Rational design and implementation of synthetic genetic digital logic circuits in Saccharomyces cerevisiae

Miles W. Gander

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Eric Klavins, Chair

Jesse Zalatan

Sreeram Kannan

Program Authorized to Offer Degree:

Electrical Engineering

University of Washington

Abstract

Rational design and implementation of synthetic genetic digital logic circuits in Saccharomyces cerevisiae

Miles W. Gander

Chair of the Supervisory Committee:
Professor Eric Klavins
Department of Electrical Engineering

Biology is capable of a wide range of amazing functions, from complex pathway synthesis of high value chemicals to embryonic development. Synthetic biology seeks to harness and control the incredible potential of living systems for the betterment of society. These incredible biological functions are governed by the interaction of complex networks of interacting genes. In principle, the underlying circuitry that control biology function can be recapitulated synthetically, allowing for precise control of biology. In this dissertation, we describe our efforts to rationally design synthetic gene circuits in *Saccharomyces cerevisiae.* The main thrust of this work is the development, analysis and demonstration of a set of single gene NOR gates based on the CRISPR/Cas9 system. Using these NOR gates we constructed logic circuits with up to seven gRNAs, including repression cascades with up to seven layers. Our NOR gates allowed for the construction of the largest eukaryotic gene circuits to date. We detail the process of development of the NOR gates, their limitations and considerations for future use and improvement of the single gene circuit components. The NOR gate technology represents an advancement of the state of the art in synthetic gene circuit design 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, therapeutics, and development.

Table of contents

List of Figures

List of Tables

Acknowledgements

I would like to thank the following people for their impact on me throughout my academic career: Eric Klavins for giving me an opportunity to pursue Synthetic Biology, for all the invaluable guidance and for countless pieces of useful advice; Kyle Havens for putting up with a young inexperienced lab assistant and teaching him how to clone; Jennifer Nemhauser for her probing, insightful questions and her support; Georg Seelig for his thoughtful input on the direction of my projects; Sreeram Kannan for his class on computer communication networks and the time and effort spent on my committee; Jesse Zalatan for sharing his expertise with the CRISPR/Cas9 system; James Carothers for his positive attitude and ability to play devil's advocate; Justin Vrana for a truly synergistic collaboration, all of his work and effort truly elevated the quality of our project; Willy Voje for his unflappable desire to do good work; Michelle Parks, I wouldn't have made it without you, I wish you all the success at UW Med next year; Chris Takahashi, David Younger, Nick Bolten, Tammy Gu, Tileli Amimeur, Yaoyu Yang, Cami Cordray, Leandra Brettner, Alberto Carignano, Arjun Khakhar, Orlando de Lange and Laura Adam, thank you all for the advice, support and hard work, you made my time at UW stimulating and fun.

Dedication

To my mother, Melanie and my father, Malcom,
I would not be anywhere without your unwavering support and love.

To my sister Maddie,
Your talent, sense of humor and drive inspire me every day.

To Hayley,
Your kind and compassionate soul speaks directly to my heart.

Chapter 1

Introduction

1.1 Motivation

Living cells process information to make complex decisions. These decision-making systems commonly take the form of genetic digital and sequential logic circuits[1–8]. Decision-making systems can operate on the function on the single and multicellular level[9] where individual cells work together to process information. Decisions-making circuits mediate some of the most important and powerful processes found in biology. Examples include lysis-lysogony choice of lambda phage[10], regulation of metabolic pathways such as the galactose and glucose utilization pathways in *s. cerevisiae*[11] and asymmetric cell differentiation that leads to distinct tissue types during human development[12].

By understanding the design principles behind these decision-making circuits, we can in principle recapitulate their function through synthetically designed genetic circuits. Synthetic biologists seek to harness the computational power of biology by rationally designing and building gene circuits. These synthetic decision-making systems could have applications in diagnostics[13,14], therapeutics[14,15] and new developmental programs for tissue engineering[16,17].

1.2 Challenges for increasing size and complexity of synthetic gene circuits

Creating synthetic cellular decision making systems that reliably process information and actuate desired responses is extremely challenging. On the single cell level, genetic circuits have been demonstrated capable of forward and sequential logic and memory[18–26], but scaling up the complexity and size of such circuits has been difficult for several reasons. Firstly, genetics parts commonly used to construct these circuits are unreliable. The parts can have high

variability[21,23], non-ideal response functions[22], have crosstalk between one another and can have retroactive effects on the chassis cells in which they are used[22,27]. Secondly, generating libraries of parts can be difficult due to an involved synthesis process or a limited number of biological components that function in a desired way[27,28]. Finally, construction of the circuits can be a slow and cumbersome process. Difficulties relating to DNA assembly strategies and organism chassis considerations has slowed the scaling up of gene circuits. Additionally, even when circuits are constructed some design paradigms require large scale screens to identify functional circuits. Large scale screens can be difficult to engineer in and of themselves. In synthetic multicellular systems, the information processing capabilities are limited by the computational power of the individual cells and the limited number of well characterized cell-to-cell communication signaling pathways.

1.3 Overview of thesis contribution

In this dissertation, we present work that furthers the state of the art of the size and complexity in genetic based cellular decision making systems. In electronics, a compositional approach has allowed the construction of digital circuits of great complexity to be quickly designed and implemented. With this approach as our inspiration we sought to develop set of low-variability genetic parts that can be routinely composed to create large digital circuits in yeast cells. Such a set of parts would be the basis for a framework that allowed for large and complex synthetic gene circuits and multicellular systems.

We created a set of genetic NOR gates, designed in *S. cerevisiae,* based on CRISPR-dCas9 that addresses two main challenges associated with scaling up gene circuits. First, we designed 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. Second, we required a consistent “OFF” state for our NOR gates allowing for digital responses in our circuits. Using the

NOR gates we were able to implement large and complex digital logic circuits, including, by our measure the largest synthetic circuit ever in eukaryotes.

In this document, we will open with a broad overview of the current state of the field of genetic circuitry and then detail the development of the NOR gates. We cover the circuits achieved and those we were unable to achieve with the NOR gates. We detail the design, build and test cycle used for our circuits and the role the Aquarium software and the UW BIOFAB played in the process. We then present a method for extending the ability to rapidly construct synthetic gene circuits in yeast. The development of the NOR gates included studies of promoter design, repression domains, mathematical modeling, RNA folding, bistability and multiplex transformation strategies. A significant portion of this work has been accepted for publication as “Digital logic circuits in yeast with CRISPR-dCas9 NOR gates” in Nature Communications.

Chapter 2

The current state of synthetic gene circuitry

2.1 Current methodologies for constructing synthetic gene networks

Rationally designed synthetic circuits have the potential to transform the way humans interact with biological systems. The ability to build reliable information processing functions in cells will allow for precise control of existing useful biological functions, such as stem cell differentiation[16,17] and high-value chemical synthesis via metabolic pathways[29]. Additionally, using these circuits new biological functions could be created, such as engineered immune cells that response to specific diseases like cancer[30]. While examples of these types of synthetic biological systems have been demonstrated, size, complexity and predictability of the underlying circuits have limited the functionality of these systems. Here we review the different methodologies that are currently being used to build circuits, their drawbacks and how researchers are attempting to improve them.

DNA binding domains

A common tool for building synthetic gene networks are DNA binding domains (DBDs). In both prokaryotes and eukaryotes DBDs are applied as TFs to create transactional circuits. The DBDs can act as repressing or activating transcription factors. The activity of the DBDs can be modulated by creating fusion proteins with transcriptional effector domains. There are many examples of circuits created with DBDs that bind to their natural binding sequence, there are also programmable DBDs that have been applied as well. Each class of DBDs has different aspects to consider when construction.

A recent publication demonstrating some extremely impressive circuits using a wide variety of DBDs in *E. coli* to build large transcriptional circuits[21]. The authors carefully characterized a library of sixteen TetR homologs based on an earlier publication[23]. The response curve of the repressors was measured and using this information gates with compatible dynamic ranges in their response were wired together to form circuits. The authors were able to create circuits similar in scale to those discussed in this document. The strategy employed in this work stands in contrast to the one employed for the circuits in this document. Here the authors spent a large amount of time and resources to curate a set of parts with variable performancel then used a design tool to put the parts together in combinations which were predicted to yield a functional circuit. In this work, we attempted to create a large set of parts that functioned consistently across variants. While we observed some performance differences between components in most cases combinations of gates chosen at random yielded functioning circuits. Later in this document (Chapter 10) we propose a semi-random rapid construction strategy for building many circuit variants at once and screening for functional circuits without rationally selecting the gates appearing in circuits. This contrast between these two methods gets at a deeper question about the future of genetic circuit design. Is it better to apply a completely rational approach or is applying a random high throughput creation and screening method superior?

Zinc Finger TFs are derived from the DNA binding domain of a *Xenopus laevis* protein that is involved in regulation of the 5S RNA subunit of RNA polymerase II in early development[31]. The “fingers” consist of 30 amino acid tandem repeats that contain a Cyc2-His2 motif that is known to bind zinc atoms. The zinc atoms in this context facilitate DNA binding. Each “finger” is capable of binding specifically to three basepairs of DNA. Subunits capable of targeting any three basepair combination have been engineered[32]. To target arbitrary sequences of DNA, arrays of these subunits are cloned together. When constructed, these arrays bind to their DNA sequences with very good specificity and have little to

no off-target binding. The main drawback of ZFs is that arrays of three or more subunits have fairly high error rates for construction[33,34]. There are a few existing proprietary methods for constructing longer arrays but as of yet there is no reliable method is available to the public[32,35,36]. This makes ZFs undesirable for building the large libraries of orthogonal sequence binding TFs, an important factor necessary for large genetic network building.

Transcriptional Activator-Like Effectors (TALEs) are another programmable synthetic TF. Like ZFs, the TALE is made up of protein tandem repeats fused to an effector domain. The TALE repeats are made up of 33-35 amino acid structures combined into an array. The array is flanked on the N and C terminal with two protein caps. The protein subunits and flanking cap were derived from *Xanthomonas* spp. protobacteria, a genus of bacteria that commonly act as plant pathogens[37,38]. In contrast to ZFs, TALEs are relatively easy to construct and bind DNA targets with relatively good specificity[33]. ZFs may possess slightly better binding specificity but with the difficulty of construction with the ZFs, TALEs may have higher potential for higher utility in larger scale genetic networks that require multiple orthogonal DNA binding sequences.

In order to use the DBDs discussed above as transcription factors commonly transcriptional effector domains are fused to the DBD protein, especially in eukaryotic cells. This allows the DBD to cause targeted transcriptional changes to circuit genes. One of the most common transcriptional activating domains used in *S. cerevisiea* is the VP16 domain[39–41]. VP16 is a protein native to the human herpes simplex virus and is responsible for regulating the lytic phase of the viral life cycle[42,43]. In yeast, VP16 has been shown to interact with the TATA-binding protein and TFIIB, both important general transcription factors that comprise the PIC[44]. VP16 also directly interacts with the SAGA histone acetylase complex, an important protein complex in general transcription that possesses histone acetylation capabilities. Overall, VP16 acts by recruiting the SAGA complex, which acetylates associated histones allowing for a more accessible

DNA structure for the associated transcriptional machinery and by recruiting the general transcription factors for the PIC to promoters at an above-basal rate[45] . Additionally, multiple copies of VP16 protein domain have been fused together to illicit a stronger activating effect[41]. In general, most activating domains work similarly to VP16, through interacting with the general transcription factors in the PIC, acetylation of histones or recruitment of the RNA Pol II holoenzyme, which consists of RNA Polymerase, a subset of general TFs, and the Sbr/Med proteins[46].

In general, transcriptional repressors fall into three categories: those that interact with a specific DNA sequence and either sterically repress or affect transcriptional machinery; those that interact with activating TFs or coactivators preventing transcriptional activation; and corepressors that interact with dedicated cellular repressors[47]. The first category of repression domains is not ideal when building larger transcriptional gene networks if the target sequence in the promoter is not programmable. Lack of programmability results in non-orthogonality due to the limited number of possible instances of DNA binding sequence in a network before introducing crosstalk. ZFs and TALEs are synthetic TFs with programmable DNA sequence binding capability and can sterically hinder RNA Pol II and associated transcription machinery if the target promoter is bound in the proper place. This hindrance is purely a physical barrier to transcription and is not always effective in complete down regulation[41,48]. The second category may also not be ideal for synthetic networks because the domains rely on deactivating context-dependent TFs and therefore may lack orthogonality due to narrow-range compatibly-interacting TFs[46]. More commonly, corepressors that interact with dedicated cellular repressors are used. The TFs are constructed by fusing transcriptionally active repressor domains to programmable DNA binding proteins such as CRISPR/cas9, ZFs and TALEs. An example of this type of repressor category is Tup1, which forms a complex with CYC8. This complex recruits histone deacetylases and associates with the RNA Polymerase II mediator complex to facilitate transcriptional repression[49,50].

Domains such as TUP1 are attractive for network building due to the nature of their repression. Although, the dynamics of general repression of transcription is sometimes promoter and promoter targeting position dependent, in some publications Tup1 has been shown to activate genes. This demonstrates that there is much yet to be understood about transcriptional regulation.

RNA based regulators

Another commonly used type of transcriptional regulators used in genetic circuit design are RNA based devices. Most these regulators function by changing the structure of an mRNA molecule to change the rate of translation. Types of RNA based regulators include, aptazymes[51], which are RNA structures that change conformation in the presence of a small molecule, small interfering RNA[52], which bind mRNA and prevent transcription, and micro RNA[53], which are small RNAs that bind mRNA and modulate transcription. Additionally, these RNA types can be combined to create Riboswitches[54], which turn transcription on or off based on a change in conformation when a small molecule or small RNA binds to a larger RNA structure on an mRNA. There are also examples of RNA binding proteins that bind to specific RNA sequences, these binding proteins can mediate RNA cleavage or be tagged with effector domains to alter expression of circuit components[55,56]. It should be noted that many of these RNA regulators could be used synergistically with CRISPR/Cas9 system gRNA to further expand their functionality[55–57].

CRISPR/Cas9 based circuits

CRISPR/cas9 is an emerging technology whose system is present in many different species of Archea and Bacteria as a sort of vial immune system[58]. Essentially, small guide RNA (gRNA) sequences bind with a large nuclease protein, cas9, and target sequences that are complementary to the gRNA. Synthetic biologists have converted cas9 into a TF by eliminating the nuclease

activity of the cas9 protein (*dCas9*), fusing effector domains to it and using the gRNA targeting capability to bring the *dcas9* protein to arbitrary sequences of DNA. There are two requirements for DNA targeting with CRISPR based system. The gRNA must have approximately 20 basepairs of sequence complementarity to the DNA target and the DNA target must be adjacent to a small PAM sequence recognized by the *dCas9* protein[41,59]. These PAM sequences vary depending on which bacterial species the Cas9 protein is derived from[60]. One reported difficulty with DNA targeting in this system is the lack of specificity for DNA binding[48]. In addition, some gRNA sequences confer stronger specificity than others, possibly making equivalent orthogonal parts difficult to construct[61]. Another difficulty with this system is the expression of the small guide RNA. In the native system, the gRNA is expressed from a RNA pol III non-protein-coding promoter. In *S. cerevisiea* these RNA Pol III promoters are weak, relatively uncommon and difficult to work with. One work around is to deliver the gRNA with and RNA pol II promoter expressing a ribozyme-flanked transcript that cleaves out a functional gRNA[57]. This brings in the added complexity of RNA folding where inevitably some target sequences will confound the functionality of the ribozymes[62].

Recombinases

Another way to create a genetic circuit involves using DNA recombinases that literally rewrite the DNA code when activated by an inducer molecule[20,63–68]. The recombinases can excise specific DNA sequences and invert specific sequences of DNA. Because of the switch like nature of these enzymes they are suited to build digital circuits. A recent publication created a large set of Boolean logic functions[67]. The drawbacks of the recombinases are that they are somewhat leaky leading to unwanted circuit activity. Additionally, their logic is irreversible in many cases which limits the applications.

Conclusion

There are many different tools available to the synthetic biologist to control transcriptional activity. As of yet, the complexity of gene networks has not reached a level theoretically possible. Many of the tools that we have discussed could be combined to scale up and create large complex gene networks. One large hurdle that stands in the way of this scalability of gene networks is the engineering of orthogonality. In this review, we have pointed out the existing strategies for building this quality into circuits. Improvements to technologies such as these are currently being developed and it is only a matter of time until the true potential of synthetic biology is realized in the form of rationally designed gene circuits that perform useful functions for humankind.

Chapter 3

NOR Gate design

3.1 NOR gate overview

We built a universal, single gene NOR gate (Figure 3.1a). NOR gates are functionally complete[69] 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. 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 Pol III promoters to express gRNAs[24,59,70,71] because they have low expression levels relative to Pol II promoters and are more difficult to engineer[72,73]. 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 which corresponded to complete or near complete repression of the of the output promoter. (Figure 5.4). To achieve this, we used the chromatin remodeling repression domain *Mxi1* to take advantage of the eukaryotic cell’s ability to repress gene expression, by fusing this domain to *dCas9*[59] (Figure 5.3). When compared with a number of repression domains, *Mxi1* showed the strongest repression (Figure 6.1). The *Mxi1* domain is thought to recruit histone deacetylases[74,75], and with it we observed strong transcriptional repression in our

circuits. The strong and consistent “OFF” behavior 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. 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[13,14], therapeutics[14,15], and development[16,76].

