UNRAVELING THE OPERATIONAL LANDSCAPE OF ENGINEERED GENETIC REGULATORY NETWORKS

by

Tyler George Quarton

APPROVED BY SUPERVISORY COMMITTEE:

Leonidas Bleris, Chair

Michael Zhang

Stephen Levene

Mark Spong

This work is dedicated to my love Patricia Quarton.

UNRAVELING THE OPERATIONAL LANDSCAPE OF ENGINEERED GENETIC REGULATORY NETWORKS

by

TYLER GEORGE QUARTON, BS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

BIOMEDICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2019

ACKNOWLEDGMENTS

First, I would like to thank my supervisor Dr. Leonidas Bleris who facilitated my scientific curiosity by allowing me to seek truths in all corners of science. I would also like to thank my committee members for sharing their knowledge and support with me throughout the years. I would like to thank my brother, Iuri Gagnidze, for being a constant source of inspiration and support. Furthermore, I would like to thank all my lab members for their enduring friendship and scientific collaboration. Lastly, I would like to especially thank my parents, whose hard work, love, and endless support positioned me to be able to pursue my dreams.

November 2019

UNRAVELING THE OPERATIONAL LANDSCAPE OF ENGINEERED GENETIC REGULATORY NETWORKS

Tyler George Quarton, PhD
The University of Texas at Dallas, 2019

Supervising Professor: Leonidas Bleris

The operation of all living cells is executed by a collection of complex biomolecular regulatory networks. Naturally occurring networks exhibit remarkable properties as they robustly operate in an ever-changing cellular milieu. In this dissertation, we present theoretical and experimental results that investigate broadly, the cellular context in which regulatory networks operate and, more specifically, the nuanced regulatory action of a class of small non-coding RNA molecules called microRNAs.

In investigating the cellular context, we commenced our work by exploring the impact that resource scarcity may have on regulatory network function by mathematically inferring gene interactions through the experimental perturbation of a synthetic gene circuit. We then investigate how the inherent stochastic environment of a cell affects protein production through the characterization of a panel of synthetic circuits stably integrated in human cells. We develop and validate a novel mathematical framework that can reliably attribute measured cellular noise to local and global fluctuation sources.

In investigating microRNA function, we commenced our work by engineering custom genetic circuits that contain microRNA-based regulation and introduced an analytical strategy that utilizes the clustering and superposition of discrete experiments to produce a "bird's-eye view" perspective capable of correlating the function of a microRNA to its concentration and the genetic circuit abundance. We then studied intragenic miRNA-mediated host transcript regulation, where the miRNA and its target are processed from the same gene transcript. To this end, a synthetic system was stably integrated in human kidney cells within a genomic cell safe harbor locus. We provide results showing robust output filtering with respect to promoter strength and a validation of the experimental observations based on stochastic population model simulations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The capacity for living systems to regenerate, replicate, and evolve uniquely manifests from the biochemical information processing capabilities of genetic material. Housed within biological cells, this genetic material naturally partitions into units of genes that are often equipped with biochemically interfaceable regulatory elements. These regulatory elements enable the information encoded within genes to interact with itself and its environment through the controlled production of biomolecular regulators. The assemblage of these regulators along with their web of interactions responsible for the orchestration of gene expression are collectively referred to as gene regulatory networks.

Enabled by high-throughput gene expression profiling technologies, the unraveling of a cell's gene regulatory network became possible through the data-driven development of various inferencing methods[1–5], including perturbation-based network reconstruction methods capable of distinguishing between global network responses caused by disturbance propagation from local pairwise interactions[6–8]. The successful inference of gene regulatory networks has been used to further our understanding of disease[9,10] and guide the rational design of drugs and therapies[11–13]. However, the inherent complexity of gene regulatory networks presents complications in which the direct characterization of the impact specific interactions have on a network's behavior and the faithful discrimination between causal and non-casual interactions remains a challenge. The fields of systems and synthetic biology approach these challenges through the analysis of rationally designed and isolated gene regulatory networks. In this dissertation, we present theoretical and experimental results that investigate broadly, the cellular context in which gene regulatory

networks operate and, more specifically, the nuanced regulatory action of a class of small non-coding RNA regulator biomolecules called microRNA.

Systems and synthetic biology are multidisciplinary fields of study that bridge concepts from physics, engineering, mathematics and biology with the intent of both understanding and utilizing the capabilities of living systems through the engineering of simplified gene regulatory networks called synthetic gene circuits. Enabled by the technological advances of both DNA synthesis and next-generation sequencing, gene circuits can be modularly constructed using a rational combination of genetic elements found naturally throughout the animal kingdom or artificially through design. As such, synthetic gene circuits can be constructed to either passively interface with natural gene regulatory networks through logical sensing[14–17] and isolated emulation[18] or actively interface with natural gene regulatory networks through direct actuation[19,20]. Additionally, with the modern advances of CRISPR technology enabling the surgically precise engineering of a genomes coupled with the intentional exclusion of endogenous regulatory elements, synthetic circuits are capable of orthogonally operating as disconnected subnetworks fully integrated within the genome of cell[18–21]. This property of orthogonality has been leveraged to explore the nuanced behavior of gene regulatory networks through a reductionist's lens by isolating interactions of interest found within complex endogenous systems[18,24,25]. One such interaction of interest presented in this dissertation is the contextual relationship established between gene regulatory networks and the cellular environment in which they operate.

Within the spatial confines of a cell's membrane, the activity of the relatively small quantity of biochemical species therein often precedes the limits of averaging afforded by the

statistical law of large numbers. As a consequence, the genetic program responsible for the maintenance and operation of all living cells is executed by the collection of complex gene regulatory networks arising from underlying, non-deterministic biomolecular interactions. Rather than this inherent variability existing as an unavoidable disruptive source of error, natural gene regulatory networks have evolved to robustly operate in, and even exploit, their stochastic environment to their operational advantage[26]. Furthermore, the production and interaction of every regulatory element within a gene regulatory networks are globally facilitated through the shared biochemical resources the cell endogenously provides. The heavy demand of these shared resources within large or operationally demanding regulatory networks can create undesired or unexpected network behaviors[27]. The consideration of the impact this imbalanced allocation of resource can have is vital in designing robustly operating synthetic circuits. In Chapter 2, we present a benchmark synthetic gene circuit that emulates a feedforward gene regulatory network motif that examines the impact potential resource scarcity has on gene network behavior[28]. Here, we present experimental and theoretical results in which reverse engineering uncovers a hidden, nontrivial regulatory interaction that was not explicitly engineered into the network. Though advances in understanding how gene regulatory networks mitigate the effects of stochastic fluctuations, the propagative effects of noise originating specifically within the natural cascade of the central dogma has yet to be dissected in an isolated synthetic system. Thus, in Chapter 3 we present the analysis of a family of three engineered synthetic circuits that variably interface with the central dogma and explore the inherent stochastic nature of the biochemical processes within the various protein production stages. Here, we provide a generalized statistical framework capable

of discriminating the effects of both global and local sources of noise present in populations of cells.

Another gene regulatory interaction of interest presented in this dissertation is the regulation of proteins via the targeted inhibition of precursor mRNA translation by microRNA. MicroRNA are a recently discovered class of small, noncoding RNAs that regulate gene expression through the disruption of post-transcriptional processes. The unique mechanism by which microRNA affect protein production confers novel regulatory properties in gene regulatory networks. As such, microRNA are of great interest as both a nuanced class of natural gene network regulators and as modular components of advanced synthetic gene circuits[16]. Though the effects of microRNA interaction with respect to its target output have been previously explored[29,30], the functional connection between intergenic microRNA activity relative to its circuit template abundance has yet been established. In Chapter 4, we present results in which we examine the mechanistic activity of microRNA operating between genes within a self-contained synthetic gene circuit[31]. Here, we elucidate the function of microRNA inhibition relative to gene circuit abundance by introducing an analytical strategy that utilizes the clustering and superposition of discrete experiments to produce a "bird's-eye view" perspective. Endogenously, microRNA have been found to both interact between genes and within a single gene, as a gene can produce both a microRNA and its mRNA target. Though ubiquitous in endogenous systems, this internal interaction edge has yet to be explored within the context of an isolated synthetic system. In Chapter 5, we present work in which we explore the impact of intragenic microRNA host regulation through the implementation and analysis of a synthetic gene circuit containing a microRNA that originates from the gene it regulates. Here we uncover a unique property of

intragenic microRNA-based filtration in which gene expression is selectively allowed through an established promoter strength threshold that naturally emerges through the intragenic microRNA interaction.

1.1 References

1. Zhang, X. *et al.* Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics* (2012). doi:10.1093/bioinformatics/btr626

2. Wille, A. *et al.* Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana. *Genome Biol.* (2004). doi:10.1186/gb-2004-5-11-r92

3. Friedman, N. Inferring Cellular Networks Using Probabilistic Graphical Models. *Science* (2004). doi:10.1126/science.1094068

4. Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. U. S. A.* (2000). doi:10.1073/pnas.220392197

5. Liang, S., Fuhrman, S. & Somogyi, R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.* (1998).

6. Bruggeman, F. J., Westerhoff, H. V., Hoek, J. B. & Kholodenko, B. N. Modular response analysis of cellular regulatory networks. *J. Theor. Biol.* (2002). doi:10.1016/S0022-5193(02)93096-1

7. Kholodenko, B., Yaffe, M. B. & Kolch, W. Computational approaches for analyzing information flow in biological networks. *Science Signaling* (2012). doi:10.1126/scisignal.2002961

8. Sontag, E. D. Network reconstruction based on steady-state data. *Essays Biochem.* (2008). doi:10.1042/BSE0450161

9. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* (2005). doi:10.1038/ng1532

10. Madhamshettiwar, P. B., Maetschke, S. R., Davis, M. J., Reverter, A. & Ragan, M. A. Gene regulatory network inference: Evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Med.* (2012). doi:10.1186/gm340

11. Li, Y., Nowak, C. M., Withers, D., Pertsemlidis, A. & Bleris, L. CRISPR-Based Editing Reveals Edge-Specific Effects in Biological Networks. *Cris. J.* (2018). doi:10.1089/crispr.2018.0018

12. Fortney, K. *et al.* NetwoRx: Connecting drugs to networks and phenotypes in Saccharomyces cerevisiae. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gks1106

13. Ghosh, S. & Basu, A. Network medicine in drug design: Implications for neuroinflammation. *Drug Discovery Today* (2012). doi:10.1016/j.drudis.2012.01.018

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

15. Ehrhardt, K., Guinn, M. T., Quarton, T., Zhang, M. Q. & Bleris, L. Reconfigurable hybrid interface for molecular marker diagnostics and in-situ reporting. *Biosens. Bioelectron.* **74**, 744–750 (2015).

16. Guinn, M. & Bleris, L. Biological 2-input decoder circuit in human cells. *ACS Synth. Biol.* **3**, 627–33 (2014).

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

18. Kang, T., Moore, R., Li, Y., Sontag, E. & Bleris, L. Discriminating direct and indirect connectivities in biological networks. in *Food, Pharmaceutical and Bioengineering Division 2015 - Core Programming Area at the 2015 AIChE Meeting* (2015). doi:10.73/pnas.1507168112/-/DCSupplemental

19. Li, Y., Moore, R., Guinn, M. & Bleris, L. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci. Rep.* (2012). doi:10.1038/srep00897

20. Moore, R. *et al.* CRISPR-based self-cleaving mechanism for controllable gene delivery in human cells. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gku1326

21. Li, Y. *et al.* Coevolutionary couplings unravel PAM proximal constraints of CRISPR-SpCas9. *Biophys. J.* (2019). doi:10.1016/j.bpj.2019.09.040

22. Nowak, C. M., Lawson, S., Zerez, M. & Bleris, L. Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Research* (2016). doi:10.1093/nar/gkw908

23. Li, Y. *et al.* Exploiting the CRISPR/Cas9 PAM constraint for single-nucleotide resolution interventions. *PLoS One* (2016). doi:10.1371/journal.pone.0144970

24. Shimoga, V., White, J. T., Li, Y., Sontag, E. & Bleris, L. Synthetic mammalian transgene negative autoregulation. *Mol. Syst. Biol.* **9**, 670 (2013).

25. Shin, Y. J. & Bleris, L. Linear control theory for gene network modeling. *PLoS One* (2010). doi:10.1371/journal.pone.0012785

26. Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: From microbes to mammals. *Cell* (2011). doi:10.1016/j.cell.2011.01.030

27. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal* (2012). doi:10.1002/biot.201200085

28. Quarton, T., Kang, T., Sontag, E. D. & Bleris, L. Exploring the impact of resource limitations on gene network reconstruction. in *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (2016). doi:10.1109/CDC.2016.7798773

29. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol. Syst. Biol.* (2011). doi:10.1038/msb.2011.49

30. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* **43**, 854–9 (2011).

31. Quarton, T. *et al.* Mapping the operational landscape of microRNAs in synthetic gene circuits. *npj Syst. Biol. Appl.* **4**, 6 (2018).

CHAPTER 2

EXPLORING THE IMPACT OF RESOURCE LIMITATIONS ON GENE NETWORK RECONSTRUCTION[†]

Authors – Tyler Quarton[1,2], Kristina Ehrhardt[1,2], James Lee[3], Srijaa Kannan[4], Yi Li[1,2], Lan Ma and Leonidas Bleris[1,2,]

Department of Biomedical Engineering

The University of Texas at Dallas

800 W Campbell Road

Richardson, Texas 75080-3021

[1] Bioengineering Department, University of Texas at Dallas, Richardson, TX, USA

[2] Center for Systems Biology, University of Texas at Dallas, Richardson, TX, USA

[3] Department of Biological Sciences, University of Texas at Dallas, Richardson, TX USA

[4] School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX, USA

2.1 Abstract

The characterization of biological networks via mathematical models often involves cycles of experimental perturbations and measurements, followed by the use of a network inference method. Here we study an engineered genetic circuit, introduced in a recent paper by the authors, and report additional analysis and interpretation. Using this synthetic network as a benchmark, we find that the application of the modular response analysis (MRA) network inference method leads to the discovery of a hidden, nontrivial "ghost" regulatory edge, which was not explicitly engineered into the network. Importantly, this result is not evident from direct inspection of the experimental measurements and global response coefficients. To probe the global to local conversion in MRA, we use conditionally randomized global response matrices to obtain distributions of local response coefficients and demonstrate that sign changes are numerically possible. Additionally, using simulations of a cascade network in a biochemical setting which does not take into account resource limitations, we show that MRA cannot return "ghost" edges, which points to the impact of the cellular milieu and in particular the use of shared resources. Taking resource availability into account during reverse engineering may allow for closer approximation of the cellular environment and points to a potential opportunity for network characterization strategies.

