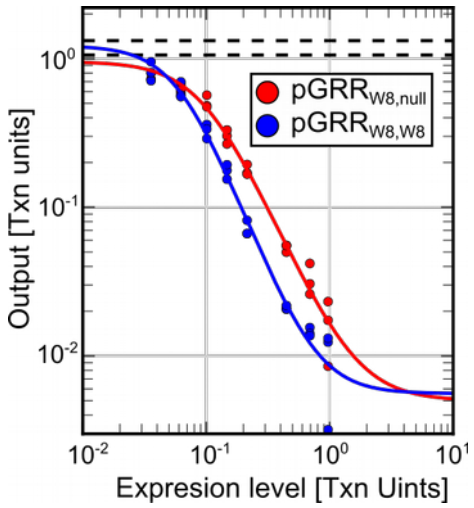


Figure D.3: Tuning CRISPRi dynamics through promoter engineering

RGR_{W8} dose response in a one layer cascade targeting either a $pGRR_{W8}$ or $pGRR_{W8,W8}$ reporter.