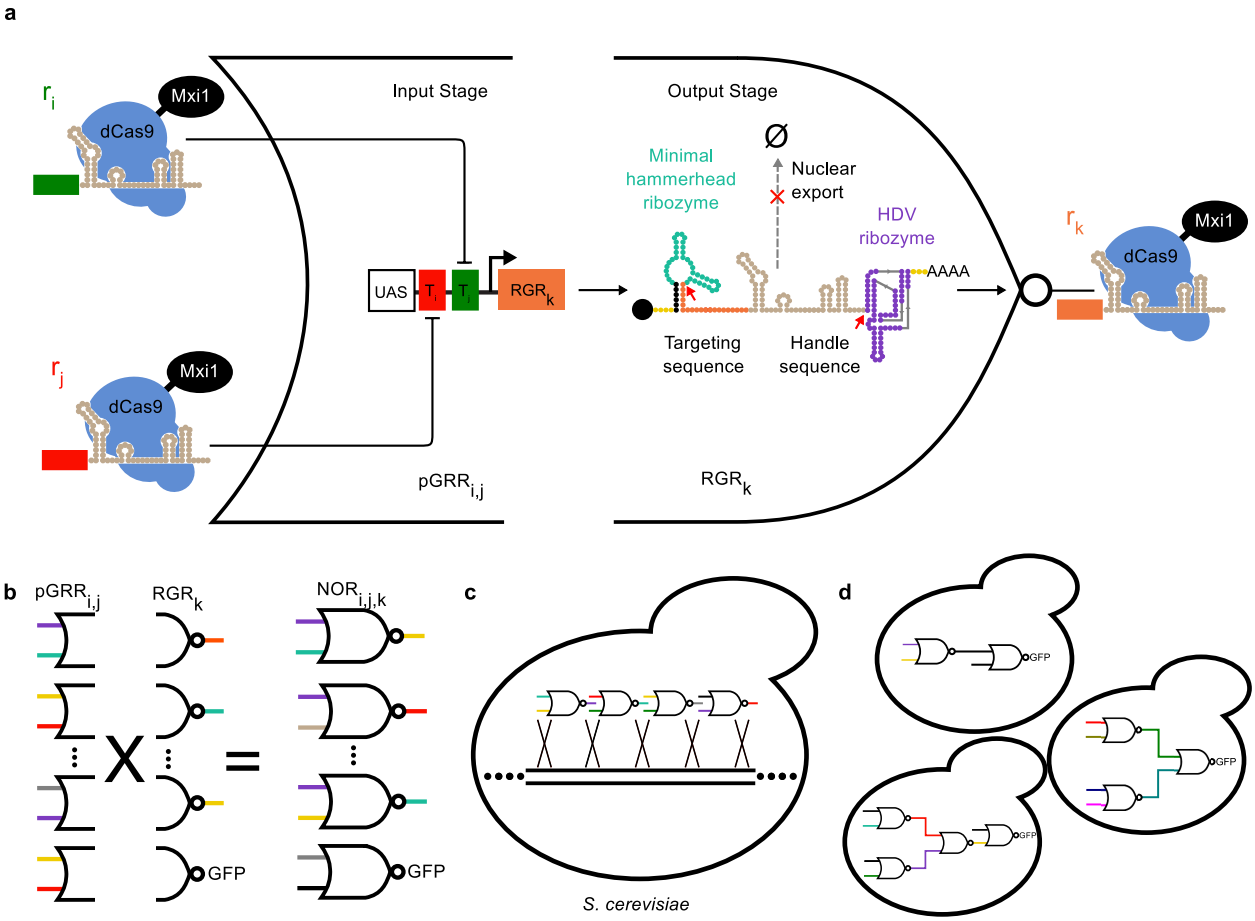


Figure 3.1 | Schematic of the NOR gate architecture and circuit composition **a** A NOR gate comprised of an input stage consisting of a Pol II gRNA Responsive Promoter (pGRR) with two distinct gRNA target sites. The pGRR promoter is fully repressed by the binding of either one or both of its cognate gRNA-*dCas9-Mxi1* complexes. The output stage of the NOR gate is a gRNA transcript, flanked by self-cleaving ribozymes (RGR). Cleavage sites 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 2-intput pGRR promoters and 20 RGR outputs, for a total of 8000 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.

3.2 NOR gate notation definition

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[59]. 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 (Table 9.2). The $pGRR_{i,j}$ promoter contains two, 20 base-pair (bp) target sites that match r_i and r_j respectively. Since we designed twenty 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_j 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.

Chapter 4

NOR gate promoter design

4.1 pGRR NOR gate promoter design

The pGRR promoter is the input stage of our genetic NOR gates. Our design goal was to create a set of promoters that could be orthogonally regulated by two input signals, had a low OFF state and had high expression in their ON states.

The $pGRR_{i,j}$ promoter is tightly repressed when gRNA-*dCas9-Mxi1* is bound to one or both of its two twenty base pair 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[41]. 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[77] was added, forming the base pGRR promoter. The UAS increased the unrepressed expression level of the pGRR output approximately three-fold 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 4.1).

A pGRR promoter map highlighting all relevant sequence features is included in (Figure 4.2). A library of 11 $pGRR_{i,j}$ promoters, with i and j chosen from the twenty guide sequences, showed limited expression variability when driving *GFP*, with an ~18% standard deviation from the mean (Figure 4.3)

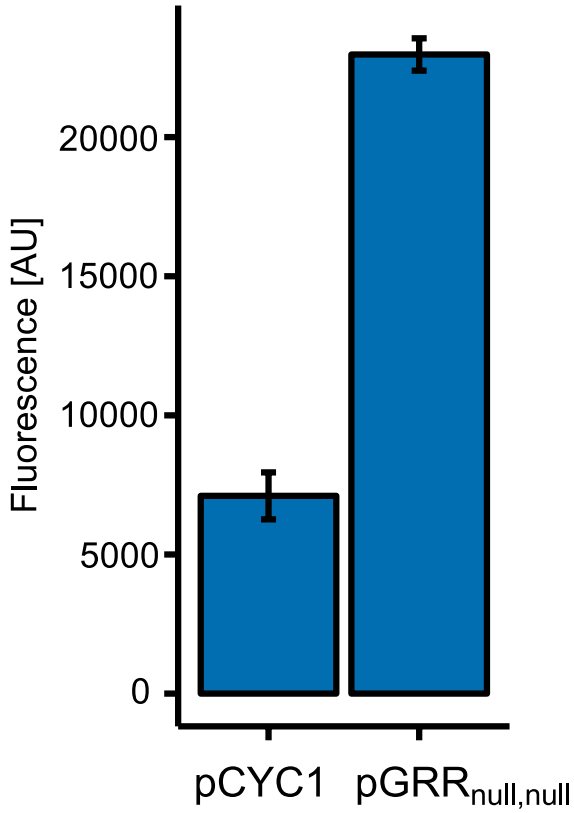
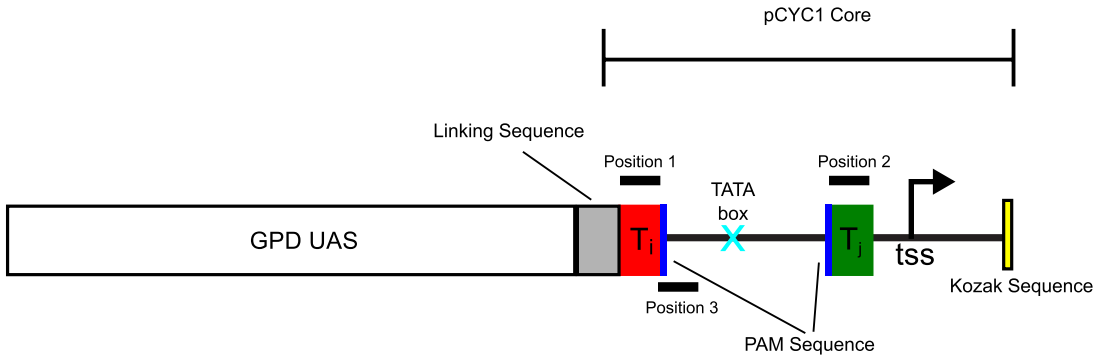


Figure 4.1 | Comparison of pCYC1 and base pGRR promoters Addition of the pGPD upstream activating sequence (UAS), to the pCYC1 minimal promoter, increases the expression of *GFP* 3.23-fold. Error bars represent the standard deviation of three biological replicates collected during one experimental run

a



b

agtttatcattatcaatactcgccatttcaaagaatacgtaaataattaatagtagtgattttcctaactttatttagtcaaaaaattagccttttaa
ttctgctgtaacccgtacatgcccaaaatagggggcgggttacacagaatatataacatcgtaggtgtctgggtgaacagtttattcctggc
atccactaaatataatggagcccgctttttaagctggcatccagaaaaaaaaagaatcccagcaccaaaatattgttttcttcaccaacca
tcagttcataggtccattctcttagcgcaactacagagaacaggggcacaaacaggcaaaaaacgggcacaacctcaatggagtgat
gcaacctgcctggagtaaatgatgacacaaggcaattgacccacgcatgtatctatctcattttcttacaccttctattaccttctgctctctctg
atttggaaaaagctgaaaaaaaaggttgaaaccagttccctgaaattattcccctacttgactaataagtaaattcctgcagcccgggtac
tgtatNNNNNNNNNNNNNNNNNNNtggcatgcatgtgctctgtatgtatataaaactcttgttttcttcttttctctaaatattctttc
ctNNNNNNNNNNNNNNNNNNNNgcataaattactatacttctatagacacacaaacacaaatacacacactaatctagat
attggattctagaactagtggatctacaaa

Figure 4.2 | 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 6.2 are marked on the diagram. **b** The sequence of a typical pGRR promoter is shown along with highlighted portions that correspond to annotations from extend data Figure 4.2a. Colors of the annotations match the diagram in Figure 4.2a. 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.

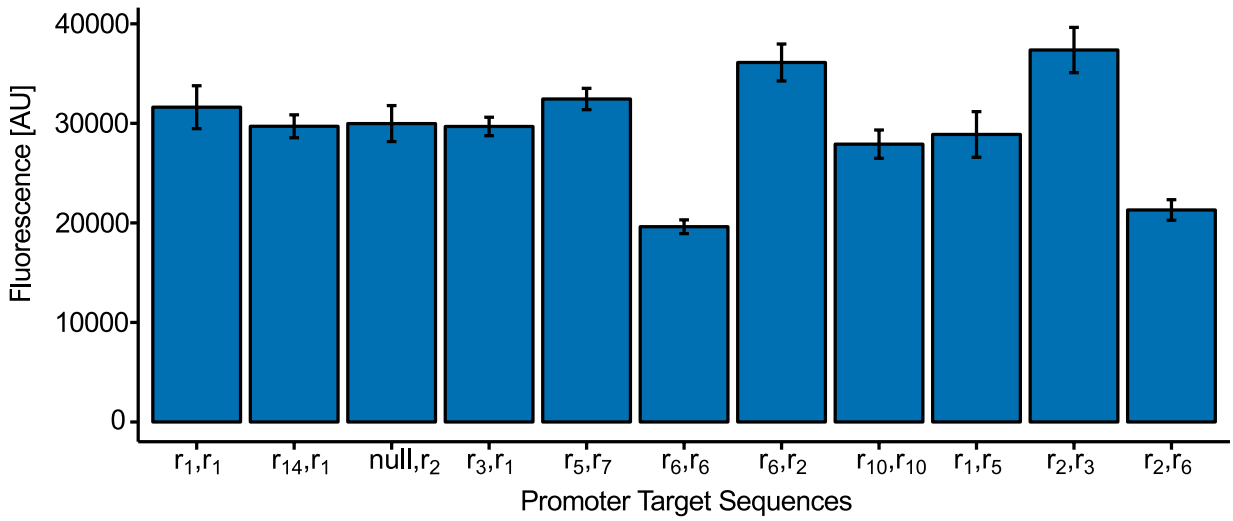


Figure 4.3 | A subset of 11 pGRR$_{i,j}$ promoters driving *GFP* This subset has a mean fluorescence of 29511.78 [AU], a standard deviation of 5357.249 [AU] and a range of 17751.67 [AU]. Error bars represent the standard deviation of three biological replicates collected during one experimental run.

4.2 pGPD pGRR prototyping

In an effort to further increase the dynamic range of the pGRR promoters a set of pGRR variants using the GPD promoter as the core promoter were created (Figure 4.4). We hypothesized that using a stronger pGPD promoter as the basis for the pGRR would create a larger separation between the ON and OFF state of the NOR gates when targeted by *dCas9-Mxi1*. To test this hypothesis three separate sites adjacent to native PAM sites in the pGPD promoter were chosen to swap in target sequences from the 20-component library used for the NOR gates. Strains were created with pGPD variants driving expression of GFP in the presence or absence of cognate gRNA (Figure 4.5).

ccttaaccagattcgaaaagcggcagtttatcattatcaatactcgccatttcaaagaatacgtaaataattaatagtagtgattttcctaactt
tatttagtcaaaaaattagccttttaattctgctgtaacccgtacatgcccaaaataggggcgggttacacagaatatataacatcgtagg
tgtctgggtgaacagtttattcctggcatccactaaatataatggagcccgcttttaagctggcatccagaaaaaaaaagaatcccagca
ccaaaatattgttttcttcaccaaccatcagttcataggtccattctcttagcgcaactacagagaacaggggcacaaacaggcaaaaaa
cgggcacaacctcaatggagtgatgcaacctgcctggagtaaatgatgacacaaggcaattgacccacgcatgtatctatctcattttctt
acaccttctattacctctctgctctctctgatttggaaaaagctgaaaaaaaaggttgaaaccagttccctgaaattattcccctacttgactaat
aagtatataaagacggtaggtattgattgtaattctgtaaatctatttcttaaacttcttaaattctacttttatagttagtctttttttttagtttaaaac
accagaacttagtttcgacggattctagaactagtggatctacaaa

Figure 4.4 | pGPD promoter sequence structure The sequence of the pGPD promoter is shown along with highlighted sequence elements. The turquoise region corresponds to target site 1, the purple regions corresponds to target site 2 and the green region corresponds to target site 3. The blue highlighted regions are native PAM sites and the highlighted yellow sequence is a Kozac sequence.

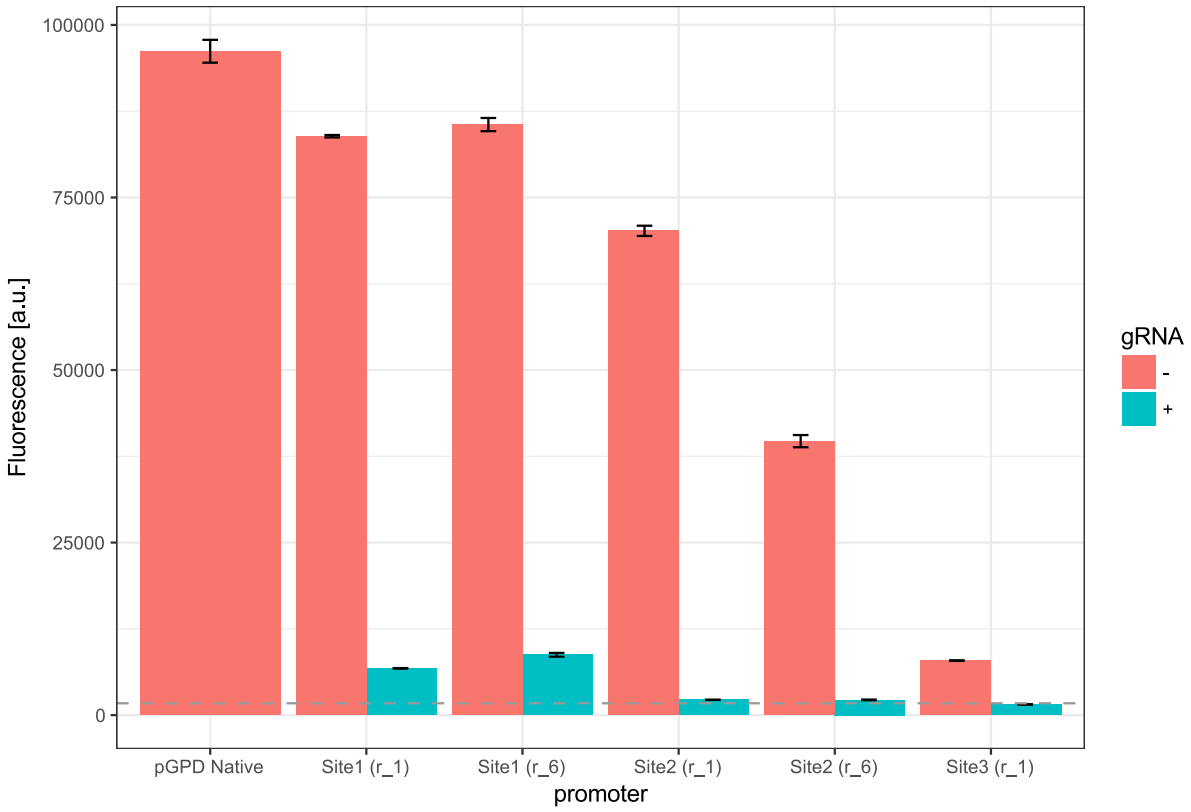


Figure 4.5 | Five variants of the GPD promoter Variants were created by replacing the three sites displayed in Figure 4.4 with either gRNA target sequence r_1 or r_6 (Sequence table). Fluorescence values for strains containing the promoter variants driving gRNA and *dCas9-Mxi1*, were collected with and without their cognate gRNA during log phase growth. The dotted grey line represents the autofluorescence level of *S. cerevisiae* 1718.63 [a.u.]. Error bars are standard deviation of fluorescence measurements from three biological replicates collected during one experimental run.

Results of the GPD variant screen revealed two interesting factors to consider when designing promoters for repression using a *dCas9-Mxi1* based repression system.