2.2 Introduction

Reverse engineering of gene regulatory networks is a fundamental area of research in systems and synthetic biology[1–4]. The availability of massive sets of gene expression data has shifted the focus from describing interactions among several genes to comprehensive analysis at

the network level[5–8]. As the scope of analysis expands to a larger system, the dynamics of common cellular resources must be factored into the analysis as well. The abundance of factors that govern the central dogma are generally understood to be monitored and regulated throughout the cell cycle[9–11], and consequently have a profound impact on the activity of the cellular network[12,13].

Mathematical frameworks that enable the analysis of resource derived network disturbances have been established. Specifically, network effects caused by molecular competition were derived using Metabolic Control Analysis[14], metabolic constraints were established and explored using Flux Balance Analysis[15], design rules to avoid intranetwork cross couplings were outlined through enzymatic competition considerations[16], and queuing models were proposed to explore the effects of ribosome bottlenecking[17,18]. In a recent work, Qian et al. developed a framework which accounts for the effects arising from the sharing of RNA polymerase and ribosomes in gene networks specifically[19].

We use as a benchmark a synthetic three-node cascade motif that is orthogonal to the cellular milieu but operates using shared resource[1]. We subject our network to iterations of perturbation, measurement and reverse engineering[2]. The synthetic network, which consists of three distinct nodes with their own fluorescent reporters, features an indirect regulation between the input and the output node (i.e. X activating Y, Y activating Z). After perturbation-based network inference, we show that a weak direct regulatory relationship between the input (X) and output node (Z) is uncovered, in addition to the intended connections. Here we study the origin of this ghost edge uncovered during network inference. This chapter discusses experimental results and numerical explorations.

2.3 Benchmark Synthetic Circuit

The model synthetic network consists of three independent nodes in the cascade motif, where node X activates node Y, which, in turn, activates node Z (Figure 2.1). Each node features a distinct reporter fluorescent protein and controls its respective downstream node via inducible transcriptional activation. Perturbation of each node in the system was made possible by introducing small interfering RNA (siRNA) that directly inhibit translation of the fluorescent protein as well as the co-expressed transactivator proteins of the node.

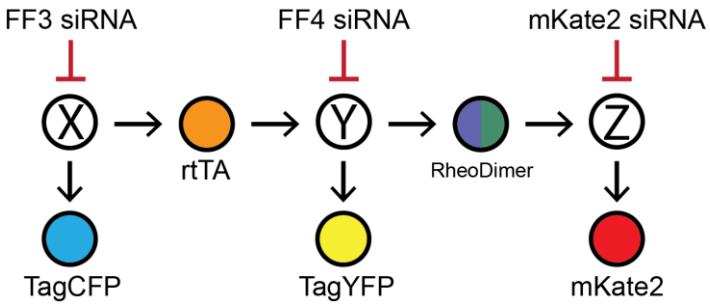


Figure 2.1: **Graphical representation of the synthetic benchmark circuit.** Each node produces fluorescent reporter protein as well as the transcriptional activator protein to its downstream node. Perturbation of each node made possible through siRNA[1].

The entire synthetic network was transiently delivered to human embryonic kidney (HEK293) cells (Figure 2.2). Node X is constitutively active, doxycycline enables node Y activation by node X, and ponasterone A mediates activation of node Z by node Y. As there are multiple transcripts that must be targeted to disrupt nodes X and Y, perturbations of these nodes are performed by synthetic siRNAs whose target sequences are located in 3′ UTR of all transcripts produced[5].

The regulatory interactions of the synthetic circuit are mediated by chemical ligands doxycycline and ponasterone A (Figure 2.2). Activity of node X, represented by TagCFP, is constitutive and produces transactivator rtTA. In combination with doxycycline, rtTA activates node Y to produce

TagYfp and both components of RheoSwitch transactivator. Finally, node Z is activated when an ecdysone receptor agonist such as ponasterone A binds to the RheoSwitch dimer complex to enable Gal4 domain activity and production of mKate2 fluorescent protein.

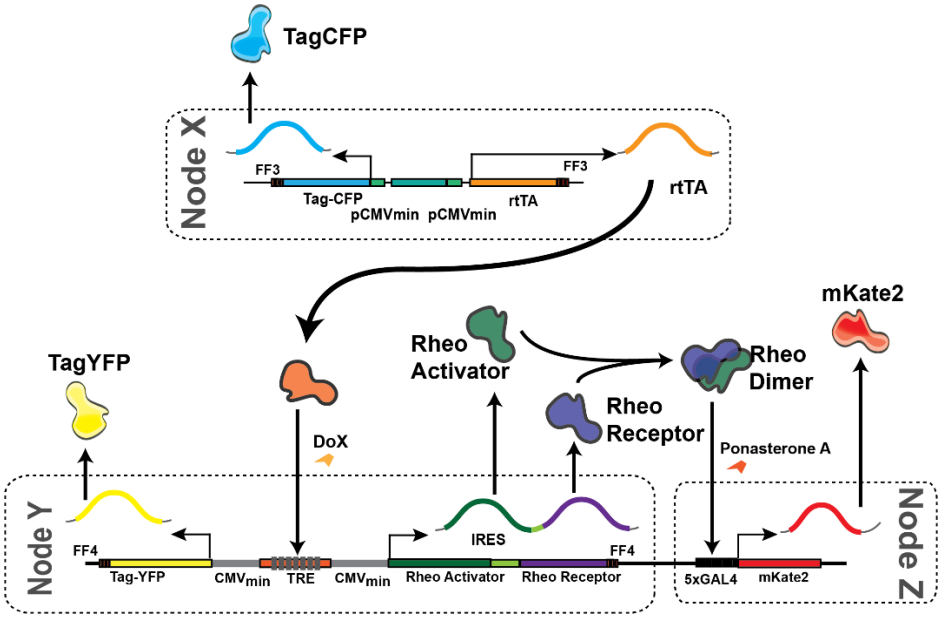


Figure 2.2: **Components of the synthetic cascade network.** Each node produces a fluorescent reporter protein as well as a transactivator protein for its downstream target activation, with the exception of node Z. Ubiquitous disruption of nodes X and Y are achieved via siRNA targets in the 3' UTR of the transcripts produced[1].

To confirm the architecture of the synthetic network, we performed two separate validation tests[1]. First, the circuit plasmids are transfected transiently and the activity of each output protein is measured 48 hours post-transfection after various combinations of the ligands that mediate node-to-node interactions (Figure 2.3)[1].

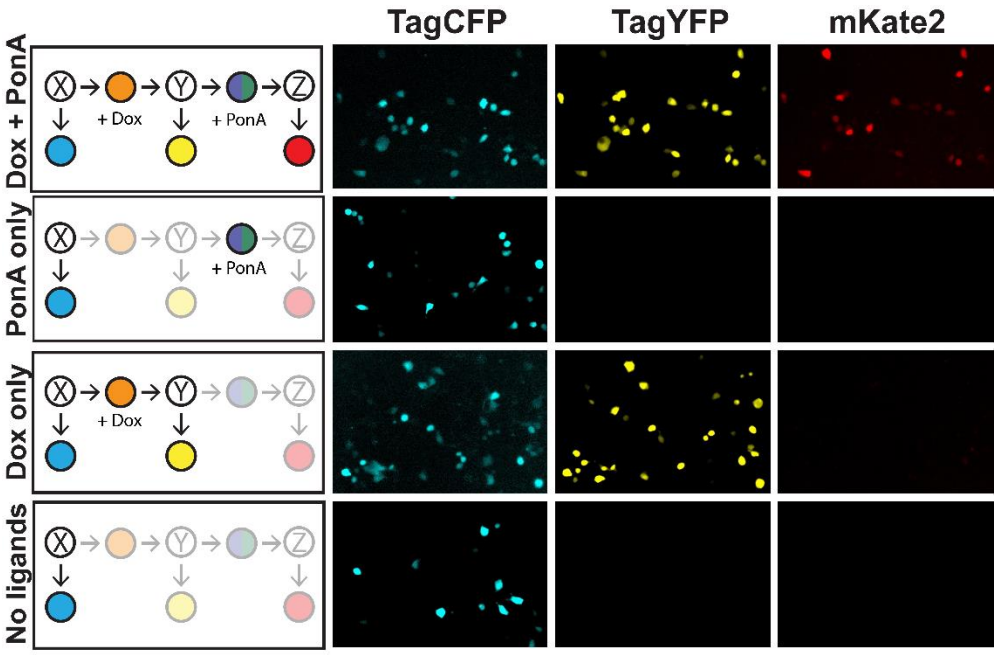


Figure 2.3: **Confirmation of the regulatory connections in the synthetic cascade network.** Network activity was measured via fluorescence microscopy under all combinations of the inducer ligands that activate the regulatory links of the network[1].

Since multiple components are produced in nodes X and Y, we opted for perturbation using synthetic siRNAs and placed their targets at the 3' UTR of every transcript produced at each node. For node X, siRNA FF3 was used, and for node Y, FF4 siRNA. For node Z, we used a custom siRNA that exclusively targets the fluorescent protein's transcript[1]. We validated this node-specific perturbation through titration of each siRNA. The siRNAs are co-transfected with the circuit, and as before the outputs are measured after 48 hours via flow cytometry (Figure 2.4).

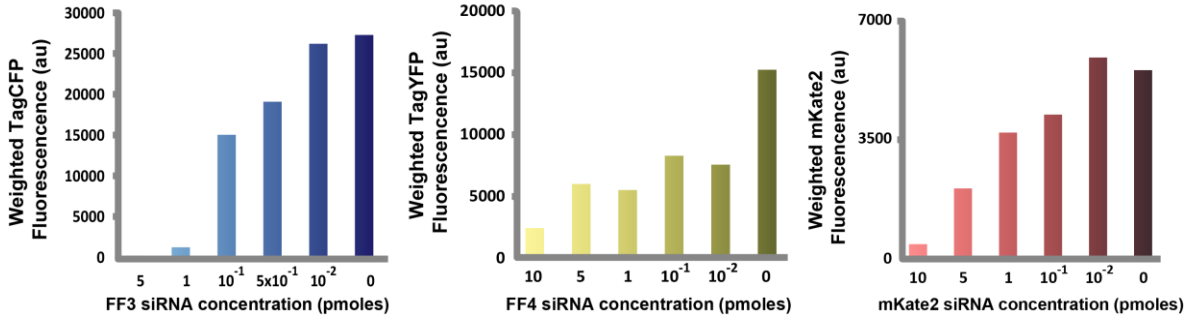


Figure 2.4: **Individual node response following perturbation with siRNA that target each node.** Each node of the synthetic cascade network is perturbed via RNA interference. To measure the efficacy of each siRNA, titrations were performed[1].

2.4 Network Reconstruction

When it comes to perturbation-based network reconstruction methods, accurately distinguishing direct interactions from indirect interactions is one of the most important and complex tasks. Perturbation to a single node may propagate through the network and results in changes at a global scale, which renders inference of pairwise interactions particularly challenging. Over the years, there have been several attempts to address this issue[20,21]. One such method is modular response analysis (MRA), which uses linear algebra techniques to derive local, pairwise interaction coefficients from global response data[22,23].

The MRA method has previously shown to be successful in characterization of the canonical Raf/Mek/Erk MAPK network pathway[24]. The procedure for reverse engineering of our synthetic cascade network using MRA is as follows: first, we capture the steady states of the system after siRNA perturbation for each node at a specified concentration, which we use to calculate the global response coefficients. After constructing a matrix consisting of the response coefficients, we then use this matrix to calculate the local response coefficients[23].

2.4.1 Modular Response Analysis

We consider a system of differential equations of the following general form:

$$\dot{x}_i = F_i(x_1, \dots, x_N, p_i), i = 1, \dots, N$$

where each F_i is continuously differentiable. The variables x_i, $i = 1, \dots, N$ denote concentrations of species, and each p_i, $i = 1, \dots, N$ is a parameter. Note that the ith parameter only affects directly the ith equation.

The MRA method amounts precisely to the following procedure. We are interested in finding the direct dependencies of each variable x_i on each other variable x_j. More precisely, we assume that $\xi = (\xi_1, \dots, \xi_N)^T$ is a steady state of the system corresponding to a nominal parameter $\pi = (\pi_1, \dots, \pi_N)^T$, that is,

$$F_i(\xi_1, \dots, \xi_N, \pi_i) = 0, i = 1, \dots, N$$

and we want to estimate the gradients ∇F_i evaluated at (ξ, π), $i = 1, \dots, N$. We assume that there is a smooth curve $x(p)$ of steady states, where p is some neighborhood of the nominal parameter π, and $x(\pi) = \xi$ and denote by

$$\frac{dx_k}{dp_j}$$

the global change in the kth coordinate, x_k, of the steady state when p_j is perturbed. Taking derivatives with respect to $p_j, j \neq i$ in $F_i(x_1(p), \dots, x_N(p), p_i) \equiv 0$, we obtain:

$$\sum_{k=1}^{N} \frac{\partial F_i}{\partial x_k} \frac{dx_k}{dp_j} = 0 \quad \forall j \neq i. \tag{1}$$

Thus, we have the orthogonality equations

$$\nabla F_i \cdot P_j = 0, i = 1, \dots, N \quad \forall j \neq i,$$

where

$$P_j = \left(\frac{dx_1}{dp_j}, \frac{dx_2}{dp_j}, \dots, \frac{dx_N}{dp_j}\right).$$

From steady-state information, we can never find the gradients of the F_i's except up to a constant multiple. As such, we attempt to estimate the normalized *local response coefficients*, defined as follows:

$$r_{ij} := -\frac{\partial F_i}{\partial x_j} \Big/ \frac{\partial F_i}{\partial x_i}$$

for $i, j \in \{1, \dots, N\}$, assuming that $\frac{\partial F_i}{\partial x_i} \neq 0$ (typically negative, hence the negative sign). Note that, by definition, $r_{ii} = -1$ for all i. The matrix $\mathbf{r}$ consisting of the r_{ij}'s is called the *local response matrix*. Observe that (Eq. 1), together with the definition of r_{ij}, gives that, for all i and all $j \neq i$,

$$\frac{dx_i}{dp_j} = \sum_{k \neq i} r_{ik} \frac{dx_k}{dp_j}. \tag{2}$$

We also write, purely formally,

$$\frac{\partial x_i}{\partial x_j} := -\frac{\partial F_i}{\partial x_j} \Big/ \frac{\partial F_i}{\partial x_i} = r_{ij}. \tag{3}$$

Intuitively, we interpret each $\frac{\partial x_i}{\partial x_j}, i \neq j$, as a perturbation in the activity x_i of species i with respect to infinitesimal perturbations in the activity x_j of species j under the idealized assumption that one could keep the activities of all other species clamped. These coefficients are all understood as evaluated at $x = \xi$ and $p = \pi$.