Firstly, replacing sequence of a native promoter can have a dramatic effect on expression level. The effects are likely dependent on the position of the sequence substitution and the content of the sequence being incorporated into the promoter. Deleterious effects on expression could be observed due to the

position of substitution because of deletion of natural TF binding sites. Sequence content substituted could negatively impact expression by introducing new native repressor TF binding sites[78] or by altering nucleosome occupancy based on GC base content of the new sequence[79].

Secondly, the maximum level of repression achieved with a gRNA-*dCas9-Mxi1* complex appears to be dependent on the expression level of the targeted promoter. The *dCas9-Mxi1* is a strong repressor and repressed almost all the pGRR promoters to near or at the level or yeast autofluorescence. In the case of these pGPD variants, it seems that if a promoter has a very high fluorescence level the gRNA-*dCas9-Mxi1* complex is not able to fully repress expression to autofluorescence levels. This is illustrated by the fluorescence values of the repressed Site1(r_1) and Site1(r_6) variants. The two variants repressed levels are significantly higher than the yeast autofluorescence. This is an important trend to consider when designing digital repression circuits using *dCas9-Mxi1*. Our modeling predicts that reducing transcriptional leak at maximal repression is paramount for scaling up the size of *dCas9-Mxi1* based circuits. Thus, this data suggests a limit to the expression level of gate promoter can have and still achieve effective repression in the OFF state for purposes of digital logic circuit construction.

Chapter 5

gRNA Design

5.1 RGR design

Two different RNA pol II expression methods were used in this work (Figure 5.1). The first was an RGR design utilizing a 5' minimal hammerhead ribozyme (mHH) and a 3' Hepatitis delta virus ribozyme (HDV), flanking the gRNA[57]. 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[80,81]. It has been shown that RNA device folding can be insulated from surrounding sequence context through computational sequence selection[82,83]. 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 5.2). 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[84].

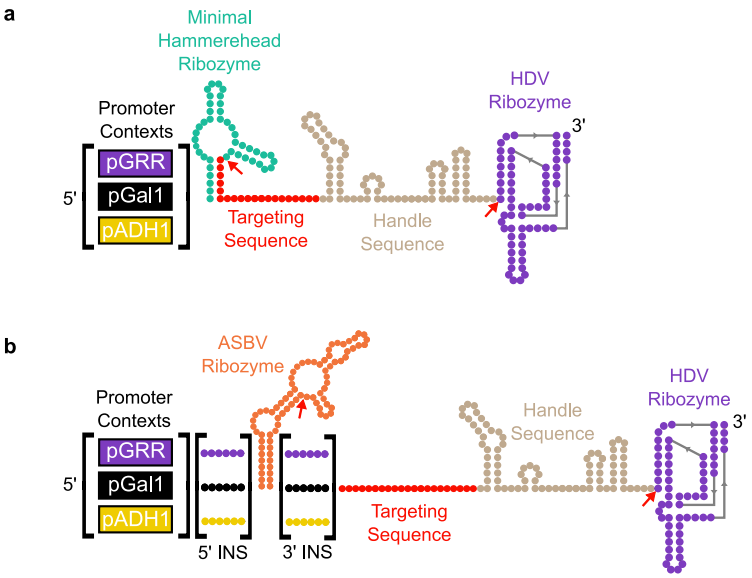


Figure 5.1 | The RGR architecture All RGR constructs have guide sequences that were computationally predicted to confer proper folding of the minimal hammerhead ribozyme in all three promoter sequence contexts used in the work. Cleavage sites are indicated by red arrows. **b** The insulated RGR (iRGR) 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.

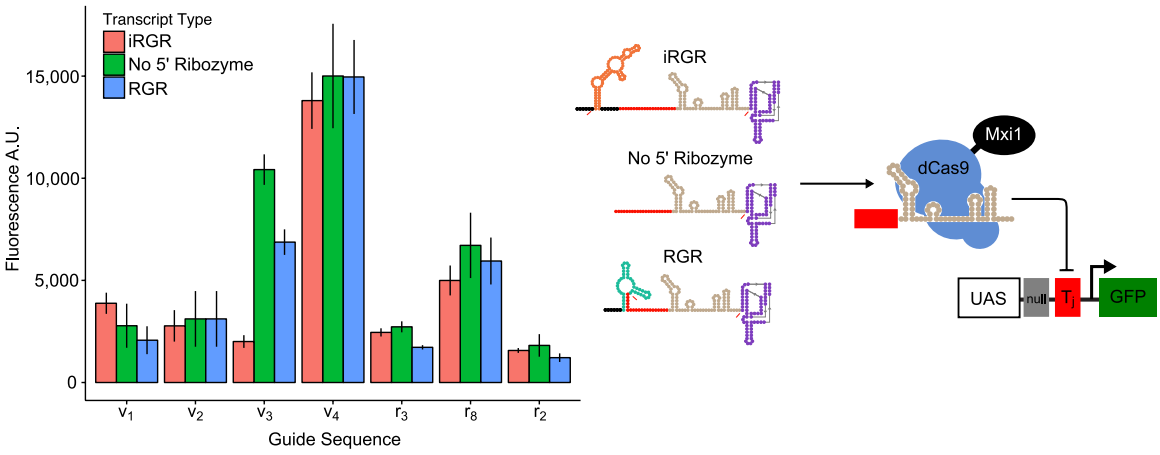


Figure 5.2 | gRNA expression system comparisons Seven gRNAs were expressed via three different designs, the RGR, the iRGR and an altered RGR design lacking the 5' ribozyme. Guide sequences r2, r3, and r8 were drawn from the gRNAs used in the main body of this paper, while guide sequences v1-v4 were randomly 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, r2 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.

RGR and iRGR sequences were computationally designed to enable the 5' 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 (co-transcriptional folding) with inputs of low and high polymerization rates of 25 nt s^{-1} and 50 nt s^{-1} respectively, helix minimum free energy = 6.346 kcal mol^{-1} and folded without pseudoknots nor entanglements. 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 Kinefolds simulation evaluation were determined by folding ribozyme sequences (Minimal HH : 5' – NNNNNNCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCNNNNNN -3' ASBV1 : 5' - GGGACGGGCCATCATCTATCCCTGAAGAGACGAAGGCTTCGGCCAAGTCGAAACGGAAACGTCGGATAGTCGCCCGTCCC -3') using RNAfold and Kinefold (melt and anneal of 1 minute), respectively. RGR targeting sequences and iRGR insulating sequences were screened in specific 5' promoter contexts (pGAL1min: AGTATCAACAAAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATACGGATTCTAGAACTAGTGGATCTACAAA, pAHD1 : CAAGCTATACCAAGCATACAATCAACTATCTCATATACAGGATTCTAGAACTAGTGGATCTACAAA, pCYC1 : ACTATACTTCTATAGACACACAAACACAAATACACACACTAATCTAGATATTGGATTCTAGAACTAGTGGATCTACAAA) and in the 3' context of the targeting sequence and the gRNA handle sequence (gRNA handle : GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT).

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 seventy-five randomly generated and ten randomly generated 20 bp guide sequences using viennaRNA and Kinefold, respectively.

5.2 Orthogonality of gRNAs

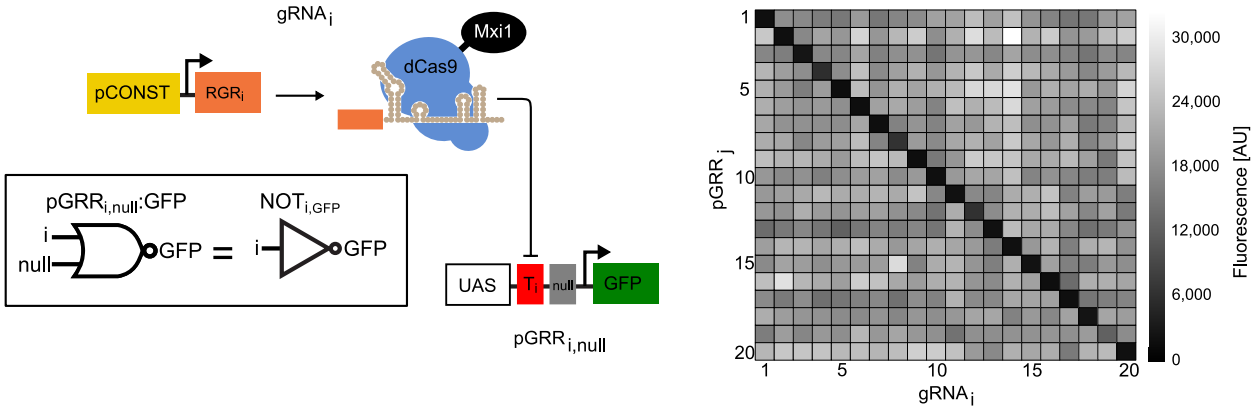


Figure 5.3 | Orthogonality via *dCas9-Mxi1* a A constitutive promoter drives expression of gRNAs paired with a combinatorial library of cognate promoters. Orthogonality of the gRNA guide sequences was tested by crossing the 20 pGRR$_{i,null}$ promoters, each expressing *GFP*, with the 20 gRNA$_i$, creating 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.

No significant crosstalk was observed when all $r_{1\text{-}10}$ (RGR design) and $r_{11\text{-}20}$ (iRGR design) were paired with all $pGRR_{1\text{-}20,null}$:*GFP* among non-cognate pairs (Figure 5.3 and 5.5). Sixteen out of twenty total RGRs ($RGR_{1\text{-}10}$ and $iRGR_{11\text{-}20}$) when targeted to their cognate $pGRR_{1\text{-}20,null}$:*GFP* constructs, repressed fluorescence to or near the level of autofluorescence for *S. cerevisiae* (Figure 5.4). These data suggest that guide sequence content of gRNAs can influence the amount of repression they are capable of when complexed with *dCas9-Mxi1.* Possible explanations for this variation of repression level could be guide sequence interference with gRNA handle sequence folding, guide sequence interference with ribozyme cleavage in the RGR context or guide effects on the rate of complexing with *dCas9-Mxi1*. Until more experiments are performed to study which sequences will confer poor repression, screening each gRNA for performance is the best method for identifying functional guide sequences.

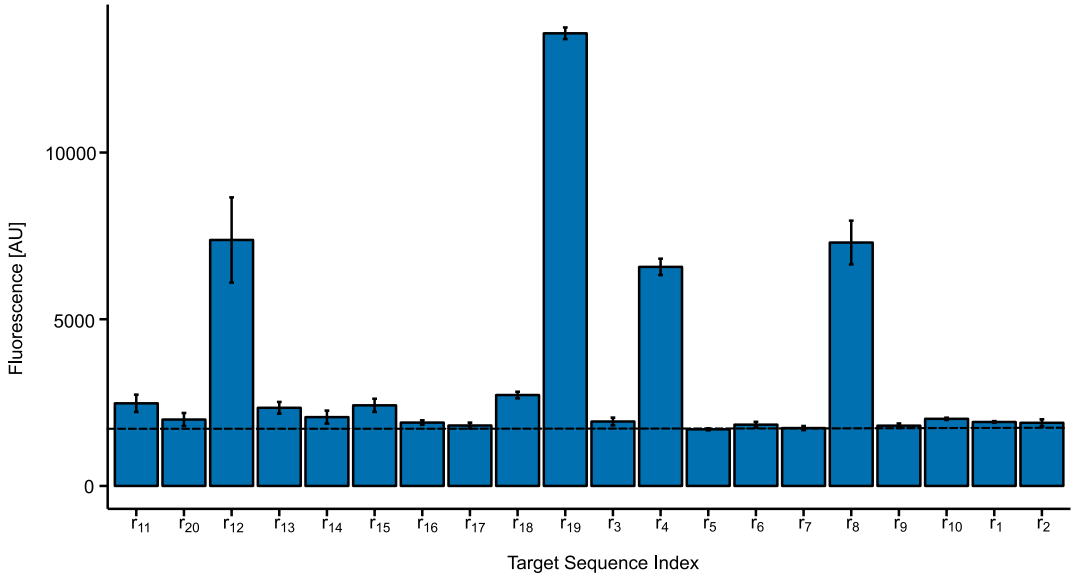


Figure 5.4 | Diagonal of orthogonality matrix repression variation Bar chart representation of the diagonal of the orthogonality matrix from Figure 5.2. Sixteen of the twenty guide sequences, when matched with their cognate promoter, show *GFP* repression near or at the level of autofluorescence for 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.

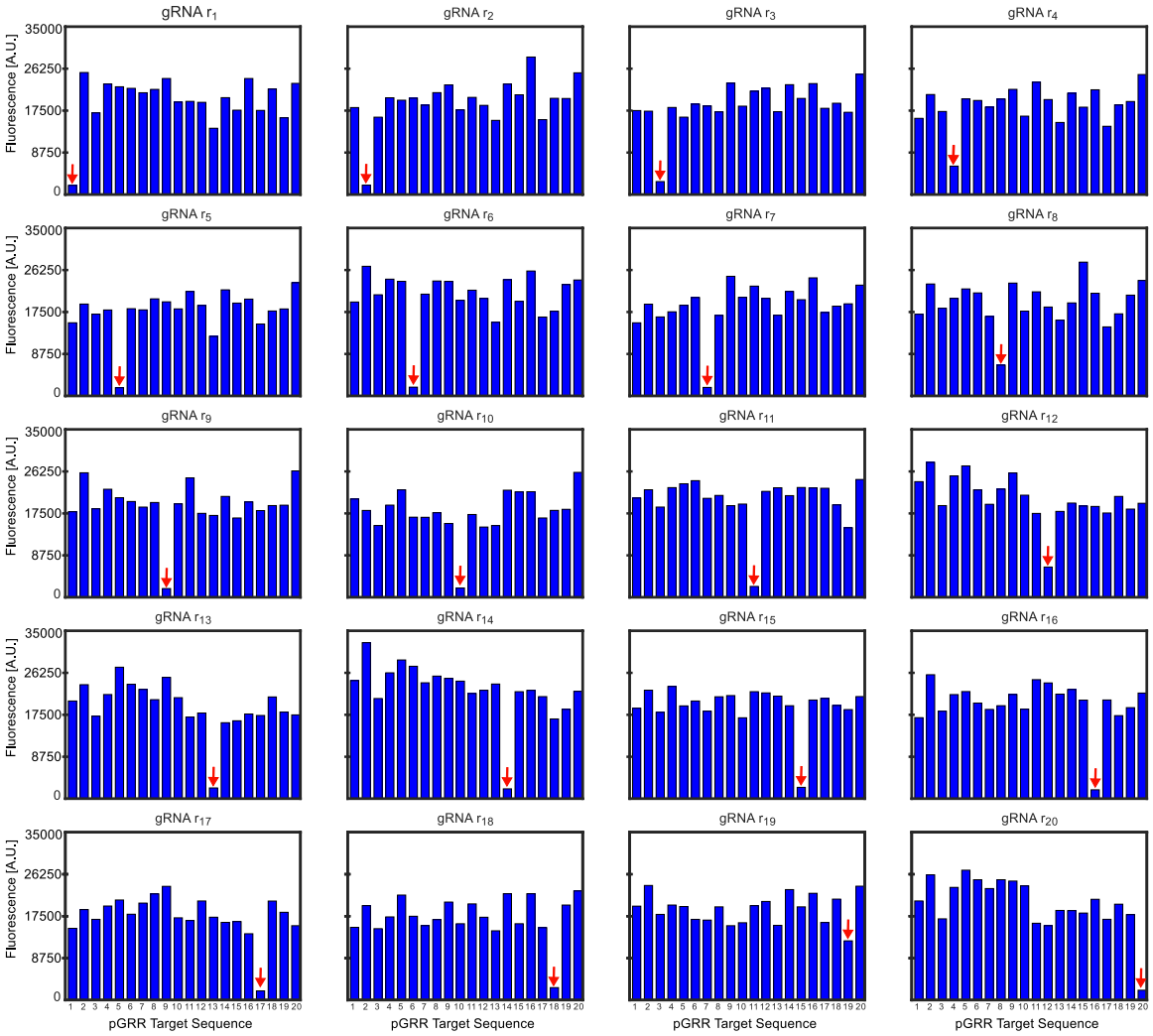


Figure 5.5 | Bar chart of fluorescence values of orthogonality matrix Fluorescence values for all 400 strains in the orthogonality matrix. The strains are segmented by the 20 gRNA target sequences. Promoter target sequence index are in the same order for each subplot. Red arrows indicate a cognate pair of gRNA and pGRR promoter.

Chapter 6

Transcriptional repression via dCas9

6.1 Identifying an effective dCas9-repression domain fusion protein

A key part of engineering digital logic circuits is to ensure that component gates have digital responses. In the of our NOR gates, this means a clear separation between the ON and OFF states. To achieve a clear OFF state, we sought to repress the output pGRR promoters as strongly and consistently as possible. In this chapter, we will review the differences between repression via steric hindrance caused by *dCas9* and repression domain mediated repression via *dCas9*. We describe the repression domains tested in our NOR gate context and show the reasoning behind our choice of *Mxi1*. Finally, we comment about the importance of transcriptional leak and how *Mxi1* reduces transcriptional leak.

It has been shown that translocating a *dCas9*-gRNA complex to promoter DNA can down regulate transcription of the targeted promoter[22,24,41,59]. It is hypothesized that this transcriptional repression is due to steric hindrance of RNA polymerase's progression along the promoter and gene sequences. This steric hindrance, however has varying degrees of repression based on what portion of the promoter sequence is bound by *dCas9* in eukaryotic cells[41]. For our NOR gates we hoped to gain consistent repression via *dCas9* when bound to the two, programmable target site in the pGRR promoter. To accomplish this, we fused several known transcriptional repression domains to the c-terminus of *dCas9* and compared their effects (Figure 6.1).

The repression domains we chose to screen came for a range of organisms including S. cerevisiae, Arabidopsis Thaliana and humans. The three repression domains screened from *S. cerevisiae* are TUP1[49], a general repressor of transcription and associated with chromatin remodeling, XTC1[85], a protein known

to interact with the RNA polymerase II holoenzyme and GAL80[86], a transcriptional regulator involved in regulation of galactose metabolism genes. From Arabidopsis LUG[50] (LEUNIG) a protein associated with the corepressor complex and TPLRD1 a subdomain of the TOPLESS[87] gene which is part of the corepressor complex. Finally, from humans we tested Krüppel associated box (KRAB) [88], a protein that promotes heterochromatin and *Mix1*[59]. These *dCas9* repression domain fusions were all tested in the presence of a gRNA targeting position 1 of a pGRR promoter driving GFP production.

A range of repression levels were observed with different fused domains, however *Mxi1* was clearly the most effective in repressing pGRR activity. *Mxi1* showed strong repressive activity with multiple gRNA binding sequences, repressing a pGRR promoter to near or at the level of *S. cerevisiae* autofluorescence with sixteen of twenty guide sequences developed for the NOR gates (Figure 5.4).

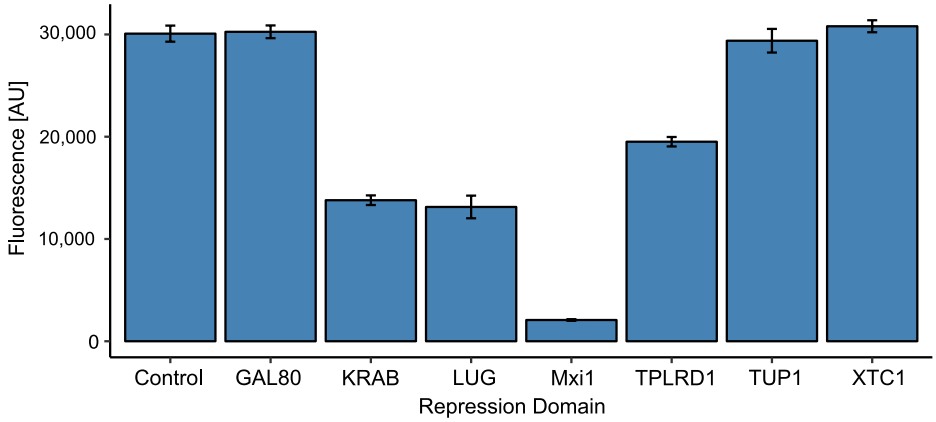


Figure 6.1 | Repression Domain Comparison Protein fusions of *dCas9* and a panel of repression domains that have been shown to function in yeast were compared for repression level. Repression domain fusions were expressed constitutively, along with a pGRR promoter driving GFP and a constitutively expressed cognate RGR. The control strain contains a pGRR promoter driving GFP alone. The repression domains tested are *GAL80*[86], *KRAB*[88], *LUG*[50], *Mxi1*[59], *TPLRD1*[87], *TUP1*[49] and *XTC1*[85]. *Mxi1* shows the greatest amount of repression. Error bars are standard deviation of fluorescence measurements from three biological replicates collected during one experimental run.

6.2 dCas9-Mxi1 mediated repression vs. dCas9 alone

Additionally, *Mxi1* showed more consistent repression at multiple positions on the pGRR promoter than repression via steric hindrance of *dCas9* alone. In Figure 6.2 and 6.3, we show that in three separate sites on the pGRR promoter the *dCas9-Mxi1* repression is lower than *dCas9* alone. We quantified the amount of transcriptional leak at each of these positions with a parameter *L,* defined as the transcriptional leak as a percentage of the maximal expression of a given gate when maximally repressed. This parameter is discussed further in the mathematical modeling section in the Chapter 7 (7.3). Using the model to fit parameter *L* we found that at all three positions the predicted *L* value is as small or smaller for *dCas9-Mix1* than for steric repression. To account for the response curve of the inducible promoter driving the gRNAs in these experiments three more plots were generated comparing *dCas9* and *dCas9-Mxi1* repression as a function of inducible promoter activation driving gRNA (Figure 6.3).

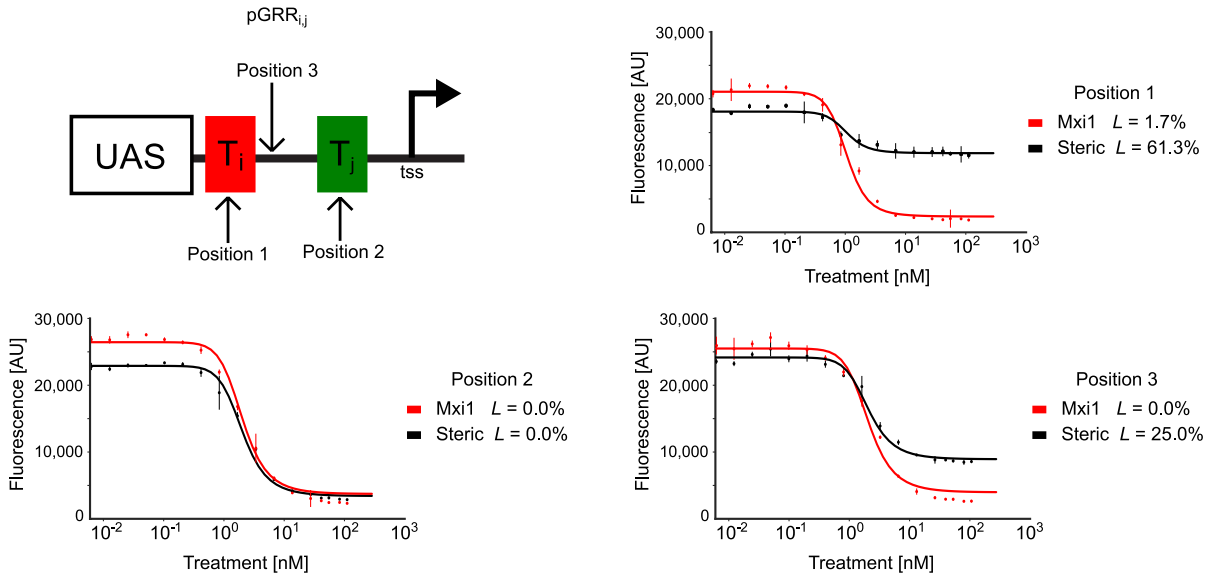


Figure 6.2 | gRNA repression dose response curves Dose response curves are shown for repression via *dCas9-Mxi1* and *dCas9* repression via steric hindrance of pGRR driving *GFP* at three separate positions in the promoter. The three positions are annotated on the pGRR promoter representation. At all 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 standard deviation of three biological replicates measured over three separate experiments.

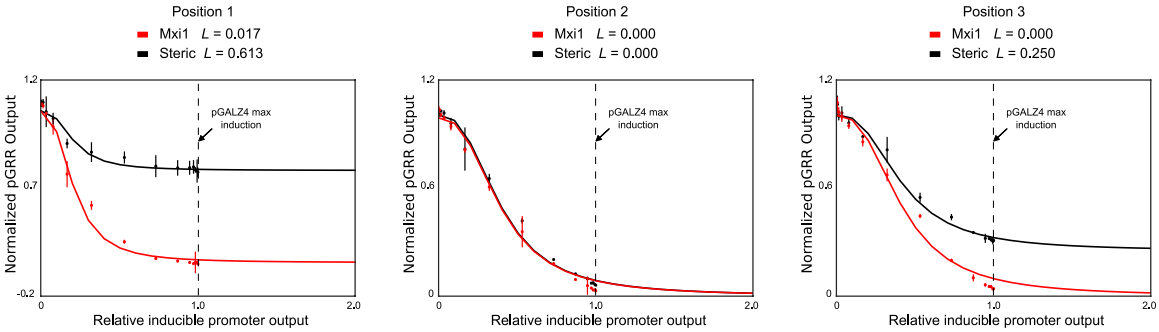


Figure 6.3 | Alternative *dCas9-Mxi1* vs. *dCas9* repression comparison dose response curves Alternative plots of the *dCas9-Mxi1* vs. *dCas9* response curves from Figure 6.2 are shown as a function of relative pGALZ4 inducible promoter output. A model fit of the pGALZ4 inducible promoter dose response 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*.

Chapter 7

Logic Circuits

7.1 2 input 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 7.1a-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 5.4). 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 7.1).

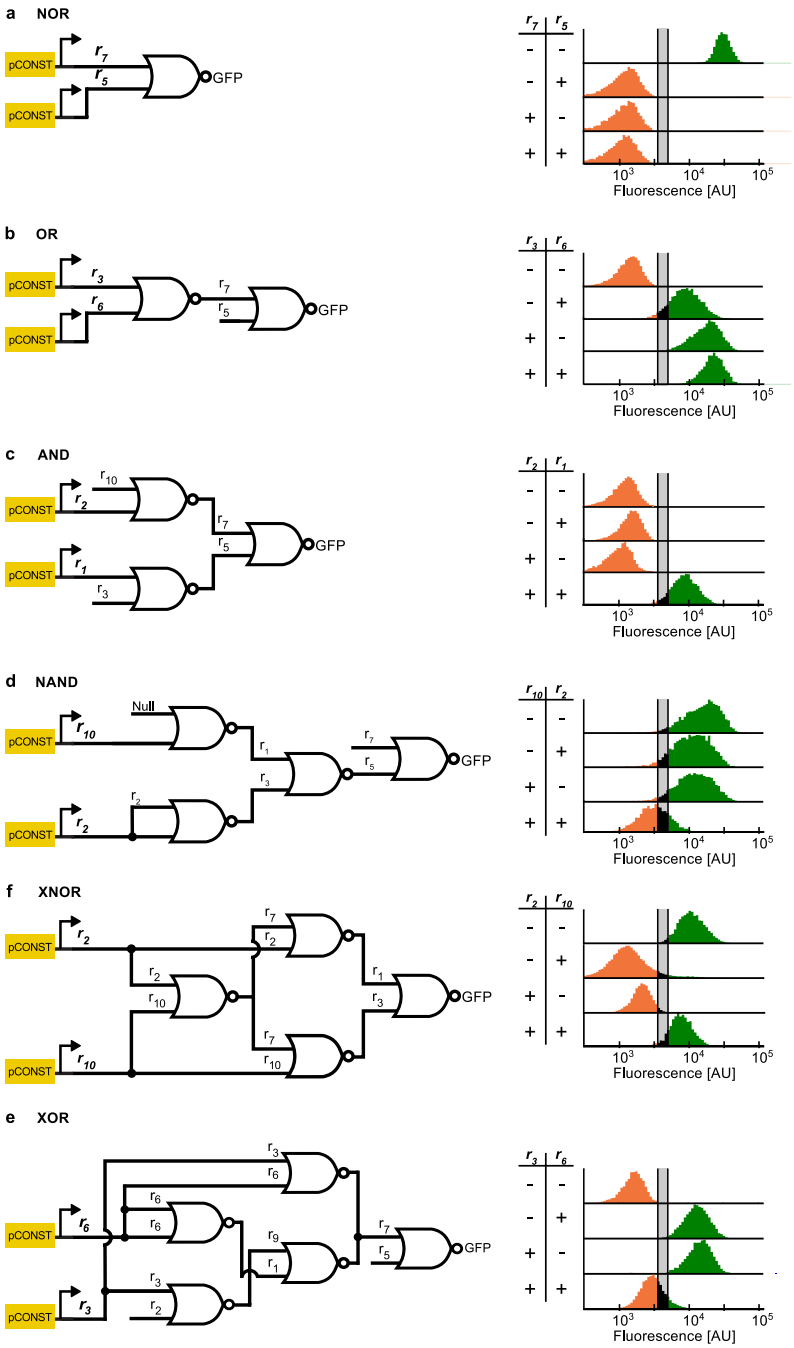


Figure 7.1 | NOR Gate-based logic circuits a-f Six different two-input logic circuits constructed by interconnecting NOR gates. For each of the four input possibilities (- -, - +, + -, + +), a distinct strain was constructed with the corresponding inputs expressed off of constitutive promoters (for logical +), 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 Table 7.2.

Table 7.1 | pCONST promoter table This table lists all constitutive promoter inputs for circuits built in this work.

Circuit	**pConst**
Orthogonality Matrix Strains	pADH1:RGR$_i$
Figure 7.1 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$_3$

We observed fluorescence intensity differences in the digital ON and OFF states in various circuits. In order 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 7.2). 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 7.2.

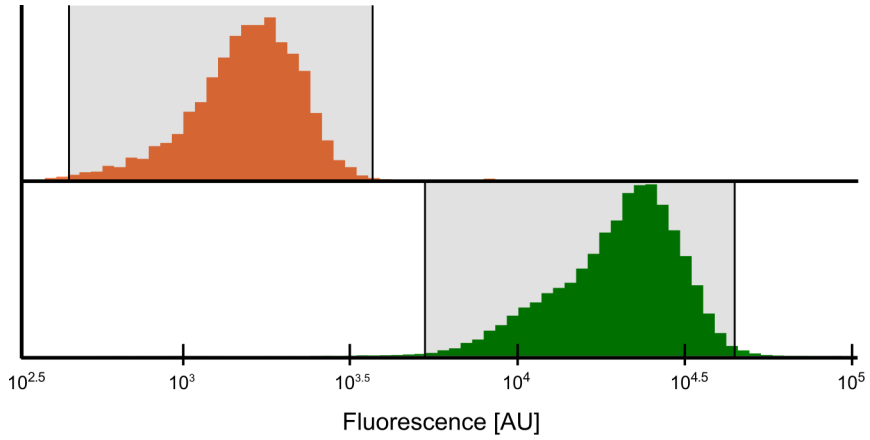


Figure 7.2 | ON OFF and Undefined fluorescence intervals The ON and OFF histograms used to specify the fluorescence intervals for the circuits are shown with their middle 99% mass ranges. The histograms were generated by measuring fluorescence of a set of strains with pGRR promoters driving *GFP* with 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 5.3 and 5.4). 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.

Table 7.2 | 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 Figures 7.1, 7.4, 7.5 and 7.6. 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

Circuits containing different NOR gate variants can exhibit a range of behaviors. For example, 15 versions of the XOR, from Figure 7.1, constructed using different NOR gates exhibited a range of performance (Figure 7.3). 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.