Fix any i. Assuming that F_i is continuously differentiable and $\frac{\partial F_i}{\partial x_i} \neq 0$ at $x = \xi$, $p = \pi$, the Implicit Function Theorem says that there is a function

$$x_i = f_i(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N, p)$$

defined in a neighborhood of (ξ, π), with the property that, as an equation on X_i,

$$F_i(x_1, x_2, \dots, x_{i-1}, X_i, x_{i+1}, \dots, x_N, p) = 0$$

has a unique solution in a neighborhood of (ξ, π), given by

$$X_i = f_i(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_N, p),$$

and furthermore, in this neighborhood

$$\frac{\partial F_i}{\partial x_k} = -\frac{\partial F_i}{\partial x_i}\frac{\partial f_i}{\partial x_k} \quad \forall k \neq i. \tag{4}$$

These formulae therefore hold also when specialized to $x = x(p)$ for parameters p near the nominal parameter π, and in particular at $x = \xi$. Thus,

$$\frac{\partial f_i}{\partial x_k} = -\frac{\partial F_i}{\partial x_j} \Big/ \frac{\partial F_i}{\partial x_i} = r_{ij} = \frac{\partial x_i}{\partial x_k} \quad \forall k \neq i$$

Now, substituting (Eq. 4) in (Eq. 1) we have

$$\frac{\partial F_i}{\partial x_i}\frac{dx_i}{dp_j} - \sum_{k \neq i} \frac{\partial F_i}{\partial x_i}\frac{\partial f_i}{\partial x_k}\frac{dx_k}{dp_j} = 0 \quad \forall j \neq i. \tag{5}$$

Let us define the *global response coefficients* as follows:

$$R_{ij} = \frac{dx_i}{dx_j} := \frac{dx_i}{dp_j} \Big/ \frac{dx_j}{dp_j} \quad \forall i, j.$$

Note that, by definition, $R_{ii} = 1$ for all i. The matrix **R** consisting of the $R_{i,j}$'s is called the *global response matrix*. Observe that

$$\frac{dx_k}{dp_j} = \frac{dx_k}{dx_j}\frac{dx_j}{dp_j} \quad \forall k, j.$$

Substituting these expressions in (Eq. 5), and then factoring out $\frac{\partial F_i}{\partial x_i}\frac{dx_j}{dp_j}$ (generically assumed nonzero), we arrive at

$$\frac{dx_i}{dx_j} = \sum_{k \neq i} \frac{\partial f_i}{\partial x_k}\frac{dx_k}{dx_j} \forall i \neq j,$$

which can also be written when using (Eq. 3) as

$$\frac{dx_i}{dx_j} = \sum_{k \neq i} \frac{\partial x_i}{\partial x_k}\frac{dx_k}{dx_j} \forall i \neq j, \tag{6}$$

or, equivalently in terms of the r and **R** notations:

$$R_{ij} = \sum_{k \neq i} r_{ik} R_{kj} \forall i \neq j. \tag{7}$$

This is the basic MRA equation that relates local and global coefficients. Assuming that the global coefficients R_{ij} are estimated from perturbation experiments, equation (Eq. 7) allows one to solve for the local coefficients, which indicate direct interactions[23].

2.4.2 Experimental Perturbation and Measurements

Using the flow cytometry data gathered after perturbation, we proceeded with the network inference using MRA (Figure 2.5)[1]. The global response coefficients (GRC) were calculated using bootstrapped mean of the gated fluorescent population, which are in turn used to calculate local response coefficients (LRC). To obtain a reasonable confidence interval associated with each LRC, we performed the calculation cycle 2,000 times and use 95% confidence interval of the resulting

LRC distributions for the final results. In other words, we only consider recovered LRC as significant if the 95% confidence interval does not intersect zero.

While we are able to recover positive, non-zero LRC in edges representing interaction from node X to node Y and node Y to node Z, we unexpectedly discovered an apparent direct interaction between node X to node Z. This reaction coefficient, though considered negligible due to its confidence interval, is significantly higher in magnitude than other unintended edges such as node Y to node X and node Z to node Y. Moreover, the same pattern was observed (Figure 2.5) when another set of siRNAs was used to perform systematic perturbation at a different level (i.e. low and high perturbation).

To summarize, MRA applied on the benchmark architecture yields a nontrivial "ghost" edge. In the next section, we study the properties of the global to local response coefficient conversion in MRA using randomized matrices and numerical simulations of cascade architectures.

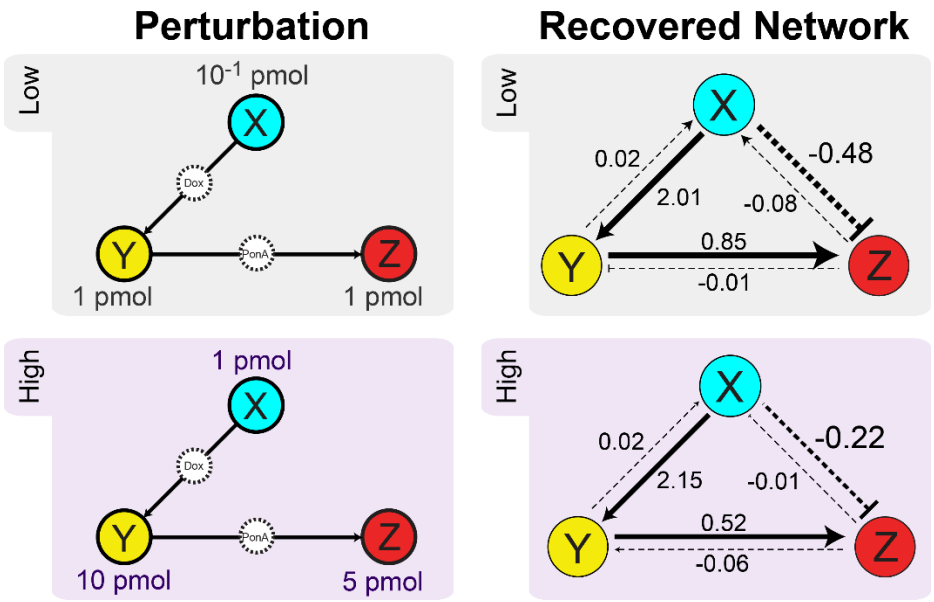


Figure 2.5: **Graphical representation of the perturbation properties.** Systematic perturbation to the network is performed twice, each with a different set of siRNA concentrations. Each siRNA concentrations are categorized as either "low" or "high" to indicate the relative strength of suppression. Interaction coefficients for all regulatory links were then calculated using modular response analysis using mean of fluorescence[1].

2.5 Randomized Cascades and Simulations

Firstly, to numerically explore the effects of processing global measurements to local responses, we performed MRA analysis on 100,000 randomized 3-node cascade networks. Using the central fractional difference to quantify perturbation results[22], given generally as

$$\frac{2(P_{pert} - P_{unpert})}{(P_{pert} + P_{unpert})}, \tag{8}$$

where P is the quantity being perturbed, we constructed corresponding global response matrices that represents all possible perturbations and after applying MRA we obtained distributions of local response coefficients. Secondly, we simulated the genetic circuit behavior and response to perturbation using a representative ODE model in MATLAB SimBiology.

2.5.1 Randomized Cascades

To study the distribution of all possible local response coefficients derived from perturbation of a cascade architecture we apply the MRA using conditionally randomized global response matrices. We restricted our analysis to direct 3 node cascades and include only the expected perturbation responses. Specifically, when X is perturbed, we expect X, Y, and Z to decrease, when Y is perturbed, we expect Y and Z to decrease, and when Z is perturbed, we expect only Z to decrease. By inspection, the maximum value of the central fractional difference (8) of two positive measurements upon negative perturbation is 0 and the minimum is - 2. Therefore, all possible direct cascade central fractional differences are contained within the range [-2,0]. Thus, the global response elements of our randomize 3-node direct cascades are designed as follows

$$\mathbf{R}_{ij} = \begin{cases} \sim U([-2,0]) & i \geq j \\ 0 & i < j \end{cases},$$

where $\sim U([-2,0])$ indicates a random value from a uniform distribution ranging from -2 to 0. After generating our cascade perturbation results, we subjected each global matrix to MRA analysis and recovered their local responses. By looking at the probability density function corresponding to the histogram of the recovered local response coefficients, we confirm that the responses r_{yx} and r_{zy} are strictly positive and, interestingly, r_{yz} has a non-zero probability of being negative (Figure 2.6).

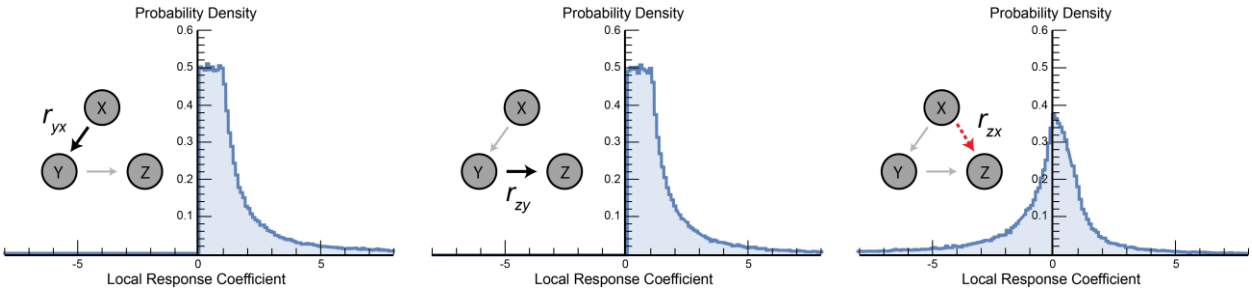


Figure 2.6: **Probability density functions of local response coefficients obtained from random cascades.** Only the probability density function for r_{zx} (bottom) includes both positive and negative coefficients.

As such, we conclude that MRA has the numerical capacity to produce a negative edge from nodes X to Z.

2.5.2 Perturbation Simulations

A genetic cascade comprised of nodes that represent DNA, RNA, and protein species were modeled using MATLAB SimBiology using previously published kinetic parameter[5]. To emulate various strengths of siRNA perturbation on nodes X, Y, and Z, we gradually increase the mRNA degradation of each respective node. The cascade's dynamics were then simulated (Figure 2.7) and steady-state values were obtained.

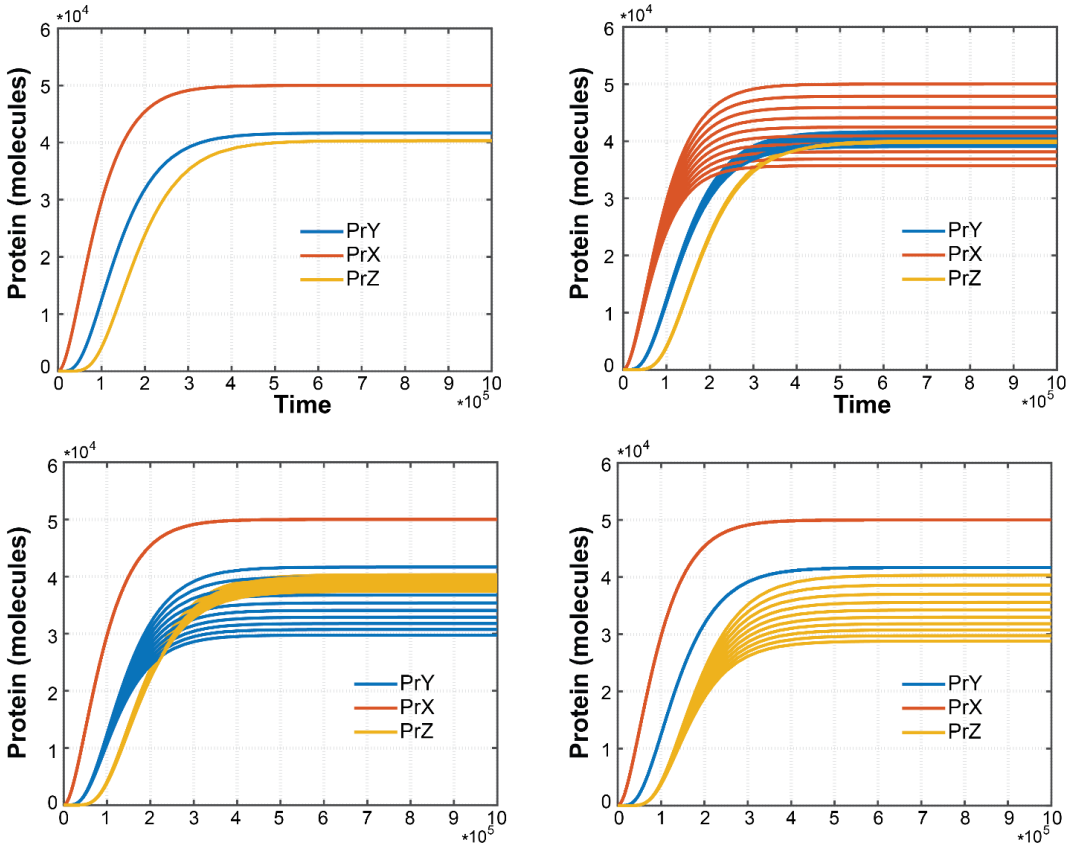


Figure 2.7: **Simbiology perturbation simulations.** The model cascade was serially perturbed at 9 different levels of increasing mRNA degradation at each respective node. Cascade dynamics for node X, Y, and Z were simulated for each perturbation case.

Using the central fractional difference (Eq. 8) as before, we calculate the global response matrices for each perturbation strength. After subjecting these global responses to MRA, we observe that in the case of our in-silico transcriptional cascade, we recover positive values for responses r_{yx} and r_{yz} while r_{zx} is zero for all perturbation strengths (Figure 2.8).

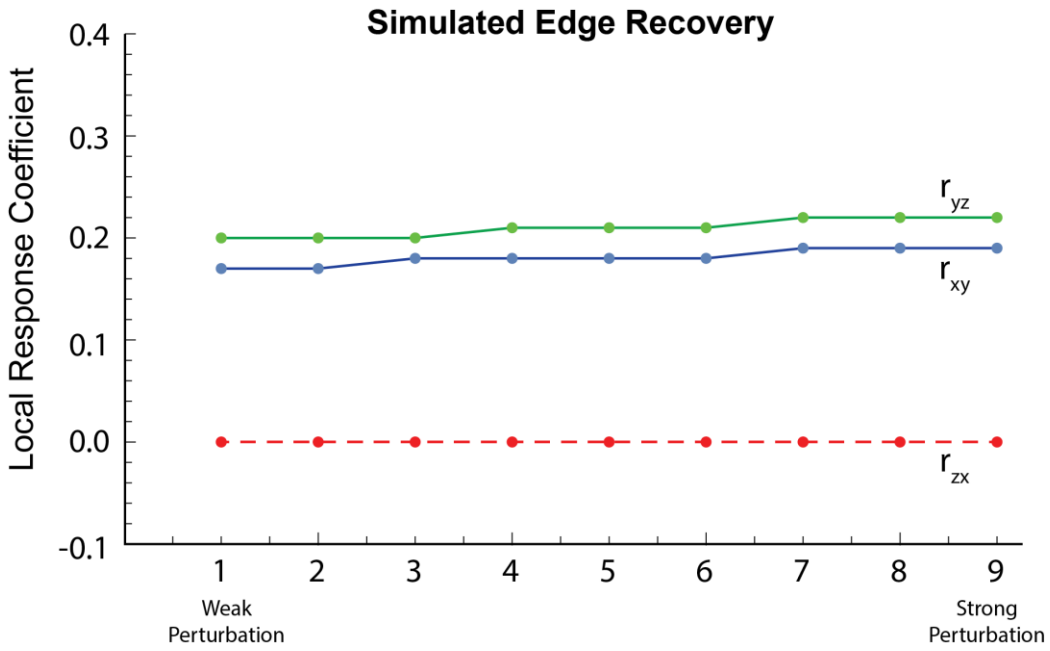


Figure 2.8: **Local response coefficients of simulated data.** Responses r_{yx} and r_{zy} were found to be positive while r_{zx} was found to be 0 for all perturbation strengths.

The fact that a negative edge was not recovered after perturbing a genetic cascade in silico, yet is numerically possible, indicates that the ghost edge may be a consequence of factors that are not explicitly modeled, including the assumed abundant shared resources.

2.6 Conclusions

Synthetic biological networks as benchmarks[1,2,25–27] have provided a unique opportunity to study biological reverse engineering methods. The reconstruction results of our circuit via MRA uncover a "ghost" regulation edge not readily apparent from experimental measurements and accordingly, requires further investigation of the cellular environment complexities. The emergent edge of the resource constrained cascade implicates the possible existence of a naturally formed genetic "dark network", where collections of hidden edges affect complex network dynamics in

unexplored and nuanced ways from the shadows of operational context. A necessary first step in uncovering the secrets of the dark genetic network would be to rigorously validate the existence of other hidden edges through modified inference methods and classify their interactions within visible genetic networks and among themselves.

2.7 References

1. Kang, T., Moore, R., Li, Y., Sontag, E. & Bleris, L. Discriminating direct and indirect connectivities in biological networks. in *Food, Pharmaceutical and Bioengineering Division 2015 - Core Programming Area at the 2015 AIChE Meeting* (2015). doi:10.73/pnas.1507168112/-/DCSupplemental

2. Kang, T. *et al.* Reverse engineering validation using a benchmark synthetic gene circuit in human cells. *ACS Synth. Biol.* (2013). doi:10.1021/sb300093y

3. Gardner, T. S., Di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science (80-.).* (2003). doi:10.1126/science.1081900

4. Marbach, D. *et al.* Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.0913357107

5. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol. Syst. Biol.* (2011). doi:10.1038/msb.2011.49

6. Shimoga, V., White, J. T., Li, Y., Sontag, E. & Bleris, L. Synthetic mammalian transgene negative autoregulation. *Mol. Syst. Biol.* (2013). doi:10.1038/msb.2013.27

7. Hayete, B., Gardner, T. S. & Collins, J. J. Size matters: Network inference tackles the genome scale. *Molecular Systems Biology* (2007). doi:10.1038/msb4100118

8. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* (2005). doi:10.1038/ng1532

9. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* (2009). doi:10.1016/j.cell.2009.01.042

10. Houseley, J. & Tollervey, D. The Many Pathways of RNA Degradation. *Cell* (2009). doi:10.1016/j.cell.2009.01.019

11. Spitz, F. & Furlong, E. E. M. Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics* (2012). doi:10.1038/nrg3207

12. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal* (2012). doi:10.1002/biot.201200085

13. Borkowski, O., Ceroni, F., Stan, G. B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology* (2016). doi:10.1016/j.mib.2016.07.009

14. de Vos, D., Bruggeman, F. J., Westerhoff, H. V. & Bakker, B. M. How molecular competition influences fluxes in gene expression networks. *PLoS One* (2011). doi:10.1371/journal.pone.0028494

15. Abedpour, N. & Kollmann, M. Resource constrained flux balance analysis predicts selective pressure on the global structure of metabolic networks. *BMC Syst. Biol.* (2015). doi:10.1186/s12918-015-0232-5

16. Rondelez, Y. Competition for catalytic resources alters biological network dynamics. *Phys. Rev. Lett.* (2012). doi:10.1103/PhysRevLett.108.018102

17. Mather, W. H., Hasty, J., Tsimring, L. S. & Williams, R. J. Translational Cross Talk in Gene Networks. *Biophys. J.* (2013). doi:10.1016/j.bpj.2013.04.049

18. Raveh, A., Margaliot, M., Sontag, E. D. & Tuller, T. A model for competition for ribosomes in the cell. *J. R. Soc. Interface* (2016). doi:10.1098/rsif.2015.1062

19. Qian, Y. & Vecchio, D. Del. Effective interaction graphs arising from resource limitations in gene networks. in *Proceedings of the American Control Conference* (2015). doi:10.1109/ACC.2015.7172024

20. Feizi, S., Marbach, D., Médard, M. & Kellis, M. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat. Biotechnol.* (2013). doi:10.1038/nbt.2635

21. Kholodenko, B., Yaffe, M. B. & Kolch, W. Computational approaches for analyzing information flow in biological networks. *Science Signaling* (2012). doi:10.1126/scisignal.2002961

22. Kholodenko, B. N. *et al.* Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. *Proc. Natl. Acad. Sci. U. S. A.* (2002). doi:10.1073/pnas.192442699

23. Sontag, E. D. Network reconstruction based on steady-state data. *Essays Biochem.* (2008). doi:10.1042/BSE0450161

24. Santos, S. D. M., Verveer, P. J. & Bastiaens, P. I. H. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat. Cell Biol.* (2007). doi:10.1038/ncb1543

25. Cantone, I. *et al.* A Yeast Synthetic Network for In Vivo Assessment of Reverse-Engineering and Modeling Approaches. *Cell* (2009). doi:10.1016/j.cell.2009.01.055

26. Marbach, D., Schaffter, T., Mattiussi, C. & Floreano, D. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *J. Comput. Biol.* (2009). doi:10.1089/cmb.2008.09TT

27. Menolascina, F., Di Bernardo, M. & Di Bernardo, D. Analysis, design and implementation of a novel scheme for in-vivo control of synthetic gene regulatory networks. *Automatica* (2011). doi:10.1016/j.automatica.2011.01.073

CHAPTER 3

PROBING TRANSCRIPTIONAL AND POST-TRANSCRIPTIONAL NOISE USING GENOME INTEGRATED ENGINEERED GENE CIRCUITS

Authors – Tyler Quarton, Vasileios Papakis, Yi Li, Khai Nguyen, Leonidas Bleris

Department of Biomedical Engineering

The University of Texas at Dallas

800 W Campbell Road

Richardson, Texas 75080-3021

3.1 Abstract

The production of proteins from genetic information that takes place within living cells is an inherently stochastic process. Every step in this cascade of biochemical reactions transpires with relatively few molecular reactants, resulting in the stochastic fluctuations of intermediary products. Moreover, the biological machinery that facilitates these reactions also varies from cell to cell. Ultimately, these series of random chemical processes compound into the observed phenomenon of "gene expression noise" where even genetically identical cells coexisting in the same environment exhibit variation in their protein levels. Here, we elucidate the impact that intermediary steps within the process of protein production have on gene expression noise by engineering custom human cell lines containing noise analyzing synthetic circuits that variably interface with the central dogma. We provide a generalized statistical framework capable of approximating the intrinsic and extrinsic noise of a population of cells expressing two-promoter systems even in the presence of imbalanced co-expression. We accompany our analysis with mathematical models of gene expression to validate and benchmark our decomposition framework. Herein, we find that gene expression is more intrinsically noisy and less extrinsically noisy when the co-expression of a two-reporter system is uncoupled earlier during the stages of the protein production process, where the converse holds true when the two reporters remain coupled until the later stages of protein expression.

3.2 Introduction

The genetic program responsible for the maintenance and operation of all living cells is executed by the collection of complex gene regulatory networks arising from genes interacting

with other genes and biomolecules within a cell. Incredibly, these naturally occurring networks robustly operate even as noisy genetic information is transduced[1]. The various biochemical sources that contribute to a gene's expression noise have been broadly categorized as either being intrinsic or extrinsic[2]. The collection of stochastic biochemical events that independently affect a gene's expression within a given cell are often called "intrinsic" or "local" noise sources. These variations can propagate along transcription network pathways where the quantitative distributions of a network's constituent proteins appear correlated. However, even proteins originating from unconnected regulation pathways show correlation. These apparently phantom correlations arises from the stochastic variations in the quantities of the reactive species that interact with of all genes (e.g. copies of housekeeping genes), and are commonly referred to as "extrinsic" or "global" noise sources. The presence of this unavoidable noise can create design constraints on improperly outfitted synthetically engineered gene circuits when deterministic outputs are desired. However, with intentional design gene circuits can operate in concert with noise to produce a wide variety of circuit functions, including stochastic state switching, circuit excitability and procrastination[3]. Biologically, gene expression noise has wide-ranging implications, including its involvement in the generation of phenotypic heterogeneity in microorganisms[4], development and cellular differentiation in higher organisms[5], and the progression of disease[6].

The categorical quantification of gene expression noise as either being intrinsic or extrinsic is intimately defined in the context of a synthetic two-reporter system[7]. The standard two-reporter design includes two, independently expressed genes absent of other regulations that produce measurable fluorescent proteins within a single cell. As both genes operate stochastically, their observed protein products exhibit variability. From a gene-centric perspective, the collection of

stochastic sources that jointly contribute to the variability of both genes are said to be extrinsic, whereas the sources of noise that independently affect both genes are said to be intrinsic. As a consequence, a strongly expressing constitutive promoter is expected to have little intrinsic noise, while a weak promoter will have high intrinsic noise. In addition, two identical, independently regulated promoters are expected to have the same extrinsic noise, which arises through global effects[2,7]. The two-reporter experimental setup designed for the intrinsic/extrinsic noise decomposition has been successfully used to study noise contributed by global effects and gene-specific effects, as well as pathway-specific[7,8].

A critical constraint of these experiments is the need for two identically regulated reporters with equal variance and mean expression[2]. To alleviate this constraint to accommodate increasingly complex two-reporter designs, we develop herein a statistical framework that generalizes the decomposition of two-reporter systems when nuanced imbalances between the expressions of the two reporters are present. Using this framework, we decompose the gene expression noise of an engineered a family of synthetic, noise analyzing cell lines designed to systematically identifying and characterizing propagative effects of the underlying protein production mechanisms of which collectively contribute to observed gene expression noise. Our engineering and theoretical approach leverages established experimental procedures to generate new temporal resolution of the propagative effects of expression noise as each engineered cell line measures protein variability by sequentially decoupling the expression of their co-expressed proteins throughout their shared biochemical production steps.

3.3 Results

3.3.1 Circuit Construction and Cell Line Panel Engineering

The panel of cell lines is comprised of a series of three engineered immortalized human colorectal carcinoma lines (HCT116) each consisting of unique synthetic gene cassettes that constitutively express a cyan fluorescent protein (CFP) and a far-red fluorescent protein (mKate2) (Figure 3.1A). Each fluorescent reporter protein is accompanied with a PEST peptide sequence, reducing each proteins half-life to sensitize their expression to noise.