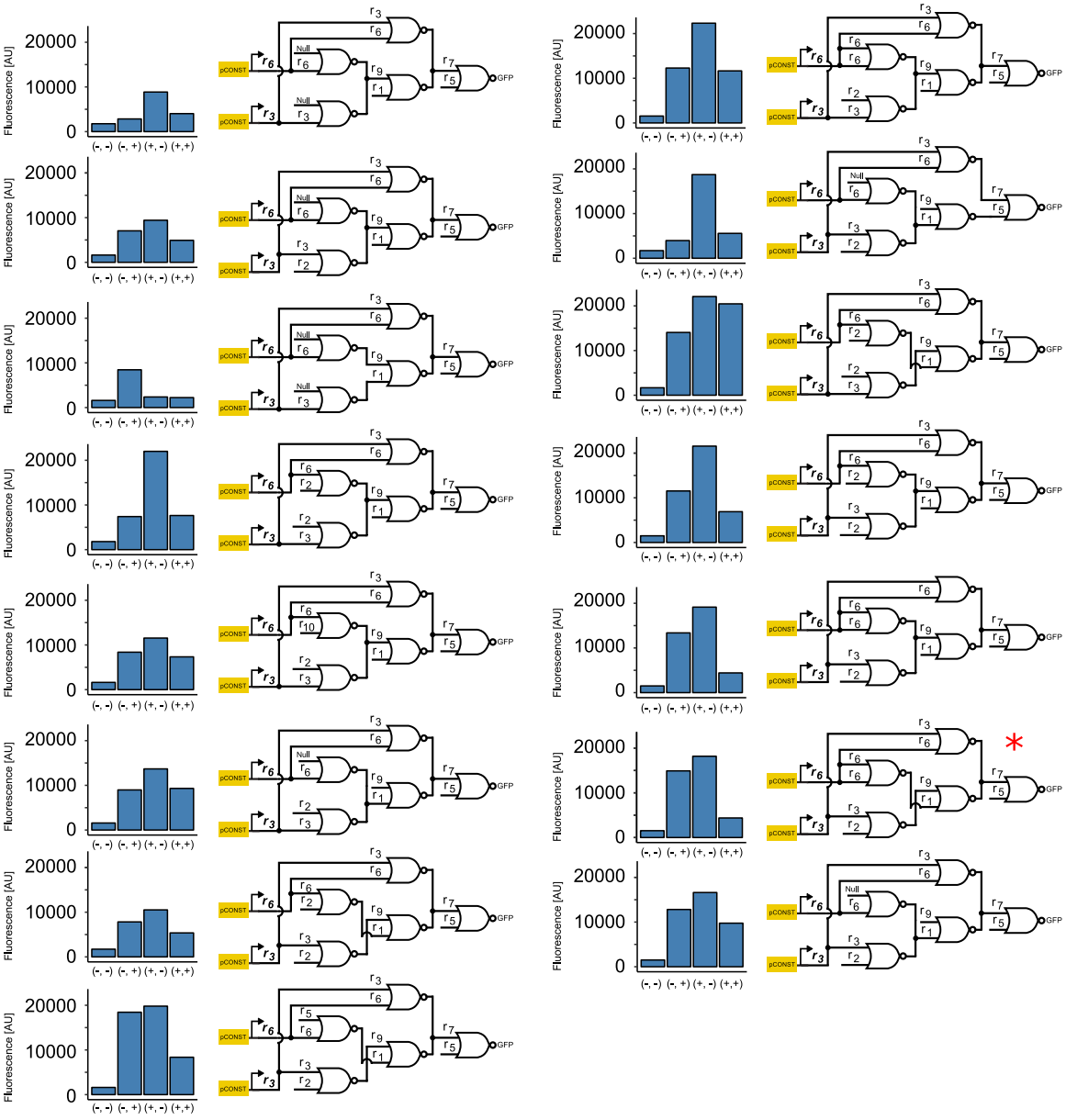


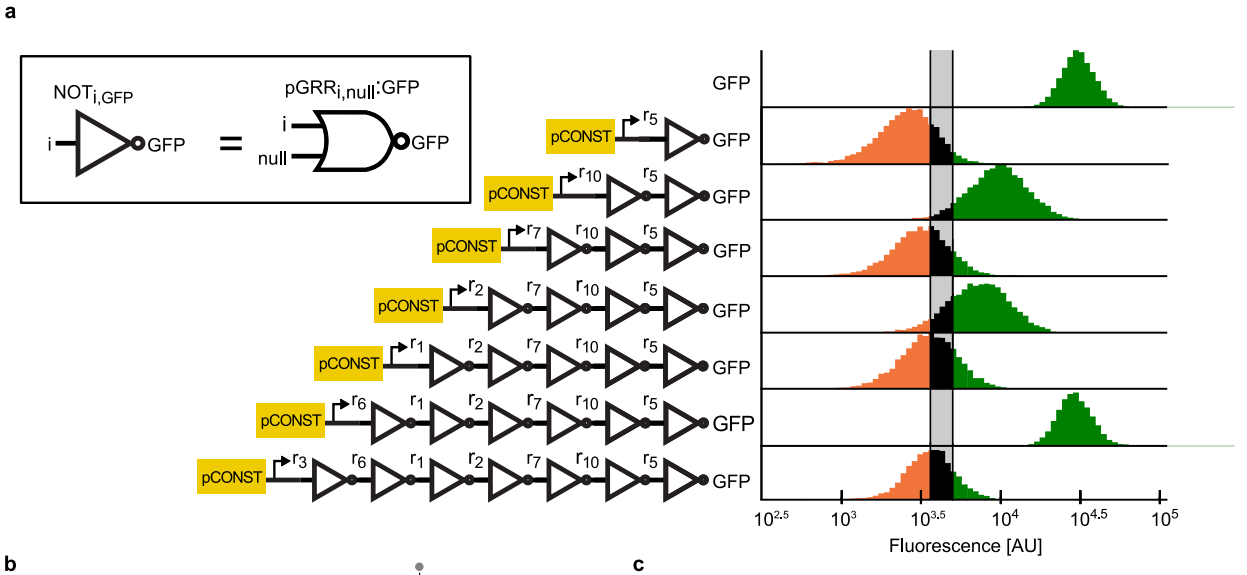
Figure 7.3 | XOR circuit performance variation Fluorescence values for 15 different XOR circuit architectures containing different NOR gates exhibit variation in their output states. This illustrates the impact the performance of the component NOR gates have on overall circuit performance. The red asterisk indicates the XOR architecture that appears in Figure 7.1e. Fluorescence values represent one measurement from cells in log phase.

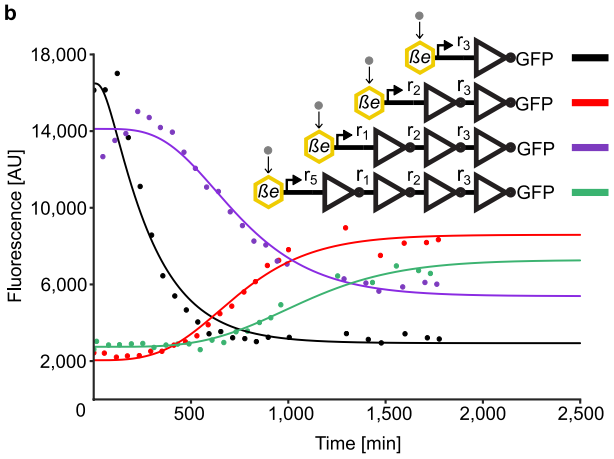
7.2 Cascade circuits

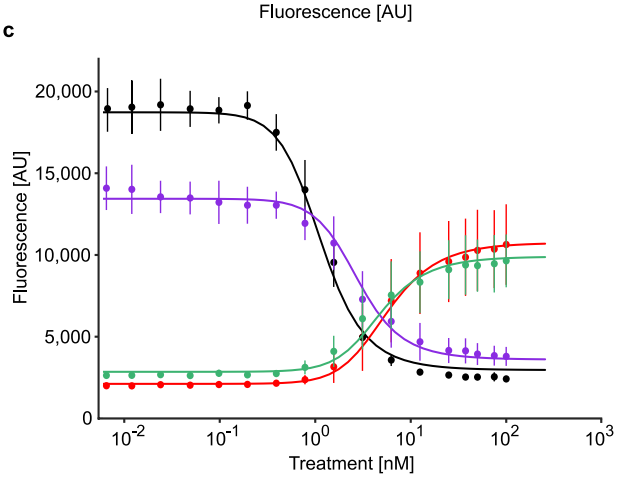
To test the limits of size and complexity our NOR gate circuits can achieve inverter cascades of depth one through seven were composed with NOT gates (Figure 7.4a). The cascade of depth D was made by the addition of a NOT gate to repress the input stage of the depth $D-1$ cascade. Each successive addition of a NOT gate inverter resulted in switching the behavior 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 behavior 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 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 behavior was also observed for other repression cascades that were constructed (Figure 7.5). Alternative versions of 6 gRNA cascades were constructed and showed variability in their levels of ON (Figure 7.6).

To investigate the temporal characteristics of the inverter cascades, we analyzed the kinetics of cascades of depth one through four. A *β*-estradiol inducible promoter[89] was used to activate transcription of the input gRNA and *GFP* expression was periodically measured over the course of about 30 hours of log phase growth (Figure 7.5b). With increasing cascade depth, a clear delay in output response was evident, with the cascades reaching half-maximal expression at 4.1±0.5 hours, 10.8±1.0 hours, 12.0±1.2 hours and 17.8±1.0 hours (residual standard 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 7.5c). 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 7.5c), suggesting a limit to the possible depth of the cascades.







d

$$\dot{r_D} = b\left(\frac{V\left(\frac{u}{K}\right)^{n_u}}{1+\left(\frac{u}{K}\right)^{n_u}} - r_D\right)$$

$$\dot{r_d} = b\left(\frac{v_d(1-L)}{1+\left(\frac{r_{d+1}}{k_d}\right)^{n}} + Lv_d - r_d\right)$$

$$d \in \{1, ..., D-1\}$$

$$\dot{G} = B\left(\frac{1-L}{1+{r_1}^{n}} + L - G\right)$$

u	Beta Estradiol input
d	d^{th} layer
D	Cascade length
r_d	gRNA signal at d^{th} layer
v_d	Expression strength at d^{th} layer
k_d	Repression strength at d^{th} layer
n	Cooperativity coefficient
b	Degradation/dilution of gRNA signal
B	Degradation/dilution of GFP
G	Arbitrary fluorescence units
L	Transcriptional leak at maximal repression
r_D	Input gRNA signal
V	Max expression level of input promtoer
K	Michaelis constant for input promoter

Figure 7.4 | Repression cascades characterization a Repression cascades of one to seven gRNA. 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 7.2. **b** Temporal dynamics for cascades of one to four gRNA. Expression of the input gRNA was induced with Beta 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 standard deviation 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.

Table 7.3 | 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 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 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 $(gates^2+connections^2)^{1/2}$	Functionally complete parts?	Medium
Cascade Circuit	**7**	**6**	**1**	**9.22**	**Yes**	***S. cerevisiae***
Nielsen et al. 2016	7	6	3	9.22	Yes	*E. coli*
Qian et al. 2011	6	5	4	7.81	Yes	*In vitro*
XOR circuit	**5**	**4**	**2**	**6.40**	**Yes**	***S. cerevisiae***
Xie et al. 2011	5	4	6	6.40	No	Mammalian
Auslander et al. 2012	5	4	2	6.40	No	Mammalian
Regot et al. 2011	5	3	2	5.83	Yes	Multicellular *S. cerevisiae*
Nissim et al. 2014	5	3	1	5.83	No	Mammalian
Stanton et al. 2014	4	3	2	5	Yes	*E. coli*
Nielsen et al. 2014	3	2	2	3.61	Yes	*E. coli*
Kiani et al. 2014	2	2	1	2.83	No	Mammalian

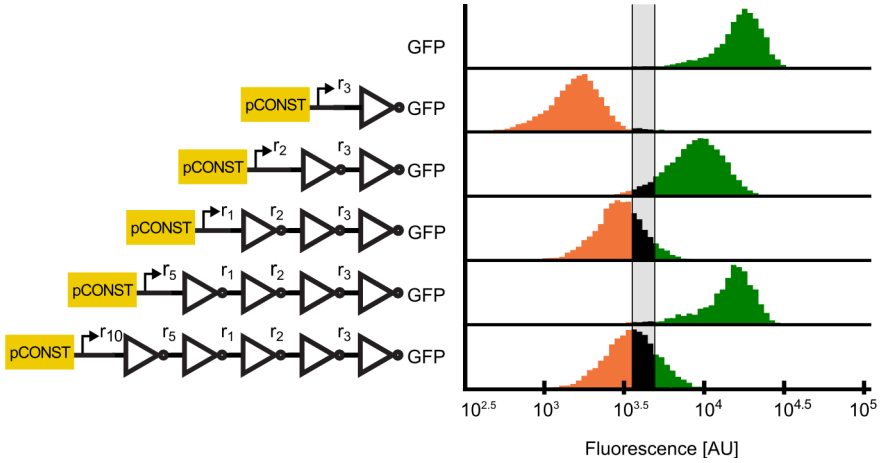


Figure 7.5 | Additional repression cascades A 5-layer repression cascade. The histograms represent population fraction from three different biological replicates measured during a single experiment. Generally, additional layers added to the cascade decrease ON state fluorescence values and OFF state fluorescence values increase.

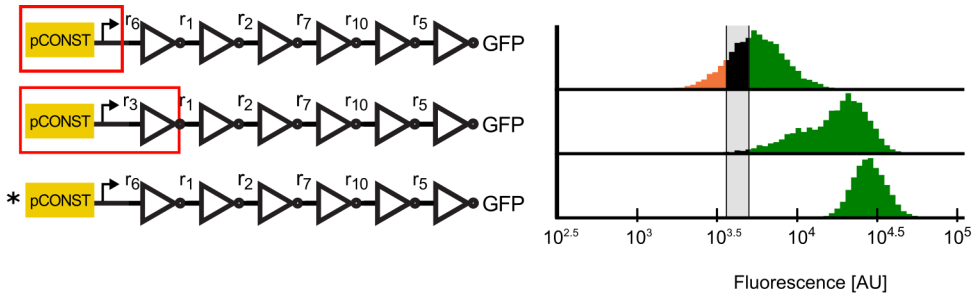


Figure 7.6 | Six layer cascade comparison The fluorescence levels of two different 6 gRNA cascades are compared to the 6 gRNA cascade from Figure 7.4a, indicated by a *. The histograms represent population fraction from three different biological replicates measured during a single experiment. Differences 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-r_6. In addition, the promoter expressing gRNA-r_1 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.

7.3 Mathematical Modeling

A kinetic model was constructed to capture the behavior 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 and degradation/dilution of gRNA-*dCas9-Mxi1* signals respectively (Figure 7.4d). The steady state dose response and kinetic time course for inducible cascade data were both fit to the model (Figure 7.4b-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 7.4b), 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 7.4). The fitting results were found to correlate well with the experimental data. The measured ~18% standard deviation from the mean for the promoter strength values matches well with the ~24% standard deviation from the mean of the promoter strength parameters, v_d (Figure 7.8).

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±0.1% (s.d.), which is equivalent to the production of roughly one transcript every five to ten cell divisions. The reported value of *L* was calculated as the average of the predicted transcriptional leak from the model fits from Figure 6.2. To demonstrate the ability of *dCas9-Mxi1* to decrease transcriptional leak compared to 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 6.2). 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[41]. 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, that *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 6.3.

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 re-sampled 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 hours (s.e.m.) minutes layer^{-1} (Figure 7.7a), consistent with our experimental results. Response delay was found to depend primarily on the degradation/dilution rate *b* of gRNA-*dCas9-Mxi1* (Figure 7.8), which 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 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, δ, representing the percent loss in dynamic range per each additional layer (Figure 7.7b).

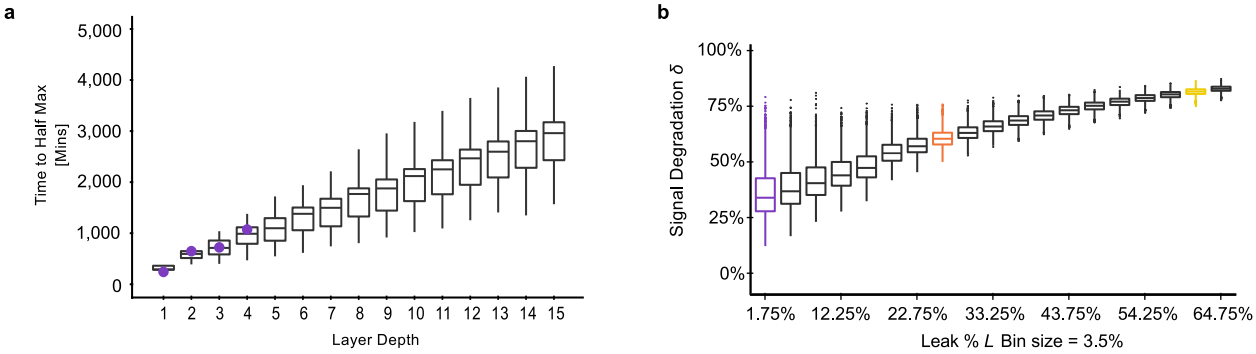


Figure 7.7 | Model predictions and analysis of repression cascades a Simulations of time to half-maximal response using the model. Increasingly layered cascades show a positive linear relationship between circuit time to half-maximal response and circuit depth, with a slope of 184.9±0.2 (s.e.m.) mins layer^{-1}. The first four data points highlighted in purple are experimental data from Figure 7.4b. **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,* $L = 0.6\pm0.1\%$ (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 6.2 of position 1, L = 25.0%, and position 3, L = 61.3%, respectively.

Signal degradation was found to be largely dependent on the transcriptional leak parameter, L (Figure 7.7b and Figure 7.8). As leak increases, δ, on average, increases. At values of $L > 80\%$, the median value of δ 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, $L = 0.6\pm0.1\%$ (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.

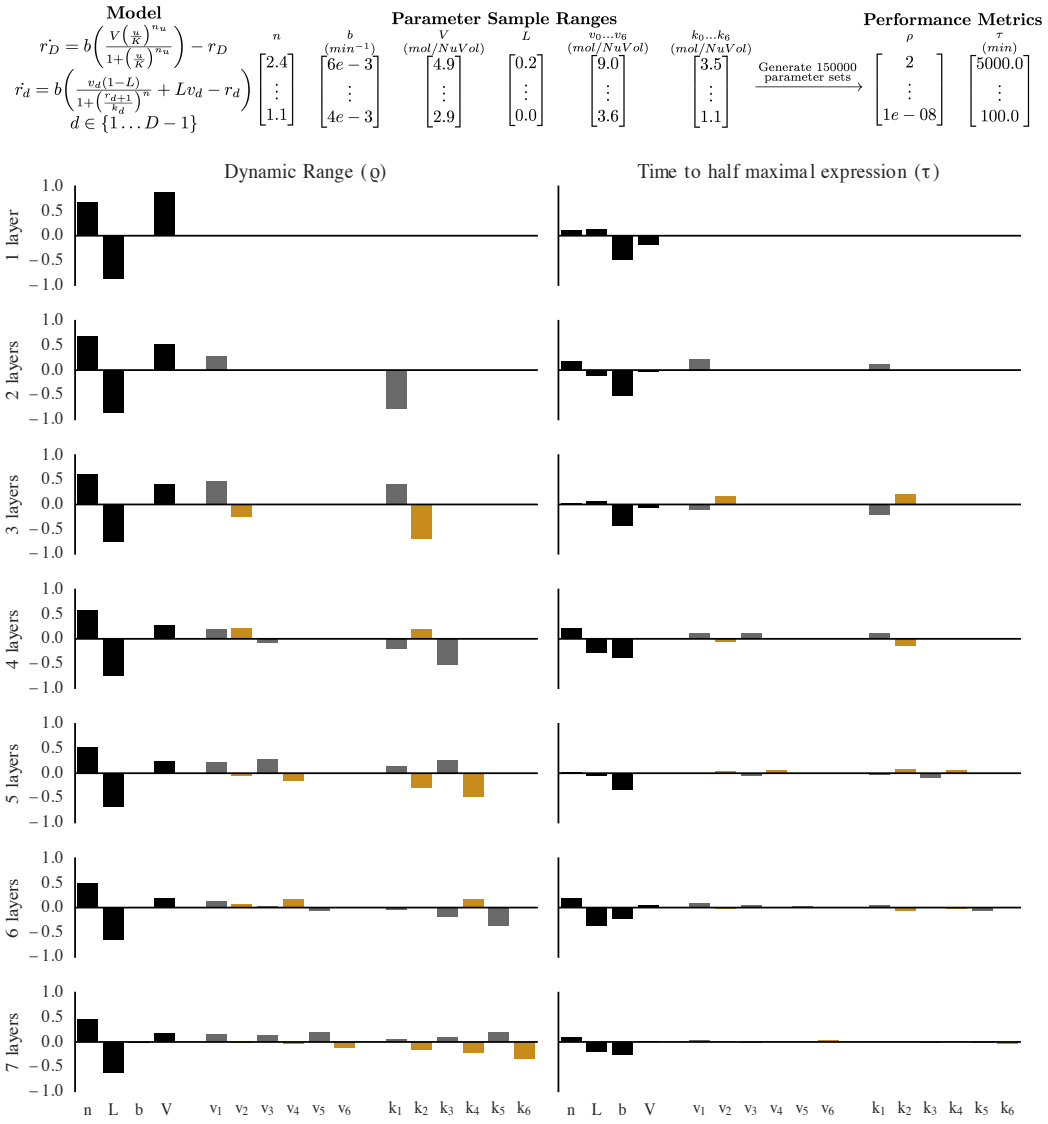


Figure 7.8 | Model parameter sensitivity a 150,000 parameter sets were resampled from a uniform distribution over the intervals shown and applied to our repression cascade model (see methods). **b** Partial rank correlation coefficient (PRCC) was used to determine the contribution of each parameter has 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.

Table 7.4 | Parameter fit values A 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[90] and the published dissociation constant of ~1.2nM for *Cas9* binding to its cognate site[91], bounds for parameters during optimization were selected based on estimates of transcription rates[19] and protein degradation rates in yeast[92]. 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*

Chapter 8

Attempts to implement bistability with the NOR gates

8.1 Building an inducible bistable set-reset latch prototype

Multistable processes govern many biological phenomenon such as sporulation of *Bacillus subtilis*[93], programmed apoptosis[94], and stem cell differentiation[95]. Engineered synthetic multistable biological circuits have been generated but are either small scale[96] or are irreversible[67]. Building up the size and complexity of synthetic multistable circuits could allow for precise control over useful biological functions.

In electronics, the set reset latch is the one of the simplest bistable elements that can be implemented with NOR gates[69]. The set reset latch takes in two signals as inputs, “set” and “reset” and has two stable states. The latch is constructed with two interconnecting NOR gates (Figure 8.1). Due to the simplicity of its topology we used the set reset latch as an initial attempt to implement bistability using our NOR gates. In this chapter, we will detail all the experimental attempts to build a bistable latch.

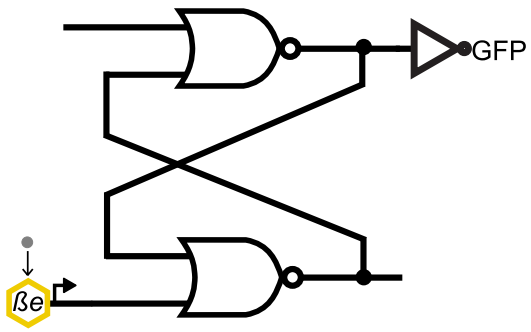
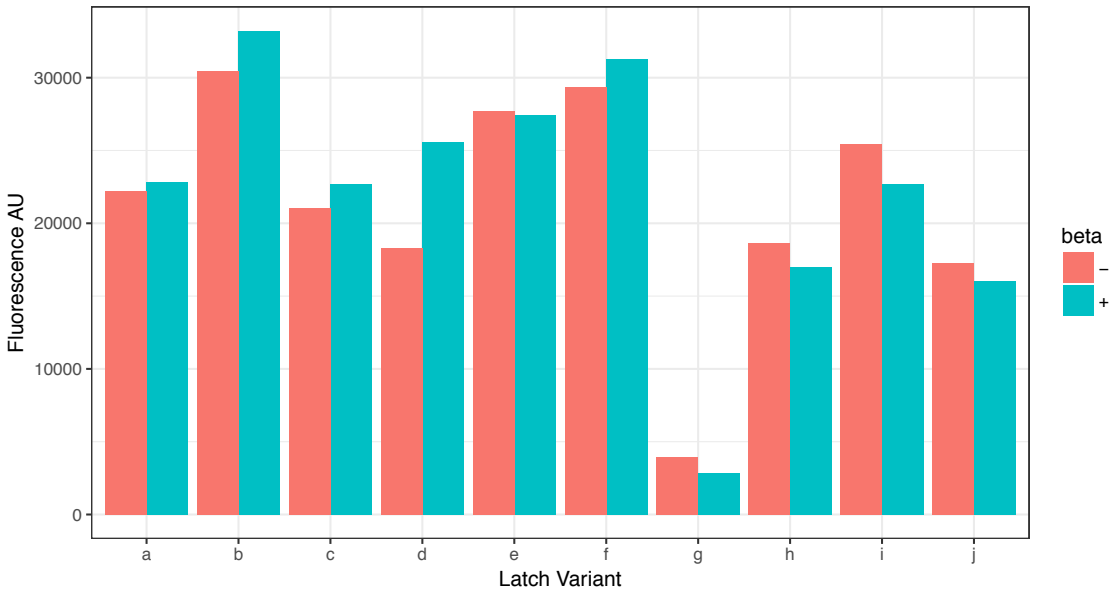


Figure 8.1 | Inducible latch prototype topology Inducible latch prototype topology. The latch is composed of two interconnected NOR gates, an inducible gRNA construct and a downstream NOT gate driving GFP expression.

The initial set of set-reset latches were built as shown in Figure 8.1. These circuits were constructed with NOR gate pairs that had the shown topology. One side of the latch would be induced with a beta estradiol promoter driving one of the input gRNA. We hypothesized that the latch would be in one of its two states, expressing one of the two output gRNAs and either repressing or allowing for expressing of a NOT gate promoter driving GFP. If the circuit was capable of bistable behavior, inducing one of the inputs should cause a drop or an increase in fluorescence of the cells, this fluorescence change should then persist after the inducer was removed from the system.

As an initial test of possible bistability of this architecture several latches were built and screened for their ability to change fluorescence level when under constant induction of the input gRNA. Because we could not predict what state the latch would be in based on the NOR gates chosen some circuit topologies would, by nature of the state they settled into, not yield a state change in the latch output upon induction even if the circuit were functioning as desired. Unfortunately, this led to constructing some circuits whose functionality would not be testable by inducing one input gRNA. Despite this caveat, we carried out these experiments with ten different latch pairs. The results are shown in Figure 8.2, along with a table describing the strain, and its expected and observed behavior.



Latch variant	Starting fluorescence level	Expected induced fluorescence level	Observed induced fluorescence level
a	ON	OFF	ON
b	ON	ON	ON
c	ON	ON	ON
d	ON	OFF	ON
e	ON	ON	ON
f	ON	OFF	ON
g	OFF	OFF	OFF
h	ON	OFF	ON
i	ON	OFF	ON
j	ON	OFF	ON

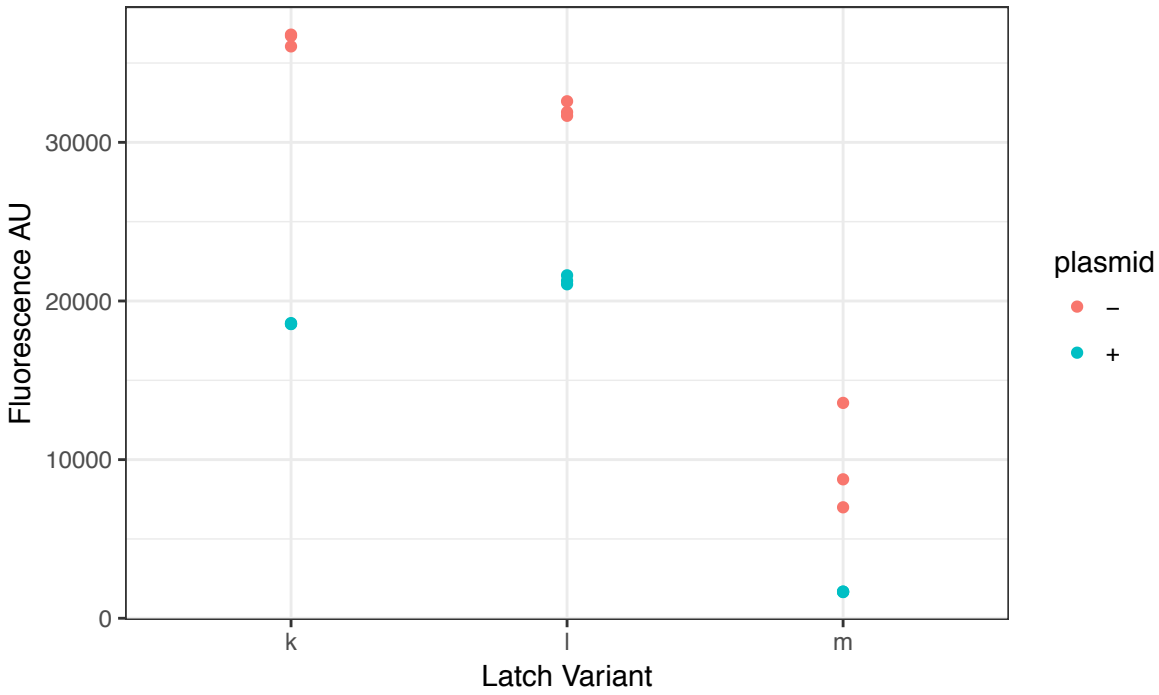
Figure 8.2 | Induction of input gRNA via beta estradiol in a set of latch circuit variants Latch variants a-j are shown in the presence and absence of 100nm beta estradiol. The Bar chat displays one biological replicate from log phase of the latch strains. A table including the desired fluorescence changes upon induction. A full strain information table of the latch circuits is included in this chapter.

Of the ten latch variants created, none showed the desired behavior. The experimental plan for the latches was to induce a state switch and then wash away the inducer to see if the state was maintained. Unfortunately, none of the latches showed a state switch under induction. The data shows that circuits are extremely stable in their initial state before induction.

8.2 Attempts to force latch state change

One hypothesis as to why we were unable to induce a state change in the latch circuits is that the inducible gRNA was not prevalent enough to change the state. To increase the level of input gRNA in the latches strains that contained 2μ plasmids expressing gRNA and strains that expressed input gRNAs. The 2μ plasmids are multi-copy and range from 40-80 copies per cell[97]. This high copy number yields high expression of genes coded on the plasmid. We hypothesized that this high expression of the input gRNA could force the circuit into another state, and if the 2μ plasmid alone was not enough we could further increase the level of input gRNA via beta estradiol induction.

To test our 2μ plasmid hypothesis, three inducible latch variants, that were found to not be induced to another state by beta estradiol induction (data not shown), were transformed with 2μ plasmids that constitutively expressed the latch's input gRNA. As shown in Figure 8.3, while there may be some movement in strain "l" and strain "k" towards the correct state, there is not a decrease in fluorescence large enough to push the latch from the ON state to the OFF state. The other latch, "m" transformed with an input gRNA expressing 2μ plasmid shows some movement further toward the OFF state.



Latch variant	Starting fluorescence level	Expected induced fluorescence level	Observed induced fluorescence level
k	ON	OFF	ON
l	ON	OFF	ON
m	OFF	OFF	OFF

Figure 8.3 | Effect of 2μ plasmid expression of input gRNA in a set of latch circuit variants Latch variants k-m are shown in the presence and absence of a 2μ plasmid expressing an input gRNA. Each dot on the plot represents a biological replicate at log phase of the strains. A table including the desired fluorescence changes upon induction with the 2μ plasmid. A full strain information table of the latch circuits is included in this chapter in Table 8.1.

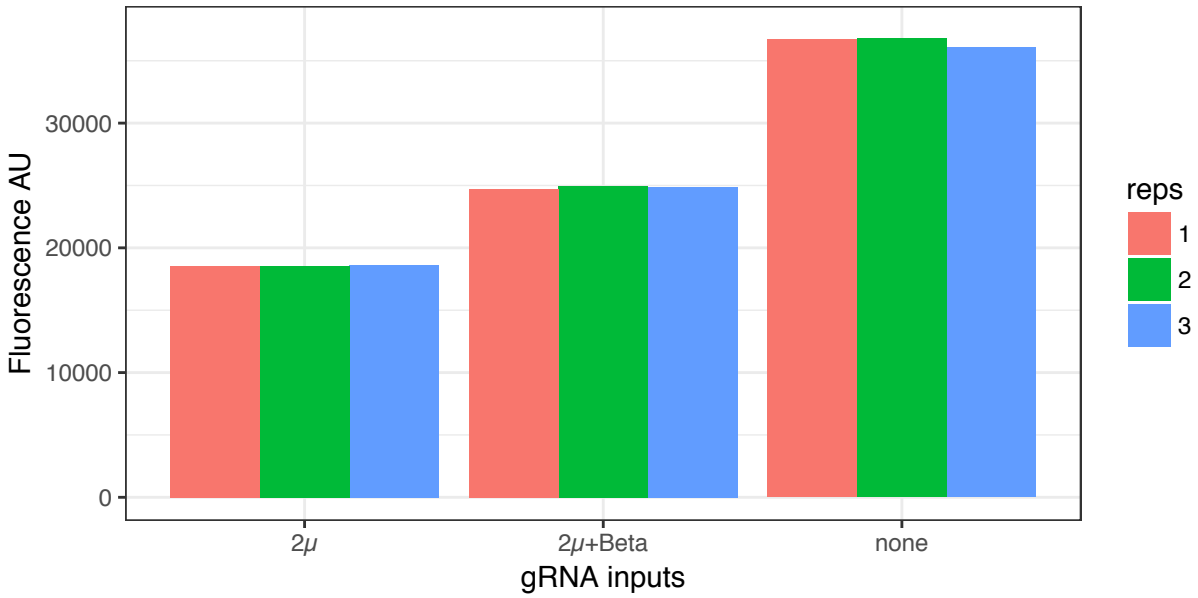


Figure 8.4 | Latch variant induced with both beta estradiol and 2μ gRNA expression Latch variant k is shown induced with a 2μ plasmid expressing a gRNA and another strain shown with both a 2μ and an inducible cassette expressing the input gRNA. The desired behavior of k is for the fluorescence to go from ON to OFF levels after induction. The replicates for each gRNA input bar are technical replicates from the same culture in log phase on the same day.

To further increase the level of input gRNA for latch “k” the inducible promoter driving the input gRNA within the strain was activated in the presence of a 2μ and compared to the strain with and without a 2μ plasmid. The induction of the beta estradiol responsive promoter did not further decrease the fluorescent output of the latch. In fact, the induction shows a higher fluorescence level likely due to a decreased cell growth rate.

The experiments with the latches described in this chapter show our failed attempts at building a bistable circuit with our NOR gates. Not only were we unable to show bistable behavior, we were also unable to show the ability of a latch to switch states by introducing input gRNA into the system. A handful of strains show movement in the fluorescence level toward the desired state. This seems to indicate that if there were enough input gRNA the state of the circuit could be switched. Reaching the level of gRNA required to switch these circuits could be extremely difficult or altogether impossible. If both circuit states would

be stable, the seemingly necessary high level of expression for a state switch renders the latches incompatible with existing circuits shown in this document. If there exists a functional latch that can be built with the existing NOR gate framework a method to build and test many in a high throughput manner, such as the one described in Chapter 10, would be necessary to identify.

Table 8.1 | Latch variant strain list

Latch variant	Plasmids incorporated	Klavins lab aquarium sample number
a	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7, pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r11-yeGFP, pMOD8-pGALZ4-URGR-r10, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-URGR-r11	13487
b	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7, pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r11-yeGFP, pMOD8-pGALZ4-URGR-r2, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-r10	13488
c	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7, pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r2-yeGFP, pMOD8-pGALZ4-URGR-r10, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-URGR-r11	13489
d	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7, pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r11-yeGFP, pMOD8-pGALZ4-URGR-r2, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-URGR-r11	13490
e	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7, pMOD6-pGRR-r10-yeGFP, pMOD-dcas9-mxi1, pMOD-HOKan-pACT1-ZEV4, pMOD8-pGALZ4-URGR-r2, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-r10	13491
f	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7, pMOD6-pGRR-r10-yeGFP, pMOD-dcas9-mxi1, pMOD-HOKan-pACT1-ZEV4, pMOD8-pGALZ4-URGR-r2, pMOD-LTR2-Bleo-pGRR-r10r2-RGR-URGR-F1	13492
g	pMOD-LTR1-Nat-pGRR-r11r10-RGR-r7,	13493

	pMOD6-pGRR-r7-yeGFP, pMOD-dcas9-mxi1, pMOD-HOKan-pACT1-ZEV4, pMOD8-pGALZ4-URGR-r2, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-URGR-r11	
h	pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r2-yeGFP, pMOD8-pGALZ4-URGR-r10, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-URGR-r11, pMOD-LTR1-Nat-pGRR-r11r10-RGR-W8	14004
i	pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r2-yeGFP, pMOD8-pGALZ4-URGR-r7, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-r10, pMOD-LTR1-Nat-pGRR-r11r10-RGR-r2	14003
j	pMODKan-HO-pACT1-ZEV4, pMOD4G-dcas9-mxi1, pMOD6-pGRR-r11-yeGFP, pMOD8-pGALZ4-URGR-r10, pMOD-LTR2-Bleo-pGRR-r7r2-RGR-URGR-r11, pMOD-LTR1-Nat-pGRR-r11r10-RGR-r2	14002
k	pMOD4G-dcas9-mxi1, pMOD-LTR2-BleoMX-pGRR-r9r5-RGR-r6, pMODKan-HO-pACT1-ZEV4, pMOD8-pGALZ4-URGR-r1, pMOD6-pGRR-r6r1-RGR-r5, pMOD-LTR1-NatMX-pGRR-r6-yeGFP	12460
l	pMOD4G-dcas9-mxi1, pMOD6-pGRR-r1r9-RGR-r5, pMOD8-pGALZ4-URGR-r7, pMODKan-HO-pACT1-ZEV4, pMOD6-r5r7-RGR-r9, pMOD-LTR1-NatMX-pGRR-r5-yeGFP	12314
m	pMOD4G-dcas9-mxi1, pMOD6-pGRR-r1r9-RGR-r5, pMOD8-pGALZ4-URGR-r1, pMODKan-HO-pACT1-ZEV4, pMOD6-r5r7-RGR-r9, pMOD-LTR1-NatMX-pGRR-r9-yeGFP	12313

Chapter 9

Cloning via Aquarium

9.1 Protocols developed

A significant portion of the recombinant DNA used in this work was created via the UW BIOFAB[98] and the Aquarium lab automation software. The UW BIOFAB is a University of Washington cost center that performs cloning services, such as plasmid and yeast strain construction, for researchers. Jobs are submitted by users via the Aquarium software in the form of codified cloning instructions. These jobs are carried out through human-in-the-loop automation by undergraduate technicians at UW. The Aquarium software allows users to design custom laboratory workflows that can be carried out by technicians. In this chapter, we will cover the workflows and protocols developed specifically for the development of the NOR gates, the naming scheme associated with the project constructs and an explanation of the specific DNA sequences used to construct the circuits.