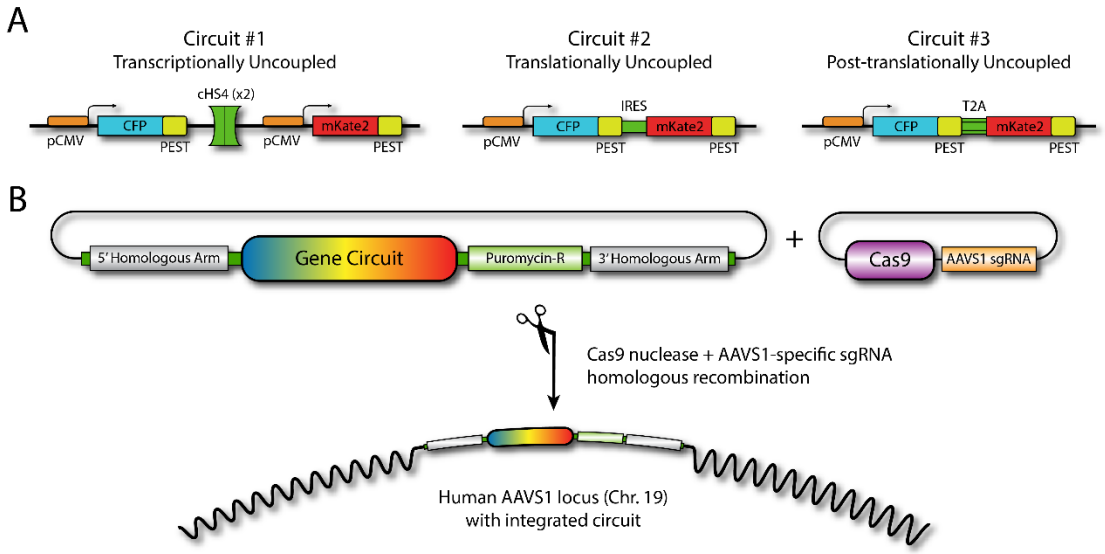


Figure 3.1: **Design and Integration of Noise Analyzing Circuits.** A) Three synthetic gene circuits that variably interface with the protein production process. The transcriptionally uncoupled circuit #1 produces CFP and mKate2 fluorescent proteins originating from two transcriptionally independently expressing genes separated by two insulating elements (cHS4). The translationally uncoupled circuit #2 produces both CFP and mKate2 fluorescent proteins originating from a single gene whose transcript contains two cistrons separated by an IRES element. The post-translationally uncoupled circuit #3 produces both CFP and mKate2 fluorescent proteins originating from a single translated polypeptide that self-cleaves into two functional proteins. B) Each circuit is stably integrated into the AAVS1 locus using Cas9-mediated homologous recombination.

Each cassette was stably integrated into the adeno-associated safe harbor locus (AAVS1) using CRISPR and interface the expression of the two fluorescent proteins with underlying biological machinery. Specifically, the transcriptionally uncoupled "*Circuit #1*" is comprised of a two constitutive minimum pCMV reporters that each independently transcribe CFP and mKate2, respectively. Both genes are separated by a series of two cHS4 insulator sites to ensure transcriptional independence given their proximity and shared orientation (citation). The translationally uncoupled "Circuit #2" features a single pCMV promoter whose transcript is comprised of both CFP and mKate2 cistrons separated by an internal ribosome entry site (IRES) (ref). What does IRES do? Finally, the post-translationally uncoupled "Circuit #3" features a T2A self-cleaving peptide that uncouples the co-expression of both reporters after their shared transcript is successfully transcribed and translated and allows for independent protein folding of both CFP and mKate2.

To perform genomic integrations, we engineered donor strands using standard recombinant DNA cloning to enable homology-directed repair (HDR) integration in the AAVS1 safe-harbor locus of immortalized human colorectal carcinoma cells (HCT116). Specifically, donor plasmids were designed to contain sequences homologous to the targets (Supplementary material,). The cassette of interest (carrying the desired two-reporter system), together with a puromycin resistance selection cassette, was separated by a cHS4 insulator and cloned within the two homologous arms (Figure 3.1B). We isolated 3 monoclonal cell lines for each of the 3 circuits (Supplementary material) and verified functional integration through microscopy and cytometry measurements (Figure 3.2).

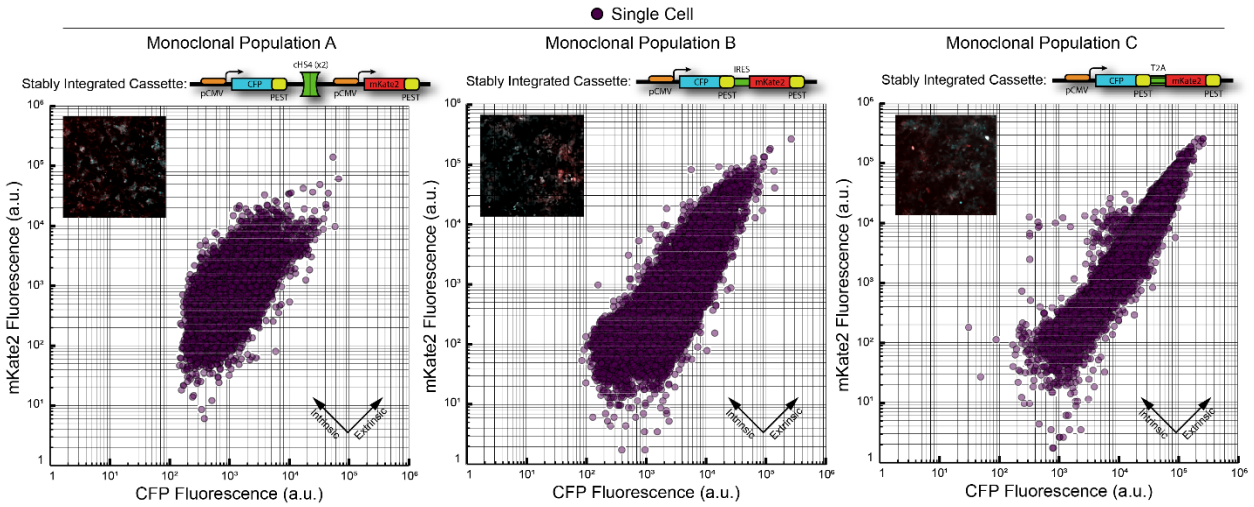


Figure 3.2: **Functional output of engineered lines.** The monoclonal populations of cells expressing circuit #1, #2, and #3 (left to right) accompanied by microscopy images that indicate functional co-expression of both fluorescent proteins in each system.

The three cassettes uncouple the output protein production at strategic locations as they are produced through the central dogma steps, allowing for the propagative effects of intrinsic and extrinsic noise to differentially manifest (Figure 3.3).

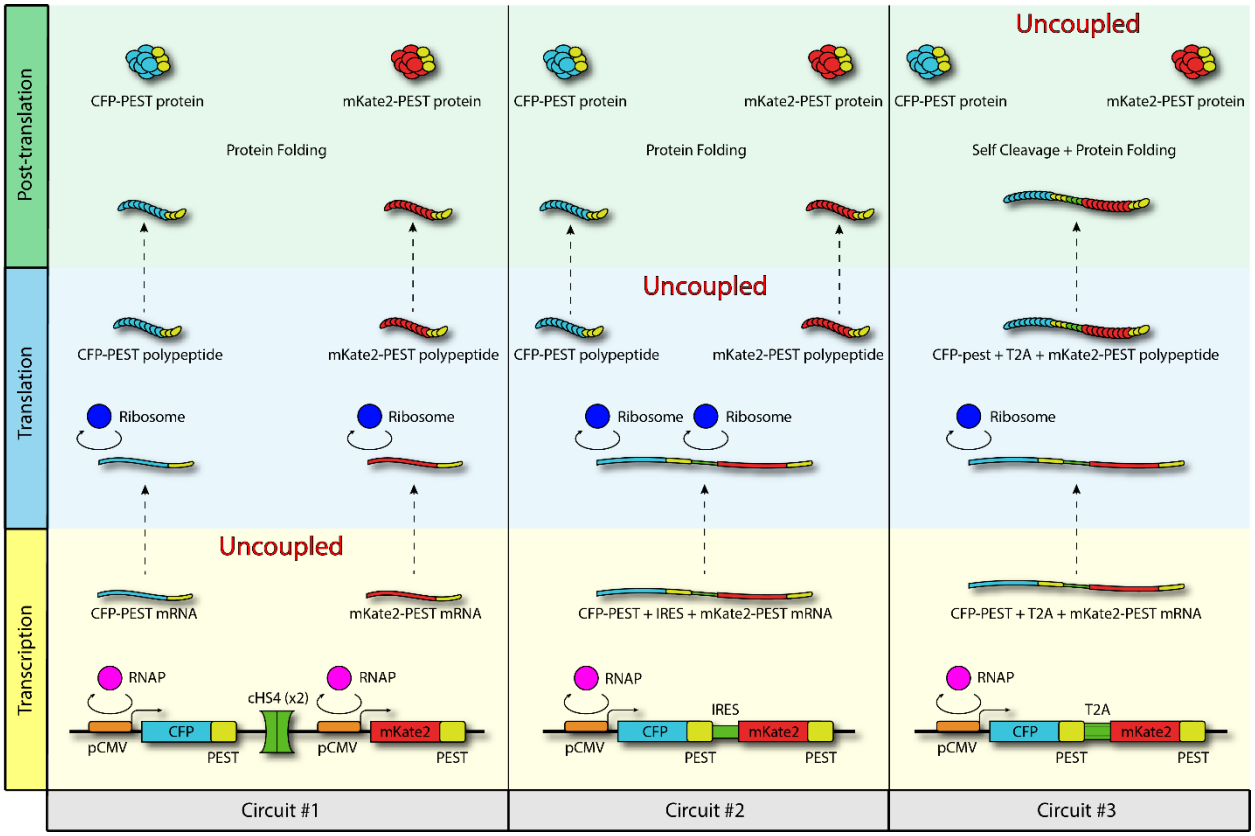


Figure 3.3: **Serial uncoupling of noise analyzing circuits.** Each circuit uncouples the expression of their respective fluorescent proteins at different production stages. In the transcriptional stage (yellow), the CFP and mKate2 dependent production elements are uncoupled in circuit #1, as two individual genes produce two individual transcripts, and coupled in circuits #2 and #3 as single genes express transcripts that contains both the fluorescent protein's cistrons. In the translation stage (blue), the CFP and mKate2 dependent production persist to be uncoupled for circuit #1, becomes uncoupled for circuit #2, as the internal ribosome entry site allows for a single transcript to produce two polypeptides, and remains coupled for circuit #3. In the post-translational stage (green) circuits #1 and #2 maintain the uncoupled co-expression of CFP and mKate2 while circuit #3 becomes uncoupled through the autocatalytic activity of the T2A element which separates the single transcribed polypeptide into two proteins.

The stable cell line containing circuit #1, independently expresses both reporter proteins ensured by the inclusion of the cHS4 insulator. At this stage of expression, both fluorescent proteins will be subjected to both global and local sources of noise within their own genes. In the case of circuit #2, both cistrons of CFP and mKate2 share the same promoter and transcript, enforcing any sources

of fluctuations that impinge either the promoter or transcript to be globally shared between both fluorescent proteins' expressions. The inclusion of the IRES element allows for the single transcript to interact with multiple ribosomes, enabling two translation events to take place within a single transcript thereby uncoupling the expression of both CFP and mKate2 at the translation stage by allowing local noise sources to independently affect CFP and mKate2 expression in addition to their shared global noise sources. Finally, the stable cell line containing circuit #2, maintains the coupling of mKate2 and CFP through the translation stage of protein production by replacing the IRES element of circuit #2 with a T2A element. Finally, the transcripts produced the stable cell line containing circuit #3, only have one ribosome initiation site in which a single transcript will be translated into a single polypeptide containing both CFP and mKate2 proteins temporarily fused together with the T2A self-cleaving polypeptide. Upon complete translation, the T2A element will cleave, uncoupling the expression of CFP and mKate2 by allowing each protein to be subjected to independent fluctuations after translation.

3.3.2 Generalized Noise Decomposition

Mathematically, extrinsic and intrinsic noise have been empirically defined through the covariant expression of two genes being expressed in symmetric two-reporter systems. As such, the application of these definitions of noise are limited to proportionally expressed and equivalently varying related reporter protein expressions. As an example, an expression bias towards the downstream protein bicistronically produced from IRES containing transcripts, as seen in our cell lines containing circuit #2, has been observed (). To accommodate for these kinds of biases present in non-deal two reporter systems in which the reporter genes of a two-reporter

system are dissimilarly processed and/or regulated, we broaden the mathematical definition of intrinsic noise using a total least squares regression approach.

Let $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ and $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$ be the ordered sets of n observations of two fluorescent proteins produced from a two-reporter noise analyzing system. We denote $\mu_{\mathbf{X}}$, $\mu_{\mathbf{Y}}$, $\sigma_{\mathbf{X}}$, $\sigma_{\mathbf{Y}}$ as the means and standard deviations of **X** and **Y** and $\sigma_{\mathbf{X},\mathbf{Y}}$ as the covariance between **X** and **Y**. In this notation, the extrinsic noise of an expressing population of cells is defined as

$$\eta_{ext} \equiv \frac{{\sigma_{\mathbf{X},\mathbf{Y}}}^{1/2}}{(\mu_{\mathbf{X}} \cdot \mu_{\mathbf{Y}})^{1/2}},$$

the root covariance of the observed two fluorescent genes scaled by their geometric mean, a natural dimensional extension of the coefficient of variation for two variables[7]. The intrinsic noise of a population has been defined as

$$\eta_{int} \equiv \frac{\sqrt{\frac{1}{2\,n}\sum_{i=1}^{n}|x_i - y_i|^2}}{(\mu_{\mathbf{X}} \cdot \mu_{\mathbf{Y}})^{1/2}}$$

where the intrinsic co-expression dispersion is captured by the root mean square distance of each ordered observation to the line of symmetric expression, $y = x$, also scaled by the fluorescent proteins' geometric mean[7]. Unlike extrinsic noise, this definition of dispersion about line of symmetric expression limits its use to populations that exhibit ideal co-expression, where the means and standard deviations of **X** and **Y** should be approximately equal. Any deviation from this ideal will result in an over approximation of intrinsic noise as the dispersion would no longer be about the central expression axis of the population (Figure 3.S1). Here, we generalize the definition of intrinsic noise to take into account the co-expressing population's position and slope in the co-expression space. To this end, we redefine the intrinsic dispersion as the root mean square

distance of each ordered observation to the line of total least squares regression. Under the assumption that the marginal distributions of both **X** and **Y** are normally distribute in the domain of observation, the line defined in the total least squares regression in two dimensions, also called Deming regression, in terms of the population statistics is

$$y = \frac{\sigma_{\mathbf{Y}}^2 - \sigma_{\mathbf{X}}^2\delta + \sqrt{(\sigma_{\mathbf{Y}}^2 - \sigma_{\mathbf{X}}^2\delta)^2 + 4\delta\sigma_{\mathbf{X},\mathbf{Y}}^2}}{2\,\sigma_{\mathbf{X},\mathbf{Y}}}(x - \mu_{\mathbf{X}}) + \mu_{\mathbf{Y}}\,, \tag{1}$$

where $\delta = \frac{\sigma_{\mathbf{X}}^2}{\sigma_{\mathbf{Y}}^2}$. Thus, the intrinsic noise defined as the dispersion about the regression line (1) is

$$\eta_{int} \equiv \frac{1}{(\mu_{\mathbf{X}} \cdot \mu_{\mathbf{Y}})^{\frac{1}{2}}}\sqrt{\frac{\alpha^2\sigma_{\mathbf{X}}^2 - 2\alpha\beta\sigma_{\mathbf{X},\mathbf{Y}} + \beta^2\sigma_{\mathbf{Y}}^2}{\alpha^2 + \beta^2}}\,,$$

the root mean square distance of each ordered observation to the total least square regression line scaled by the fluorescent proteins' geometric mean where $\rho_{\mathbf{X},\mathbf{Y}}$ is the correlation coefficient of **X** and **Y** and α and β are introduced for cleanliness (Supplement Section Derivation of Generalized Intrinsic Noise).

3.4 Stochastic Population Models

To further investigate the propagation effects of extrinsic and intrinsic noise within the central dogma, we construct three stochastic models that are representative of our gene circuits. To introduce noise, we follow the error-in-variables (EIV) modeling approach and tandemly add two noise sources to the kinetics of the central dogma (Supplemental Material, Stochastic Modeling). At each production stage, the protein precursor species were subjected to two sources of noise; an extrinsic noise, which additively perturbed the production kinetic parameters of both protein precursor species identically and intrinsic noise, which additively perturbed both protein

precursor species independently. Procedurally, both intrinsic noise parameters within a given production stage are paired random sampling events from a Gaussian distribution with mean 0 and a defined standard deviation in which two unique values are perturb their respective protein precursors kinetic parameter independently. On the other hand, the single extrinsic noise parameter within a given production stage is a unique random sampling from a Gaussian distribution with zero mean and a defined standard deviation in which both kinetic parameters of each respective protein precursor are perturbed simultaneously and identically (Figure 3.4).

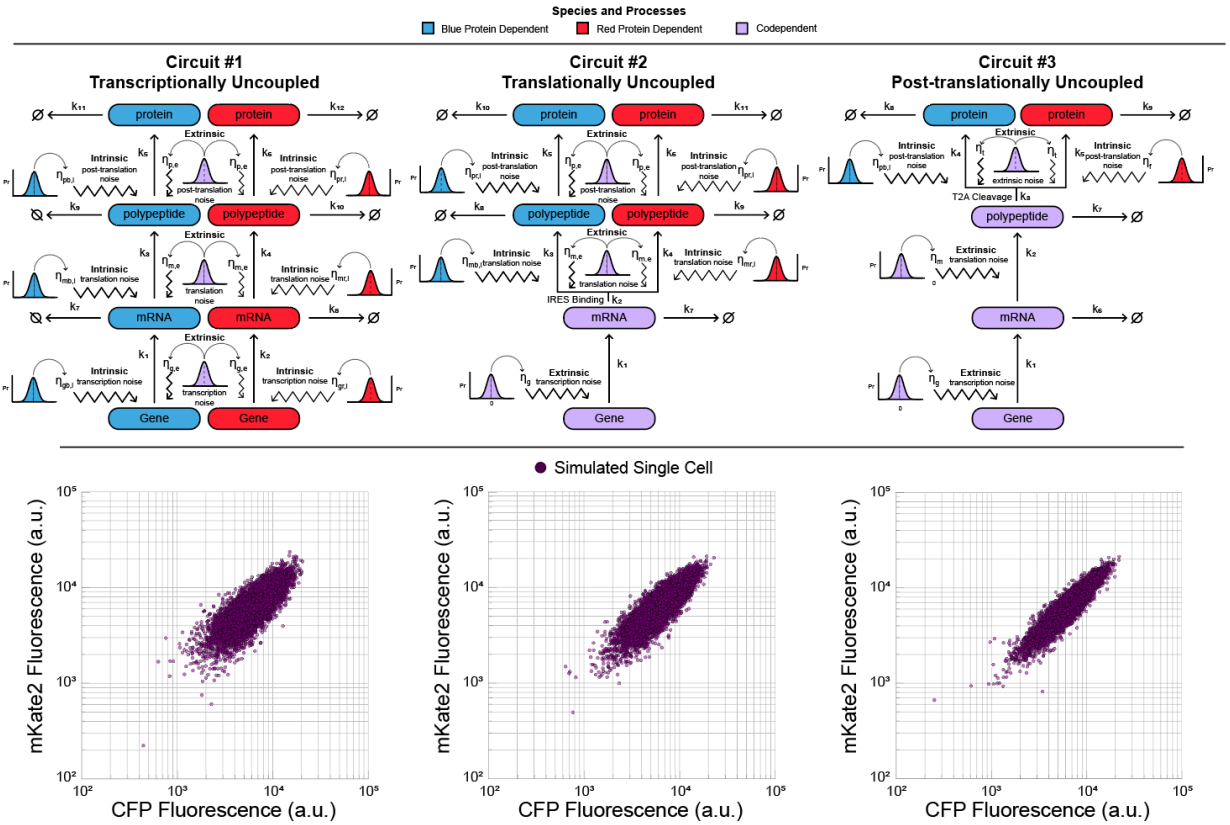


Figure 3.4: **Stochastic population modeling and simulation results.** A) Diagrammatic representation of error in variables ODE model of each synthetic architecture. Each random sampling of an error variable is indicated by a representative Gaussian distribution. The intrinsic noise sources at each production stage are sampled simultaneously, whereas the extrinsic noise sources are sampled once and applied to both noisy kinetic parameters within a given stage. The reaction species and sampling events that are solely responsible for the production of mKate2 or CFP are indicated in red and blue, respectively, whereas the reaction species and sampling events that are jointly responsible for the production of mKate2 and CFP are indicated in light purple. B) Simulation results of stochastic population models. The magnitude of noise for each intrinsic sampling was set vary within 10% of the parameter it affects, whereas the extrinsic sampling was set to vary within 20% of the parameter it affects in each simulation.

The variance of each noise distribution is chosen to scale with the parameter it is perturbing such that the noise of the parameter, as measured by the coefficient of variation (CV), remains constant throughout the production stages of the simulation (Supplementary Material). For our simulations we chose our intrinsic noise to generate a 10% variation and our extrinsic noise to

generate a 20% variation about the parameter(s) they are affecting, respectively. We chose the magnitudes of these variation to best represent a fully induced mammalian two-reporter system as seen in a previous publication[9].

A population of *N* cells is generated by sequentially evaluating the steady state co-expression of both fluorescent proteins for a given stochastic model *N* times. The population is then decomposed using the generalized noise decomposition as detailed above. For each circuit model, we simulated 100 populations of 1000 cells.

3.5 Decomposition Results

Both the experimentally engineered populations and their analogous simulated populations were decomposed following the generalized noise decomposition method. We find that as the coupling of the two-fluorescent proteins persists through the production stages of the central dogma, the intrinsic noise increases as the extrinsic noise decreases for both the experimental and simulated populations (Figure 5).

We reason that as a production step becomes coupled, two individual protein precursors susceptible to independent noise fluctuations merge into a single precursor responsible for the eventual production of both proteins. This causes any intrinsic noise sources that would have affected both independent precursor specie (had they been uncoupled) to only affect a single precursor. Furthermore, this single precursor now experiences both intrinsic and extrinsic fluctuations indistinguishably as there is no longer a companion precursor independently fluctuating within the given production stage. This causes a combined increase in the observed extrinsic fluctuation as both the biological sources of the intrinsic and extrinsic fluctuations remain present yet are now indistinguishable. The resulting shared fluctuation then proportionally

propagates through the remaining production stages, causing an eventual correlation between the final protein products to manifest.

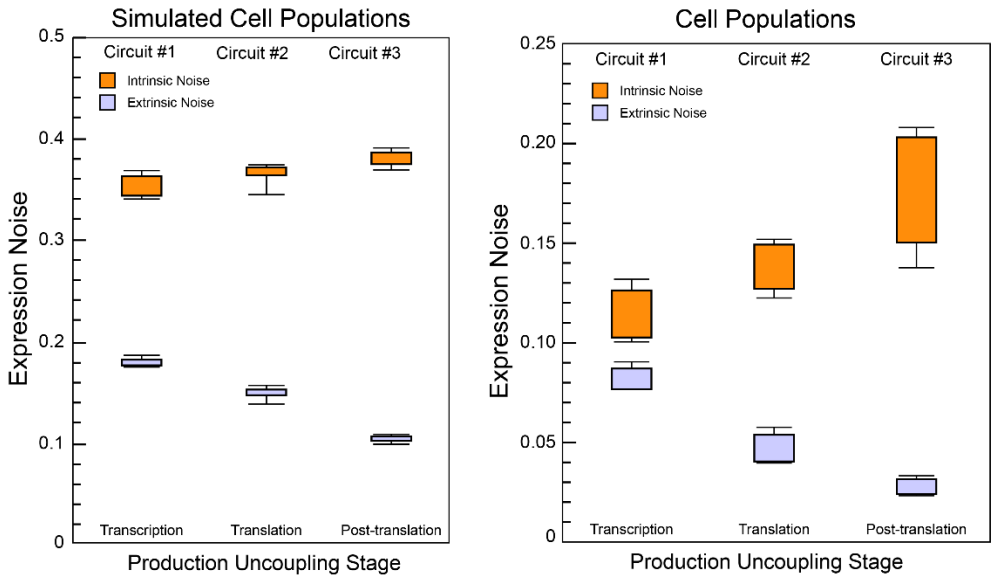


Figure 3.5: **Noise Decomposition.** Extrinsic (blue) and intrinsic (orange) noise values obtained monoclonal populations (left) and stochastic population simulations (right). The box height indicates the 75% (top) and 25% (bottom) quantiles and the whiskers bound the upper and lower range of values observed. The experimental noise was obtained from the decomposition of three monocolonal populations of each circuit-containing population. The experimental noise was obtained from 100 simulations of 1000 stochastic single cells for each circuit model. As the stage in which the two fluorescent proteins become uncoupled progresses, the intrinsic noise increases as the extrinsic noise decreases.

In the case the populations that contain the transcriptionally uncoupled circuit, the independent protein precursors for each central dogma production stage exists allowing for the most opportunities for intrinsic fluctuations resulting in the observation of a high intrinsic noise and lower extrinsic noise of a population. In the case of the populations that contain the translationally coupled circuit, the translation stage protein precursors are merged, such that any fluctuation in the translation stage is eventually experienced by both proteins, increasing the extrinsic noise and decreasing the intrinsic noise. Similarly, in the case of the post-translationally

uncoupled circuit, the merger of the all independent protein precursors of each stage into shared precursors results in the observation of an even more correlated co-expression (higher extrinsic noise) and a lower intrinsic noise.

3.6 Methods and Procedures

3.6.1 Molecular Cloning

All restriction enzymes, as well as Q5 High-Fidelity 2X Master Mix (NEB, catalog #M0492) and T4 DNA Ligase (NEB, catalog #M0202S) were purchased from New England Biolabs. All primers were synthesized by Sigma Aldrich. PCR and Gel Purification were performed using PCR Purification (QIAGEN, catalog #28104) and QIAquick Gel extraction kits (QIAGEN, catalog #28704) according to manufacturers' protocols. All intermediate and final plasmids were transformed into NEB-5alpha competent E. coli (NEB, catalog #C2987H).

For the assembly of all genetic constructs a donor plasmid flanked by 1kb homology arms to the AAVS1 safe harbor locus was used. Plasmid integrity was confirmed with direct Sanger sequencing and restriction enzyme digestion. Sanger sequencing was carried out by Genewiz. Bacterial culture media and agar were purchased from BD Biosciences (BD Biosciences, catalog #DF0446-07-5). Miniprep was performed using QIAprep Spin Miniprep kit (QIAGEN, catalog #27104).

3.6.2 Cell Culture and Transfection

HCT116 cells were gifts from Dr. Michael A. White (University of Texas Southwestern Medical Center), cell lines were maintained at 37C, 100% humidity and 5% CO_2. Cell lines were

grown in Dulbecco's modified Eagle's medium (Invitrogen, catalog #11965-1181) that was supplemented with 10% Fetal Bovine Serum (Invitrogen, catalog #26140), 0.1 mM MEM non-essential amino acids (Invitrogen, catalog #11140-050) and 0.045 units/mL of Penicillin and 0.045 units/mL of streptomycin (Invitrogen, catalog #15140). When 50%-90% confluent, cells were passed by washing with Dulbecco's Phosphate Buffered Saline (Mediatech, catalog #21- 030-CM), then trypsinizing with 0.25% Trypsin with EDTAX4Na (Invitrogen, catalog #25200) and diluted in fresh medium.

For transient transfection, approximately 150,000 cells were seeded into each well of 12-well cell culture treated plastic plates (Griener Bio-One, catalog #665180) and grown for 20-24 hours. For transfections using jetPRIME as transfection agent 1µg of DNA was added to 75ml jetPRIME buffer and 1.75µl of jetPRIME (Polyplus, catalog #114-01). Each transfection solution was incubated for 10 minutes and added to the cells and mixed with gentle shaking.