An important aspect of the high throughput circuit construction strategy discussed in Chapter 10 is a multiplex transformation strategy. As shown by a recent publication[99], it is possible to integrate multiple double stranded pieces of DNA in a single transformation by leveraging CRISPR-Cas9's targeted endonuclease activity. One of the main workflows the UW BIOFAB offers is yeast transformation using restriction digest linearized DNA. To make the CRISPR-Cas9 multiplex integration protocol compatible with the yeast transformation we developed a DNA combination protocols that allows the user to define a pool of DNA sample types, including plasmids and fragments and their amounts that is routed through the yeast transformation protocol.

The multiplex transformation strategy demonstrated by Horwitz et al. takes advantage of S. cerevisiae DNA gap repair and multi-targeted DNA cleavage to allow for multiple pieces of double stranded donor DNA to be integrated into the yeast genome at once. The basic scheme for this technique is to transform three different types of double stranded DNA into yeast cells at a single time. The three types are plasmid backbone, gRNA cassettes that have plasmid backbone homology and can be gap repaired inside the cell to create functional plasmids and donor DNA that would be homologously recombined into the genome to repair cuts made by Cas9.

All three of these DNA components must be transformed together and an Aquarium task was developed to accomplish this. The task, named “Plasmid Combinings” consists of a single protocol that allows the user to define an array of DNA sample types, the amounts in nanograms of the specific DNA sample type. The task then is run by a technician who pipettes the relevant samples together into a desired plasmid pool stock that is compatible for use in the Yeast Transformation workflow. This workflow was successfully used during the Winter 2017 offering of Laboratory Methods in Synthetic Biology at UW.

9.2 Naming scheme for constructs used throughout this work

Almost all the plasmids constructed for the NOR gate were created using the pMOD cloning system. The pMOD plasmid cloning system was developed by former members of the Klavins lab Kyle Havens and Nick Bolten. The plasmid cloning system allows for modular assembly of gene cassettes using standard linkers between sequence elements in a Gibson assembly reaction[100]. The plasmids are set up to express a single gene cassette. The plasmids are designed for single integration into the genome of the *S. cerevisiae* strain W303 after linearization with the pMEI endonuclease enzyme. The plasmid suite used in this work has eight different genomic integration sites and eight different selection markers, four auxotrophic and four antibiotic.

The standard linker sequences were optimized for Gibson assembly and have names that denote the position in which they appear in the gene cassette. Linker sequence content was designed to have melting temperatures near 50º C, which is the temperature that Gibson assembly. The linker names are Promoter Suffix (PS), Terminator Prefix (TP), Terminator Suffix (TS), Promoter Prefix (PP2), Genomic Homology Region upstream prefix (G1P), Genomic Homology Region suffix (G1S) and Genomic Region downstream suffix (G2S) (put in a table with the sequences). Sequence elements are bordered by these linkers; 5' genomic homology by G1P and G1S, the selection marker gene by G1S and PP2, the promoter by PP2 and PS, the gene coding sequence by PS and TP, the terminator by TP and TS and the 3' genomic homology by TS and G2S. If the selection marker used for a given plasmid is auxotrophic then the 5' genome homology and the selection gene are one and the same and would be boarded by the G1P and PP2 linker.

The eight selection markers consist of four auxotrophic, tryptophan, uracil, histidine and leucine and four antibiotic resistance genes, BleoMX, HygMX, KanMX and NatMX. The eight integration sites include the four auxotrophic marker gene loci, the HO locus and three long terminal repeat regions named LTR1, 2 and 3.

Each portion of a pMOD plasmid's name has a specific meaning. The name takes the following form "pMOD"-"genomic locus"-"selection marker"-"promoter"-"gene". As an example, consider the plasmid "pMOD-LTR1-NatMX-pGRR-r_3-RGR-r_{10}". The genomic locus this plasmid integrates into is LTR1, it expresses the resistance gene NatMX, the gene cassette promoter is pGRR-r_3 which drives the expression of the RGR-r_{10} gene. There are commonly used identifiers for the loci, selection marker genes and promoters for the plasmid names which are listed in the below tables. Note that for auxotrophic markers the selection marker and loci are the same sequence element.

Table 9.1| Plasmid naming scheme explanation

Promoter identifier in plasmid name	Promoter full name
A	pADH1
G	pGPD/pTDH3
$pGRR_{i,j}$	$pGRR_{i,j}$
pGALZ4	pGALZ4
pCYC1	pCYC1

Locus Name	Genomic locus
6	URA
8	HIS
4	TRP
LTR2	LTR2
LTR1	LTR1
LTR3	LTR3
HO	HO
9	LEU

Selection marker name	Marker
6	URA
8	HIS
4	TRP
BleoMX	BleoMX
NatMX	NatMX
HygMX	HygMX
KanMX	KanMX
9	LEU

The 2 gRNA input promoters of the NOR gates have a specific naming scheme that denotes their sequence. The name pGRR stands for promoter gRNA responsive. The pGRR promoter has two programmable target sites flanked by PAM sequences that can bind gRNA. When a pGRR promoter is listed in a plasmid name it is followed by the name of the two target sequences in a 5’ to 3’ orientation toward the beginning of transcription. The guide sequence names and their sequence are shown in the table below. It should be noted that the most 3’ target site on the pGRR promoters has a PAM sequence orientation such that the sequence present there is the reverse compliment of what appears in the table. For the gRNA RGR gene the guide sequence of the RNA is listed after the RGR. A sequence element map for the pGRR promoters is show in Figure 4.2.

Table 9.2 | Guide sequence table This table lists all the guide sequences used in this work. We have included the sequence index for this document as well as the sequence index that appears in the Klavins lab aquarium software inventory system.

gRNA index	Sequence	Klavins Lab Aquarium gRNA index
r_1	GGAACGTGATTGAATAACTT	W5
r_2	ACCAACGCAAAAAGATTTAG	W8
r_3	CATTGCCATACACCTTGAGG	W10
r_4	GAAAATCACAACTCTACTGA	W14
r_5	GAAGTCAGTTGACAGAGTCG	W17
r_6	GTGGTAACTTGCTCCATGTC	W19
r_7	CTTTACGTATAGGTTTAGAG	W20
r_8	CGCATTTCCTATTCAAACTT	W22
r_9	GCAACCCACAAATATCCAGT	W34
r_{10}	GTGACATAAACATTCGACTC	W36
r_{11}	GGGCAAAGAGACGCTTGTCG	F1
r_{12}	GAAGTCATCGCTTCTTGTCG	F2
r_{13}	GAGTTGACAAAGTATAACTT	F3
r_{14}	GAAGTTTCAGAATCTCGACG	F4
r_{15}	GGCTAGGATCCATCTGACTT	F5
r_{16}	GCAACCATAGACTCTCCAGG	F6
r_{17}	ACCACAACTGAGTCGAACCT	F7
r_{18}	GGGTAGCAACACTCGTACTT	F8
r_{19}	GTAAAAGATAACTCTGTTGC	F9
r_{20}	TCTACCCGAGACTCAAACGG	F10
v_1	GTACATACAGTAGGATCCTA	V1
v_2	TTTGGCACTACCGACACGAA	V2
v_3	TGGTCAAAAGTGCGGCTTTC	V3
v_4	CTTTCACAATCTTGACCTGC	V4

Chapter 10

Multiplex Transformation Circuit Construction

10.1 Problem statement

There are two major challenges in creating functional large scale circuits using the CRISPR/dCas9 NOR gates. First, individually Genomically integrating each gate into *S. cerevisiae* is an inherently slow process. With each genomic integration typically taking on the order of three days, circuits with many components will take a significant amount of time just to create one prototype. Second, as we have discussed earlier the NOR gates all have some degree of performance variability, in terms of repression and expression level. The variable performance of the gates can confound circuit function when components are wired together as shown in Figure 7.3. These two challenges have restricted the size of circuit we can implement using our NOR gates and broadly apply to gene circuit construction across the field of synthetic biology.

In general, to overcome these two issues, time consuming construction and component variability when engineering functional genetic circuits synthetic biologists generally employee one of two strategies. One method is to carefully characterize each circuit component's behavior and use the characterization data to determine the optimal circuit designs. The other is to develop a system to rapidly create circuit variants using random combinations of circuit components and then screen the variants for the desired behavior. Both methods have their drawbacks, individual characterization of component libraries requires a large time investment and multiplex construction and screening strategies require the component library to be amenable to high throughput screening, in terms of connectivity and measurability of circuit output.

Here we outline a strategy for multiplex transformation circuit construction and screening using our NOR gates. This approach is designed to use a semi-random one-pot multiplex transformation to overcome the two broad challenges described above. The approach will create a many circuit variants in a single transformation that will be screened for desired output behavior via Fluorescence-activated cell sorting (FACS) or life/death selection.

10.2 Multiplex Transformation

To reduce the lead time to a functional circuit variant with the NOR gates a transformation method that allows multiple components to be incorporated into a circuit in single reaction must be employed. A recent paper outlined a method to incorporate up to eleven genes at a time using a CRISPR-Cas9 based strategy[99]. By integrating in donor DNA and multiple gRNAs into a strain expressing Cas9 the authors showed they could use the Yeast gap repair mechanism to integrate the donor DNA in the loci cleaved by the gRNA-Cas9 complex. This multiplex transformation method could be applied to integrate NOR gate genes into the yeast genome.

An interesting caveat arises with the Cas9 multiplex transformation strategy to build gene circuits with our NOR gates. Because the NOR gates are regulated by *dCas9-Mxi1*, and express gRNA that target downstream NOR gates, there is the possibility of unwanted cleavage of NOR gate promoters if the NOR gate genes are integrated with a Cas9 strategy. Unwanted cleavage in the NOR gate promoters would automatically render any circuit faulty. Three possible solutions to this issue are as follows, first, as shown in a recent publication[101], truncating the guide sequence length of a gRNA from 20 basepairs to 16 basepairs can prevent the ability of Cas9 to cut with a given gRNA. Thus, unwanted cleavage by Cas9 can be avoided by using two separate sets of gRNAs, one truncated set expressed by the NOR gates and another 20 basepairs set for making genomic double stranded cuts with Cas9. Second, rather than using Cas9 to make double

stranded cuts there are several other RNA guided nucleases that could be employed[60,102]. These alternative RNA targetable nucleases have different PAM recognition sequences making them orthogonal to *dCas9*. Finally, by altering the NOR gate promoters to be regulated by a universal activator controlled by a small molecule inducer, such as pGALZ4[89]. This would allow for controlled activation of the circuit after Cas9 has completely been vacated from the cell.

10.3 Mathematical Investigation of multiplex construction strategy

Broadly, the construction scheme involves choosing a desired circuit behavior, mapping out the topologies that yield the desired behavior then combining libraries of NOR gates in a one pot assembly. This one pot assembly will create many different circuit versions on a single piece of double stranded DNA. With *a priori* knowledge of the desired circuit topology the libraries of NOR gates used in the one pot reaction can be chosen to randomly create many versions of a circuit. Once the topology is known, gRNAs are chosen for the “wires” and a pool of NOR gate libraries can be chosen to combine.

A NOR gate library is a pool of $NOR_{i,j,k}$ gates that whose i, j and k are allowed to be fixed or varied across the 20 component library in any combination. For example, the NOR gate library, (r_1, *, *) contains 400 different NOR gate species that all contain the r_1 target site at the c3 position, the NOR gate library (r_3, *, r_4) contains 20 different NOR gates, that express gRNA-r_4 and has the r_3 target sequence at site 1 of the pGRR. NOR gate libraries are made using the 8000 full component NOR gate library as a master template and Gibson assembly of chosen sets of RGR cassettes and promoter binding site fragments. Using primers that bind to specified to guide sequences, subsets of the 8000 full component NOR gate library can be created and wired together in multiplex transformation.

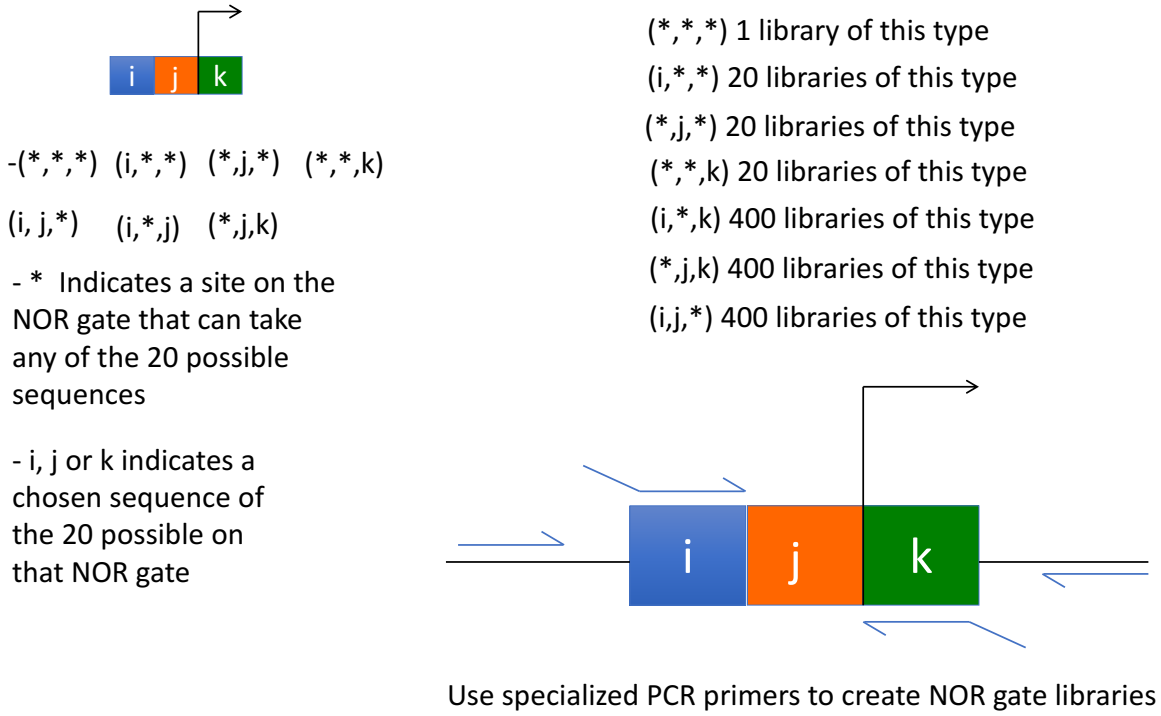


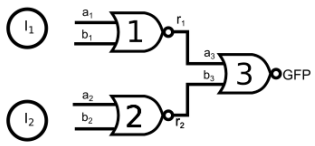
Figure 10.1 | Schematic depiction of a NOR gate library NOR gate libraries with fixed binding sites or gRNA guide sequences can be created using Gibson assembly and PCR.

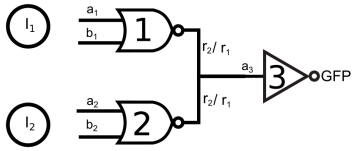
As an example, we will use the construction process of a three-component AND gate. We define our chassis strain as a yeast strain that contains two inducible gRNA cassettes as inputs to the circuits. To support our choice of a semi-random multiplex strategy we will compare the likelihood of a valid circuit topology If a completely random three gate construction strategy were applied with the whole 8000 NOR gate library to the semi random approach. If three gates are chosen at random to be transformed, the probability of an AND topology is (2691400) ÷ (256 x 10^{10}) or about 1 in 9512. With a typical yeast transformation with a µg of DNA yielding 5 x 10^{6} transformants, it follows that about 525 cells per transformation will contain a valid AND topology. This equates to about five percent of the total valid AND topologies possible with the 20 component library.

Considering that the probability of obtaining a valid topology for a circuit falls as the number of circuit components grows with the random strategy and that screening methods for identifying a functional circuit will not be 100% reliable, this random strategy is not likely to be tractable.

In contrast the semi-random strategy yields a much higher probability of a valid AND topology. First, we will identify all topologies of NOR gates that execute logical AND. Each topology has a specific set of gRNA and target site connections in the 20 component library. When the input gRNA and internal wires are fixed, there are two possible topologies that execute the desired function. Once the topologies are fixed members of the 20 component library must be assigned to the fixed connections. For the AND gate one topology contains three unique gRNAs and the other contains four. Next, a pool of gate libraries must be selected that contains the fixed gRNA and target sites for both topologies. The AND has 10 NOR gate libraries that make up its pool. This pool creates 12800 different circuit versions. The fixed gRNA and target sites define a set of constraints that must be satisfied for the finished circuit to have a valid topology. These constraints along with the NOR gate inputs to the reaction define the probability that a given three gate combination will have valid topological connections. For the AND example 9808 of 12800 (76.6%) are logically valid.