For generation of stable cell lines, approximately 10 million human HCT116 cells were seeded onto a 10cm petri dish. After 16 hours, cells were transiently transfected with 1µg of each construct and 10µg of pCMV-spCas9-t2a-mKate-U6-gRNA using jetPRIME as transfection agent. 48 hours later puromycin was added to the final concentration of 2µg/ml (ThermoFisher Scientific, catalog #A1113802). Selection lasted roughly two weeks and generated polyclonal stable cell lines for each construct.

To generate monoclonal stable cell lines, FACS was performed through the UTD Flow Cytometry or Cell Characterization core. For FACS, the core utilizes BD Aria Fusion flow cytometer. FSC (forward scatter)/SSC (side scatter) gates were generated using wild type HCT116. Single cells were grown in 96-well plates (Griener Bio-ONE, catalog #650101) using cultured

media. Cultured media was prepared for each distinct cell line by harvesting their media when fully confluent and passing it through a syringe filter (Fisher Scientific, catalog #50-202-072).

3.6.3 Flow Cytometry

For each cell line, approximately 150,000 cells were seeded into each three wells of the 12-well cell culture treated plastic plates. After 1 day, the cells were trypsinized with 0.1ml of 0.25% Trypsin-EDTA at 37oC for 3 minutes. The Trypsin-EDTA was neutralized using 0.9ml of complete medium. The cell suspension was centrifuged at 1000 rpm for 5 minutes and, after the removal of supernatant, the cell pellets were re-suspended in 0.5ml of Dulbecco's Phosphate Buffered Saline (Mediatech, catalog #21- 030-CM).

The cells were analyzed on a BD LSRFortessa flow analyzer. TagCFP was measured with a 445-nm laser and a 470/20 band-pass filter and mKate2 with a 561-nm laser, 600 emission filter and 610/20 band-pass filter. For all experiments 100,000 cells were collected and were performed in triplicates. A FSC (forward scatter)/SSC (side scatter) gate was generated using wild type HCT116 cells. The mKate and CFP readings from wild type HCT116 cells were set as baseline values and were subtracted from all other experimental samples. The data were processed in FlowJo and Mathematica.

3.6.4 Microscopy

Fluorescence microscopy was carried out using an Olympus IX81 microscope in a Precision Control environmental chamber. The images were captured using a Hamamatsu ORCA-03 Cooled monochrome digital camera. The filter sets used were ET436/20x (excitation) and ET480/40 m (emission) for CFP, ET560/40x (excitation) and ET630/75 m (emission) for mKate.

Slide book was used to process and analyze the images. All images within a given experimental set were collected with the same exposure times and underwent identical processing.

3.7 References

1. MacNeil, L. T. & Walhout, A. J. M. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research* (2011). doi:10.1101/gr.097378.109

2. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* (2002). doi:10.1073/pnas.162041399

3. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* (2010). doi:10.1038/nature09326

4. Booth, I. R. Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *Int. J. Food Microbiol.* (2002). doi:10.1016/S0168-1605(02)00239-8

5. Paldi, A. Stochastic gene expression during cell differentiation: Order from disorder? *Cellular and Molecular Life Sciences* (2003). doi:10.1007/s00018-003-23147-z

6. Cook, D. L., Gerber, A. N. & Tapscott, S. J. Modeling stochastic gene expression: Implications for haploinsufficiency. *Proc. Natl. Acad. Sci. U. S. A.* (1998). doi:10.1073/pnas.95.26.15641

7. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (80-.).* (2002). doi:10.1126/science.1070919

8. Pedraza, J. H. & Van Oudenaarden, A. Noise propagations in gene networks. *Science (80-.).* (2005). doi:10.1126/science.1109090

9. Shimoga, V., White, J. T., Li, Y., Sontag, E. & Bleris, L. Synthetic mammalian transgene negative autoregulation. *Mol. Syst. Biol.* (2013). doi:10.1038/msb.2013.27

CHAPTER 4

MAPPING THE OPERATIONAL LANDSCAPE OF MICRORNAS IN SYNTHETIC GENE CIRCUITS[†]

Authors – Tyler Quarton[1,2], Kristina Ehrhardt[1,2], James Lee[3], Srijaa Kannan[4], Yi Li[1,2], Lan Ma and Leonidas Bleris[1,2,]

Department of Biomedical Engineering

The University of Texas at Dallas

800 W Campbell Road

Richardson, Texas 75080-3021

---

[1] Bioengineering Department, University of Texas at Dallas, Richardson, TX, USA

[2] Center for Systems Biology, University of Texas at Dallas, Richardson, TX, USA

[3] Department of Biological Sciences, University of Texas at Dallas, Richardson, TX USA

[4] School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX, USA

[†] Quarton, T., Ehrhardt, K., Lee, J., Kannan, S., Li, Y., Ma, L., & Bleris, L. (2018). Mapping the operational landscape of microRNAs in synthetic gene circuits. NPJ systems biology and applications, 4(1), 6.

4.1 Abstract

MicroRNAs are a class of short, noncoding RNAs that are ubiquitous modulators of gene expression, with roles in development, homeostasis, and disease. Engineered microRNAs are now frequently used as regulatory modules in synthetic biology. Moreover, synthetic gene circuits equipped with engineered microRNA targets with perfect complementarity to endogenous microRNAs establish an interface with the endogenous milieu at the single-cell level. The function of engineered microRNAs and sensor systems is typically optimized through extensive trial-and-error. Here, using a combination of synthetic biology experimentation in human embryonic kidney cells and quantitative analysis, we investigate the relationship between input genetic template abundance, microRNA concentration, and output under microRNA control. We provide a framework that employs the complete operational landscape of a synthetic gene circuit and enables the stepwise development of mathematical models. We derive a phenomenological model that recapitulates experimentally observed nonlinearities and contains features that provide insight into the microRNA function at various abundances. Our work facilitates the characterization and engineering of multi-component genetic circuits and specifically points to new insights on the operation of microRNAs as mediators of endogenous information and regulators of gene expression in synthetic biology.

4.2 Introduction

MicroRNAs (miRNAs) are endogenously expressed in animals, plants, and viruses and regulate the expression of nearly 30% of all protein-coding genes[1,2]. The means by which miRNA mediate protein synthesis primarily depends on the degree of complementarity between the seed

sequence of the miRNA and its target, located in messenger RNAs (mRNAs)[3,4]. Once processed, a mature miRNA embedded within an RNA-induced silencing complex (RISC) guides the complex to a mRNA and binds to its corresponding target through Watson-Crick base pairing[5]. In the case of perfect or near perfect complementarity as observed in plants, the endonucleolytic activity of an Argonaute protein residing within the RISC complex initiates resulting in the cleavage of the mRNA transcript through the RNA interference (RNAi) pathway[6,7]. In animals, the vast majority of miRNA-guided RISC complexes bind to partially complementary mRNA target sequences. This partial complementarity still causes mRNA destruction through a variety of means including the blocking of translation initiation, recruitment of translation blockers, the deadenylation of the 3' untranslated region, and/or the decapping of the 5' untranslated region[8]. Quantitative proteomic profiling in human cells reinforced the view that the majority of all miRNA-based protein regulation in mammals was due to mRNA destabilization[9] whereas only 11-16% of miRNA-based regulation was shown to be caused by ribosomal translational efficiencies[10].

Gene networks have evolved to be able to exploit their stochastic environment to their operational advantage[11] where miRNA are thought to have important role in conferring robustness to endogenous processes[12]. Mathematical models that probe the properties of miRNA regulation have unveiled a collection of nuanced regulatory effects[13–15]. Experimentally, gene circuits equipped with endogenous miRNA sensors demonstrate that both perfectly and partially complementary miRNAs reduce noise for low expression genes, but increased noise for high expression genes[16]. Regulatory miRNA-based feedforward and feedback loops, frequently observed motifs in mammalian cells[17], are shown to provide genetic template adaptation[18], dosage compensation[19], and noise buffering[20,21] properties. In addition to the more subtle aforementioned

network regulatory effects, miRNA that participate in positive feedback loops have been show to directly control the p53-MDM2 core[22] and state switching between epithelial and mesenchymal states in metastasis[23]. At an additional layer of complexity, the network of miRNAs and their target mRNAs allows for cross-regulation between different endogenous mRNAs by sequestering the shared miRNAs. These sequestering properties are largely determined by the relative abundance and the binding strength of miRNA and mRNAs[24].

The adoption of RNAi in synthetic biology led to significant advancements in gene circuit functionality[25–28]. Specifically, engineered synthetic miRNAs are frequently used as regulators in gene circuits[29–33]. Additionally, gene circuits equipped with passive miRNA sensing elements establish an interface between a computing system and endogenous miRNAs. The first demonstration of an interface between siRNAs and engineered RNAi targets established a universal logic evaluator capable of computing conjunctive and disjunctive normal form expressions with up to five input variables[34]. Subsequently, endogenous miRNAs were used as inputs in complex sensor/computation architectures that differentiate between cancer and healthy cells[35]. A molecular diagnostics circuit was engineered to sense both endogenous miRNA and transcription factor abundances and perform logic operations[36]. There are a number of recent papers that explore a combination of theoretical and experimental avenues to optimize design parameters for genetic circuits operating in cells[37–41].

As a general observation, the relative abundance of miRNA and their associated target mRNA markedly impacts protein repression and genetic circuit functionality. Here, we shed new light on the relationship between input genetic template abundance, miRNA concentration, and output under miRNA control. We use the complete operational landscape of a synthetic gene

circuit and we derive a phenomenological model that provides insight into the miRNA function at various abundances.

4.3 Results

We engineered a custom genetic circuit that consists of the following components (Figure 4.1a): (a) an inducible synthetic miRNA co-expressed with an mKate2 fluorescent protein, (b) the constitutively produced fluorescent protein TagYFP (YFP) containing synthetic miRNA targets, and (c) the constitutively produced fluorescent protein TagCFP (CFP). mKate2 and its accompanying synthetic miRNA-FF4[18,42] are under direct transcriptional control via an inducible TRE promoter. The TRE promoter is activated by the endogenously expressed transcription factor rtTA complexed with a small molecule, doxycycline (Dox). On the same plasmid, we engineered the miRNA-FF4 sensing element by placing three adjacent FF4 targets[34] in the 3' untranslated region (UTR) of the fluorescent protein YFP, which is constitutively produced by the cytomegalovirus (CMV) promoter. The plasmid also harbors a constitutively produced fluorescent protein CFP. The CFP levels primarily depend on the plasmid copy number while the mKate2 fluorescence is proportional to miRNA-FF4 levels, as they are produced from the same transcript (Supp. Figure 4.S1). As a control, we engineered an additional circuit which was an exact replicate of the fully operational circuit in all aspects apart from the exclusion of miRNA-FF4 targets on the 3' UTR of YFP (Supp. Figure 4.S2).

The expression of the fluorescent proteins can be viewed as a collection of interdependent functions whose arguments are composed of quantities corresponding to the various abundances of active circuit elements (Figure 4.1b). From this perspective, CFP fluorescence is functionally mapped to plasmid copy number, mKate2 fluorescence is a function of both Doxycycline

concentration and plasmid copy number, and YFP fluorescence is a function of both miRNA concentration and plasmid copies. From this function-based vantage point, we aim to quantitatively map the complete three-dimensional fluorescent expression space, or operational landscape, of the plasmid's output with respect to its constituent circuit elements (Figure 4.1c) and decouple the respective contributions.

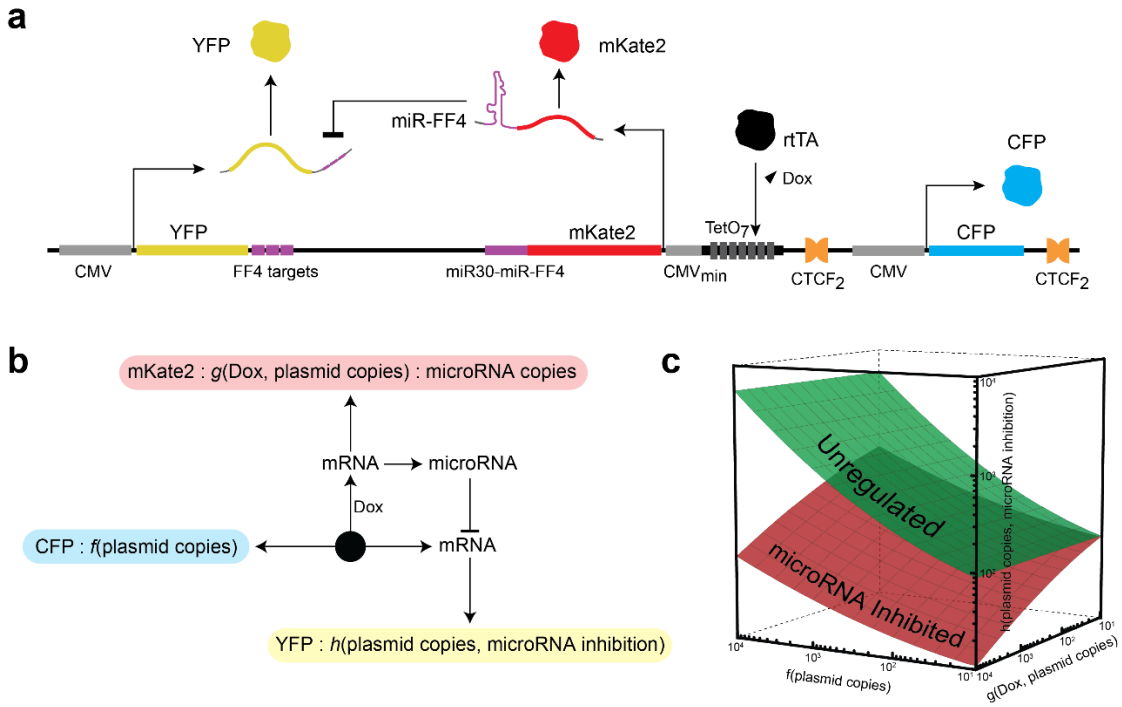


Figure 4.1: **Circuit architecture and function.** a) Biological schematic of the miRNA repression circuit. The CFP fluorescent protein is constitutively produced and is used to quantify the plasmid copy number. The synthetic miRNA-FF4 is produced in response to Dox and is quantified by mKate2 fluorescence. The miRNA-FF4 represses the mRNA of the output fluorescent protein, YFP. b) Functional graph representation of the circuit. Each observed fluorescent quantity is assumed to be a function of the underlying activity of the circuit elements. c) Visual representation of function-dependent output space of circuit. The green and red surfaces represent the expected output space of the circuit in the absence and presence of miRNA regulation, respectively.