The XOR gate can serve as another example. The XOR circuit consists of five components. There are four total topologies that lead to a valid XOR wiring. The XOR pool has eighteen NOR gate libraries in its pool. The pool creates 204800 circuits and 123136 valid XOR topologies. This equates to a 60.13% topology success rate.





If $r_1 \neq r_2$:

$I_1 = a_1$ or b_1, $I_1 \neq a_2$, b_2, a_3, b_3
$I_2 = a_2$ or b_2, $I_2 \neq a_1$, b_1, a_3, b_3
$r_1 = a_3$ or b_3, $r_1 \neq a_1$, b_1, a_2, b_2
$r_2 = a_3$ or b_3, $r_2 \neq a_1$, b_1, a_2, b_2

If $r_1 = r_2$:

$I_1 = a_1$ or b_1, $I_1 \neq a_2$, b_2, a_3, b_3
$I_2 = a_2$ or b_2, $I_2 \neq a_1$, b_1, a_3, b_3
$a_3 \neq I_2$, I_1

Pool of NOR gate libraries that will yield this topology
NOR Gate 1: (*, I_1,r_1) , (I_1,*,r_1) , (*, I_1,r_2) , (I_1,*,r_2)
NOR Gate 2: (*, I_2,r_1) , (I_2,*,r_1) , (*, I_2,r_2) , (I_2,*,r_2)
NOR Gate 3: (r_1, r_2, GFP) and (r_2, r_1, GFP)

If $r_1 \neq r_2$:

NOR Gate 1: (*, I_1,r_1) , (I_1,*,r_1) , (*, I_1,r_2) , (I_1,*,r_2) 34 valid gates
NOR Gate 2: (*, I_2,r_1) , (I_2,*,r_1) , (*, I_2,r_2) , (I_2,*,r_2) 34 valid gates
NOR Gate 3: (*, r_1,GFP) , (r_1,*,GFP) 2 valid gates

2312*2 total valid topologies

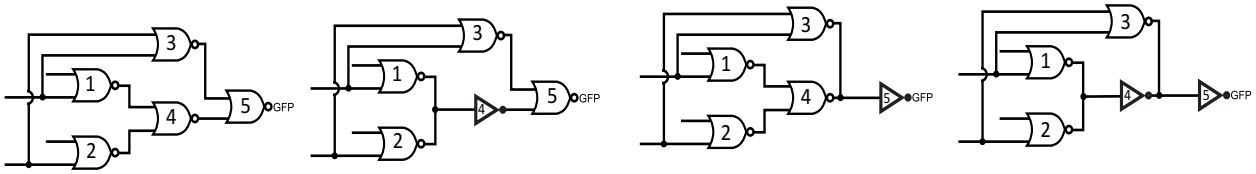
If $r_1 = r_2$:

NOR Gate 1: (*, I_1,r_1) , (I_1,*,r_1) , (*, I_1,r_2) , (I_1,*,r_2) 36 valid gates
NOR Gate 2: (*, I_2,r_1) , (I_2,*,r_1) , (*, I_2,r_2) , (I_2,*,r_2) 36 valid gates
NOR Gate 3: (*, r_1,GFP) , (r_1,*,GFP) 2 valid gates

2592*2 total valid topologies

9808/12800 = 0.76625 chance of valid topology

Figure 10.2 | Schematic of the AND circuit example construction and the relevant probabilities and library coverage



Two inputs I_1=r_1 and I_2=r_2. 6 total gRNA r_1, r_2, r_3, r_4, r_5, r_6. 5 total gates

NOR Gate pool:	Valid gate ratio
Gate 1: (*,r_1,r_4) (r_1,*, r_4) (*,r_1,r_5) (r_1,*, r_5)	if r_4=r_5, 2 * 32/40 if r_4≠r_5, 2 * 30/40
Gate 2: (*,r_2,r_4) (r_2,*, r_4) (*,r_2,r_5) (r_2,*, r_5)	if r_4=r_5, 2 * 32/40 if r_4≠r_5, 2 * 30/40
Gate 3: (r_1,r_2,r_3) (r_2,r_1,r_3) (r_1,r_2,r_6) (r_2,r_1,r_6)	4/4
Gate 4: (r_4,r_5, r_3) (r_5,r_4,r_3) (r_4,r_5, r_6) (r_5,r_4, r_6)	4/4
Gate 5: (r_3,r_6,GFP) (r_4,r_3,GFP)	2/2

when r_4=r_5 32768 *2 valid topologies when r_4≠r_5 28800 * 2 valid topologies

123136/204800 60.125% chance of valid topology

Figure 10.3 | A schematic of the XOR circuit example construction, the relevant probabilities and library coverage

10.4 Screening

Finally, after a circuit strand pool has been successfully transformed into yeast cells a there must be a way to differentiate between properly and improperly performing circuits. The screening method would need to be able to observe all input and output combinations of the circuit. One approach would be to run multiple rounds of FACS and sort the same population of cells under all possible input combinations and binning cells that have the expected fluorescence output. The binned cells from each input FACS round would then be sequenced using a high throughput or next gen sequencing to identify individual circuits components. If a specific combination of NOR gates is found to exist at a high concentration in these pools then it can be individually tested to confirm its proper behavior. It should be noted that as the size of the circuits increase it may be necessary to barcode and pre-sequence the NOR gate libraries due to the read length restrictions of sequencing techniques.

Chapter 11

Conclusions and future work

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 Figures 7.1 and 7.4 required that only a handful of gate combinations be screened to identify a functional design, and others required only one.

Table 7.3 compares our technology to selected published circuits. We measured circuit complexity with a combination of two metrics: The number of gates and the number of connections among gates. 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 *E. 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 behavior of upstream gates to which it is not connected by design[103–105]. 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[21] 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 less than 9.0. In such a case and when circuit sizes are small (fewer than twenty 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 approximately 60 bp of DNA long, and transcriptional regulation is a fairly well-understood process[106]. In contrast, the size of promoter and regulatory regions vary widely and can range from 250 bp to 10kb in yeast and other eukaryotes. Transcriptional regulation in eukaryotes is complex, involving a variety of mechanisms including chromatin remodeling[42,43,107,108,79], and understanding it remains a highly active area of research[109]. 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 7.3, is difficult. Nevertheless, we believe that because CRISPR/dCas9 functions in mammalian cells[24,41,57,59,70,71], and the human *Mxi1* repression domain has been used in synthetic contexts to regulate transcription in human cells [59,74,75], our NOR gates

could be to 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, 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[110], such as the strategy detailed in Chapter 10. 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, which decreases the possibility of interference between gates due to chromatin remodeling[42,111] and of yeast's tendency to recombine nearby homologous regions[112].

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 analog or mixed analog/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.

Bibliography

1. Shmulevich, I., Dougherty, E. R. & Zhang, W. E. I. From Boolean to Probabilistic Boolean Networks as Models of Genetic Regulatory Networks. **90,** (2002).

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

3. Hinkelmann, F. & Laubenbacher, R. Boolean Models of Bistable Biological Systems. *ArXiv e-prints* (2009).

4. 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).

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

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

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

8. 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).

9. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469,** 212–215 (2011).

10. Tian, T. & Burrage, K. Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage λ. *J. Theor. Biol.* **227,** 229–237 (2004).

11. Entian, K.-D. & Barnett, J. A. Regulation of sugar utilization by Saccharomyces cerevisiae. *Trends Biochem. Sci.* **17,** 506–510 (1992).

12. Boyer, L. A. *et al.* Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122,** 947–956 (2005).

13. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science (80-.).* **333,** 1307–1311 (2011).

14. Shankar, S. & Pillai, M. R. Translating cancer research by synthetic biology. *Mol. BioSyst.* **7,** 1802–1810 (2011).

15. Ye, H., Aubel, D. & Fussenegger, M. Synthetic mammalian gene circuits for biomedical applications. *Curr. Opin. Chem. Biol.* **17,** 910–917 (2013).

16. Marson, A. *et al.* Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. *Cell* **134,** 521–533 (2008).

17. Warren, L. *et al.* Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).

18. 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,** 8617–8622 (2011).

19. 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).

20. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31,** 448–52 (2013).

21. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science (80-.).* **352,** (2016).

22. Nielsen, A. A. K. & Voigt, C. A. Multi-input CRISPR / Cas genetic circuits that interface host regulatory networks. *Mol. Syst. Biol.* 1–11 (2014).

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

24. Kiani, S. *et al.* CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11,** 723–6 (2014).

25. Didovyk, A. & Tsimring, L. Orthogonal Modular Gene Repression in Escherichia coli Using Engineered CRISPR/Cas9. (2016). doi:10.1021/acssynbio.5b00147

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

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

28. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Meth* **5,** 374–375 (2008).

29. Stephanopoulos, G. Synthetic Biology and Metabolic Engineering. *ACS Synth. Biol.* **1,** 514–525 (2012).

30. Kalos, M. & June, C. H. Adoptive T Cell Transfer for Cancer Immunotherapy in the Era of Synthetic Biology. *Immunity* **39,** 49–60 (2013).

31. Miller, J., McLachlan, a. D. & Klug, a. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *J. Trace Elem. Exp. Med.* **14,** 157–169 (2001).

32. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11,** 636–646 (2010).

33. Beumer, K. J. *et al.* Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in Drosophila. *G3 (Bethesda).* **3,** 1717–25 (2013).

34. Ramirez, C. L. *et al.* Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* **5,** 374–375 (2008).

35. Segal, D. J. *et al.* Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42,** 2137–2148 (2003).

36. Sander, J. D. *et al.* Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* **8,** 67–69 (2011).

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

38. Dennis, M. K. *et al.* Targeted DNA Demethylation and Endogenous Gene Activation Using Programmable Tale-TET1 Fusions. *Nat. Biotechnol.* **127,** 358–366 (2012).

39. Wysocka, J. & Herr, W. The herpes simplex virus VP16-induced complex: The makings of a regulatory switch. *Trends Biochem. Sci.* **28,** 294–304 (2003).

40. Chasman, D. I., Leatherwood, J., Carey, M., Ptashne, M. & Kornberg, R. D. Activation of yeast polymerase II transcription by herpesvirus VP16 and GAL4 derivatives in vitro. *Mol. Cell. Biol.* **9,** 4746–4749 (1989).

41. 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).

42. 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).

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

44. Hall, D. B. & Struhl, K. The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *J. Biol. Chem.* **277,** 46043–46050 (2002).

45. Sterner, D. E. *et al.* Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction. *Mol. Cell. Biol.* **19,** 86–98 (1999).

46. Hahn, S. & Young, E. T. Transcriptional regulation in saccharomyces cerevisiae: Transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* **189,** 705–736 (2011).

47. Gaston, K. & Jayaraman, P. S. Transcriptional repression in eukaryotes: Repressors and repression mechanisms. *Cell. Mol. Life Sci.* **60,** 721–741 (2003).

48. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31,** 833–8 (2013).

49. Wu, J., Suka, N., Carlson, M. & Grunstein, M. TUP1 Utilizes Histone H3/H2B–Specific HDA1 Deacetylase to Repress Gene Activity in Yeast. *Mol. Cell* **7,** 117–126 (2001).

50. Sridhar, V. V, Surendrarao, A., Gonzalez, D., Conlan, R. S. & Liu, Z. Transcriptional repression of target genes by LEUNIG and SEUSS, two interacting regulatory proteins for Arabidopsis flower development. *Proc. Natl. Acad. Sci. United States Am.* **101,** 11494–11499 (2004).

51. 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).

52. Morris, K. V, Chan, S. W.-L., Jacobsen, S. E. & Looney, D. J. Small

Interfering RNA-Induced Transcriptional Gene Silencing in Human Cells. *Science (80-.).* **305,** 1289–1292 (2004).

53. Bird, A. *et al.* Multi-Input RNAi-Based Logic Circuit. 1307–1312 (2011).

54. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5,** 451–463 (2004).

55. 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).

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

57. Gao, Y. & Zhao, Y. Self-pr ocessing of r ibozyme-flanked RNAs into guide RNAs in vitr o and in vivo for CRISPR- mediated genome editing. 1–17 (2013).

58. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-.).* **321,** 960–964 (2008).

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

60. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10,** 1116–21 (2013).

61. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* (2014). doi:10.1038/nbt.3026

62. Wedekind, J. E. & Mckay, D. B. CRYSTALLOGRAPHIC STRUCTURES OF THE HAMMERHEAD RIBOZYME : Relationship to Ribozyme Folding and Catalysis. 475–502 (1998).

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

64. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci.* **109,** 8884–8889 (2012).

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

(2014).

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

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

68. Bowyer, J. *et al.* Mechanistic Modeling of a Rewritable Recombinase Addressable Data Module. **10,** 1161–1170 (2016).

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

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

71. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Squence Specific Control of Gene Expression. *Cell* **152,** 1173–1183 (2013).

72. Huang, Y. & Maraia, R. J. Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human. *Nucleic Acids Res.* **29,** 2675–2690 (2001).

73. 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).

74. 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,** 777–86 (1995).

75. 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).

76. Warren, L. *et al.* Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. *Cell Stem Cell* **7,** 618–630 (2016).

77. 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).

78. Tompa, M. *et al.* Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotech* **23,** 137–144 (2005).

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

80. Lewis, J. D. & Izaurralde, E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **247,** 461–469 (1997).

81. 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).

82. Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res.* **41,** 5139–5148 (2013).

83. Carothers, J. M., Goler, J. a, Juminaga, D. & Keasling, J. D. Devices to Quantitatively Program. *Science (80-.).* **36498,** 1716–1719 (2011).

84. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154,** 1380–9 (2013).

85. Traven, A., Staresincić, L., Arnerić, M. & Sopta, M. The yeast protein Xtc1 functions as a direct transcriptional repressor. *Nucleic Acids Res.* **30,** 2358–2364 (2002).

86. Flick, J. S. & Johnston, M. Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **10,** 4757–4769 (1990).

87. Pierre-Jerome, E., Jang, S. S., Havens, K. A., Nemhauser, J. L. & Klavins, E. Recapitulation of the forward nuclear auxin response pathway in yeast. *Proc. Natl. Acad. Sci.* **111,** 9407–9412 (2014).

88. Witzgall, R., O'Leary, E., Leaf, A., Onaldi, D. & Bonventre, J. V. The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 4514–4518 (1994).

89. McIsaac, R. S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41,** e57–e57 (2013).

90. Jorgensen, P. *et al.* The Size of the Nucleus Increases as Yeast Cells Grow. *Mol. Biol. Cell* **18,** 3523–3532 (2007).

91. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34,** 339–44 (2016).

92. Christiano, R., Nagaraj, N., Fröhlich, F. & Walther, T. C. Global Proteome Turnover Analyses of the Yeasts S. cerevisiae and S. pombe. *Cell Rep.* **9,** 1959–1965 (2014).

93. Veening, J.-W., Hamoen, L. W. & Kuipers, O. P. Phosphatases modulate the bistable sporulation gene expression pattern in Bacillus subtilis. *Mol. Microbiol.* **56,** 1481–1494 (2005).

94. Bagci, E. Z., Vodovotz, Y., Billiar, T. R., Ermentrout, G. B. & Bahar, I. Bistability in Apoptosis: Roles of Bax, Bcl-2, and Mitochondrial Permeability Transition Pores. *Biophys. J.* **90,** 1546–1559 (2006).

95. Harris, R. E., Pargett, M., Sutcliffe, C., Umulis, D. & Ashe, H. L. Brat Promotes Stem Cell Differentiation via Control of a Bistable Switch that Restricts BMP Signaling. *Dev. Cell* **20,** 72–83 (2011).

96. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).

97. Gnügge, R. & Rudolf, F. Saccharomyces cerevisiae Shuttle vectors. *Yeast* **34,** 205–221 (2017).

98. UW BIOFAB. at <http://www.uwbiofab.org/>

99. Horwitz, A. A. *et al.* Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst.* 1–9 (2015). doi:10.1016/j.cels.2015.02.001

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

101. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing , activation and repression. *Nat. Methods* 1–6 (2015). doi:10.1038/nmeth.3580

102. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* (2015). doi:10.1016/j.cell.2015.09.038

103. Del Vecchio, D., Ninfa, A. J. & Sontag, E. D. Modular cell biology: retroactivity and insulation. *Mol. Syst. Biol.* **4,** n/a--n/a (2008).

104. Jayanthi, S., Nilgiriwala, K. S. & Vecchio, D. Del. Retroactivity Controls the

Temporal Dynamics of Gene Transcription. *ACS Synth. Biol.* **2,** 431–441 (2013).

105. Mishra, D., Rivera, P. M., Lin, A., Del Vecchio, D. & Weiss, R. A load driver device for engineering modularity in biological networks. *Nat Biotech* **32,** 1268–1275 (2014).

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

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

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

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

110. Werner, S., Engler, C., Weber, E., Gruetzner, R. & Marillonnet, S. Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioengineered* **3,** 38–43 (2012).

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

112. Shinohara, A. & Ogawa, T. Homologous recombination and the roles of double-strand breaks. *Trends Biochem. Sci.* **20,** 387–391 (1995).

Appendix A

Sequence and strain information

All strain and sequence information is stored as a Microsoft Excel file at "https://images.nature.com/original/nature-assets/ncomms/2017/170525/ncomms15459/extref/ncomms15459-s2.xlsx"