In order to probe the impact of plasmid and miRNA copy number to the expression of the output protein, we partitioned the three-dimensional space generated from flow cytometry

measurements of the circuit's three fluorescent proteins. We quantitatively characterize the impact of plasmid copy number and miRNA concentration on the regulated expression of the output protein by exploring the features resulting from the mapped output surface in this expression space. The approach we employed generated a broadly-bounded expression space with fine-grained resolution which facilitated the development of a descriptive phenomenological model of the circuit output.

We performed a Dox titration (0.001-1 ng/mL) in HEK293 Tet-On (HEK293t) cells which contain a constitutively produced rtTA stably integrated into their genome. Approximately 150k cells were plated per well and were transiently transfected 24 hours later. Immediately following transfection, the cells were induced with Dox and were allowed to grow an additional 48 hours before obtaining measurements using flow cytometry.

An initial qualitative assessment of the cytometry data in a three-dimensional fluorescence space for the selected concentrations of Dox generally confirmed our expectations for mKate2 and CFP, where mKate2 was seen to be proportionally expressed with respect to CFP, while also revealing a unique copy number and miRNA concentration dependent YFP (output) expression (Figure 4.2) that was not observed in the control (Supp. Figure 4.S3).

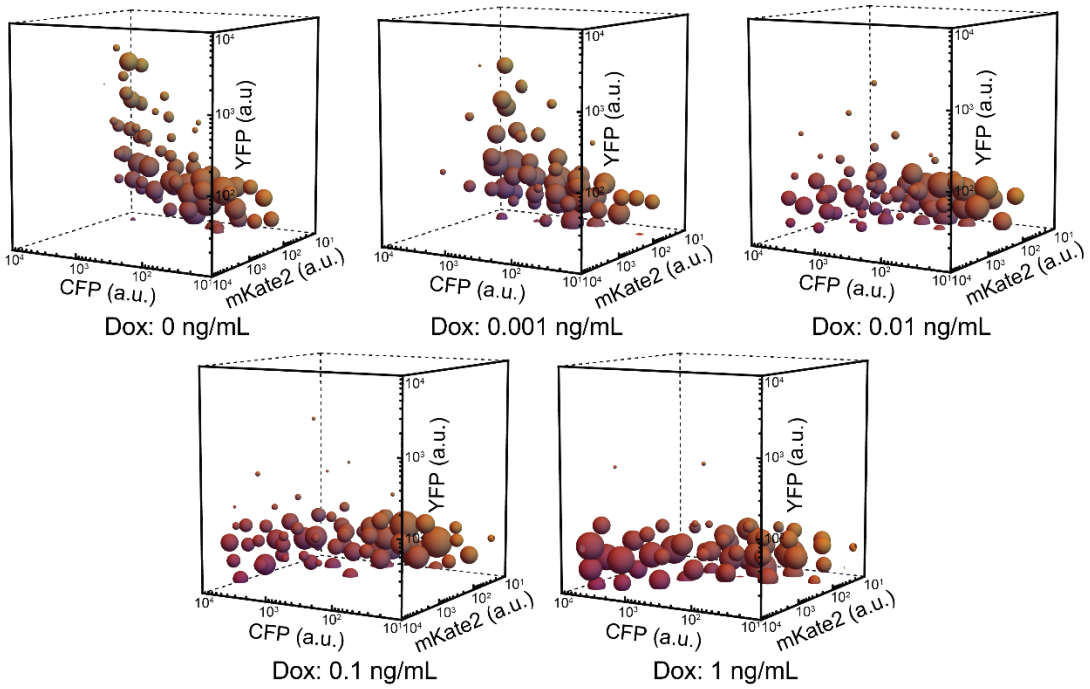


Figure 4.2: **Dox titration cluster plots.** Each sphere is a cluster of cells resulting from the k-means clustering algorithm. The color of a cluster is a blend of yellow, red, and blue in proportion to the cluster's mean values of YFP, mKate2, and CFP fluorescent arbitrary units, respectively. The diameter of a cluster is proportional to the number of cells it contains.

4.3.1 Pairwise Expression Relationships

To understand how the combined influences of plasmid copies and miRNA concentration effect the output protein's expression, we examine the pairwise relationships between each of representative fluorescent signals (Figure 4.1c). In order to discretize CFP fluorescence to define the plasmid copies, we partitioned the cytometry data into fixed bins of CFP fluorescence intensity (100 arbitrary units) and statistically averaged the YFP and mKate2 expression values of the cells collected in each bin, respectively. For example, the first bin, whose width corresponds to CFP fluorescent values ranging from 0-100 a.u. collects all the single cell whose observed CFP expression falls within the aforementioned range. These cells, which also express YFP and

mKate2, are reduced to a single data point by averaging the observed expressions of YFP and mKate2. This procedure is repeated for each bin of CFP fluorescence with a minimum of 100 cells to ensure reliable statistics. Effectively, this binning process condenses the three-way fluorescence expression of single cell sub-populations into representative data points partitioned with respect to their plasmid copy number[18].

The binned data confirmed that miRNA-FF4 concentration and plasmid copies share a linear relationship at all Dox concentrations in both the circuit and its control (Figure 4.3a**,** Supp. Figure 4.S4). We observed the YFP fluorescence trending from an initially uninhibited linear response at 0 ng/mL Dox to a saturating-like response at 1 ng/mL Dox, where at high plasmid copies, the output's expression becomes increasingly invariant to plasmid copy numbers[18] (Figure 4.3b, Supp. Figure 4.S5) where a strict linear response was observed in the control (Supp. Figure 4.S6).

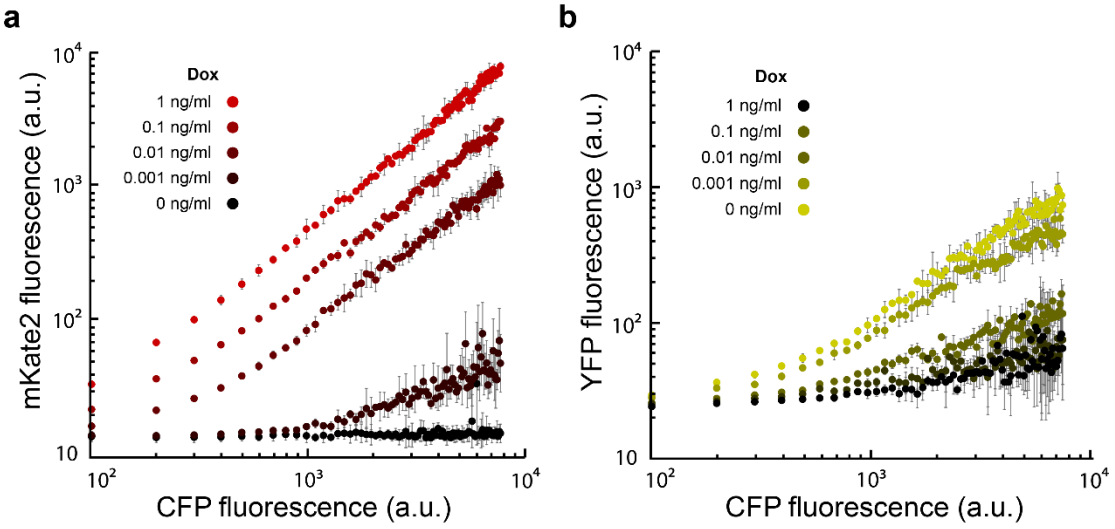


Figure 4.3: **Pairwise Output Circuit response.** a mKate2 fluorescence increases as CFP increases in a Dox dependent manner. b YFP has a nonlinear saturation at higher CFP fluorescence where the degree of saturation is Dox dependent. The error bars correspond to the standard deviation of an experimental triplicate.

The relative expression of mKate2 in each bin grew exponentially as a function of increasing Dox for both the circuit and its control whereas the relative expression of YFP exhibited saturation in the absence of Dox (Figure 4.4a, Supp. Figure 4.S7). Additionally, we confirmed the output decrease in the presence of increasing miRNA concentration, where the output was seen to be inversely proportional to miRNA concentration at a given plasmid copy number (Figure 4.4b) as compared to the output's miRNA independence observed in the control (Supp. Figure 4.S8).

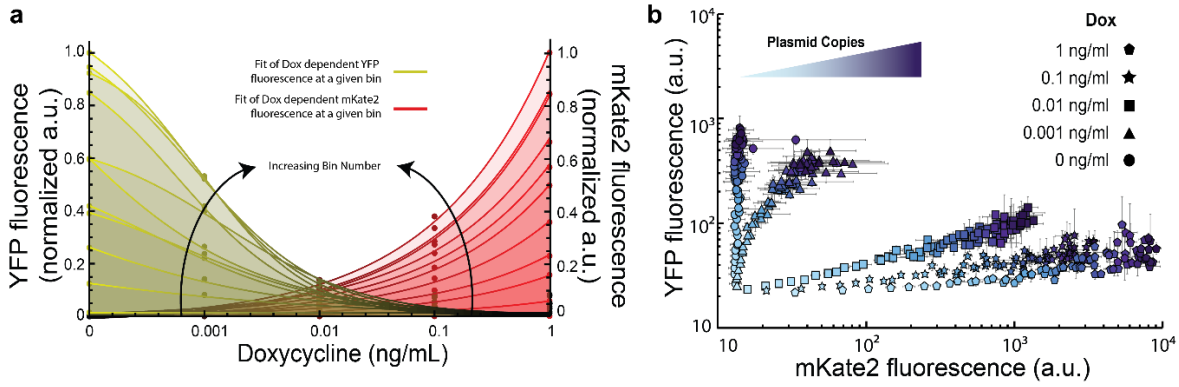


Figure 4.4: **Circuit response.** a) Relative YFP and mKate2 fluorescence in response to Dox. 10, linearly spaced CFP bins ranging from 100 a.u. to 104 a.u. were selected at each Dox concentration for both YFP and mKate2, depicted as dots. The fluorescent values of YFP and mKate2 were then normalized with respect to the highest CFP bin. Bins corresponding to the same CFP fluorescence throughout each Dox concentration were then fit. The relative expression of YFP follows a saturating function where a saturation effect at 0 Dox is observed whereas the relative expression of mKate2 exponentially increases as a function of Doxycycline. b) YFP is overall suppressed as Dox increases where, at a given copy number, YFP is inversely proportional to mKate2. The error bars correspond to the standard deviation of an experimental triplicate.

4.3.2 miRNA Repression Efficiency

In order to examine the copy number dependence on miRNA-based output suppression, we compared the relative change of mean regulated output expression of each of the Dox induced cells as compared to the uninduced state at each copy number bin. The above-mentioned binning procedure was applied to both the circuit and its corresponding control.

As an example, the efficiency of miRNA repression at the lowest copy numbers, i.e. the subpopulation of cells which fell within the first bin, was calculated by taking the ratio of the circuit's YFP value at its first bin to the control's YFP value at its first bin and subtracting the resulting fraction from 1. In the extreme case, if there was no difference in YFP expression between the circuit and the control, the ratio would equate to 1, which would correspond to an efficiency of 0 and the effects that miRNA had on the circuits output expression at that copy number would be conclusively null. On the other hand, if there was a dramatic difference between the YFP expression of the circuit as compared to the control at that bin, the ratio would be effectively 0 which reflects an efficiency of 1 where the presence of miRNA in the circuit's environment had a significant effect on the circuit's output protein expression.

As such, the efficiency of repression, ρ, for each bin was calculated as

$$\rho_{b,d} = 1 - \frac{n_{b,0}}{n_{b,d}} \frac{\sum_i^{n_{b,d}} i_{YFP}}{\sum_i^{n_{b,d}} i_{cYFP}} - \varepsilon = 1 - \frac{\mu YFP_{b,d}}{\mu cYFP_{b,d}} - \varepsilon$$

where indices b and d refer to the bin $(1, 2, \ldots, 80)$ and Dox $(0, 0.001, 0.01, 0.1, 1\, ng/mL)$ concentrations, respectively, n refers to the total number of cells, i_{YFP} and i_{cYFP} are the YFP fluorescence values of the circuit and control of cell i, respectively, and ε is an term which corrects for the absolute difference in expression levels of the circuit and the control, given as the relative difference in expression of the control and circuit of each bin at 0 ng/mL Dox

$$\varepsilon = 1 - \frac{n_{b,0}}{n_{b,0}} \frac{\sum_i^{n_{b,0}} i_{YFP}}{\sum_i^{n_{b,0}} i_{cYFP}}.$$

Initially, miRNA repression efficiency increased as copy numbers increase, reaching a maximum at copy numbers corresponding to ~ 1000 a.u. for each dox case. As plasmid copy number continues to increase, the repression efficiency gradually decreases (Figure 4.5a). The behavior of

the repression efficiency, η, was best fit with a two parameter exponentially decaying, saturating function, modeled as

$$\eta(c) = e^{-\gamma c}\frac{c}{\omega + c} \quad (1)$$

where γ is the decay constant, ω controls the rate of saturation with respect to c, which represents the plasmid copies. Using Pearson's correlation, the coefficient of determination (R-Squared) values obtained from the fits for the 1, 0.1, and 0.01 ng/mL Dox concentrations were all above 0.95 where the 0.001 ng/mL Dox induced group had an R-Squared value above 0.8. Interestingly, the decay constant γ was found to be the same at all Dox concentrations, thus attributing the differences in behavior of the repression efficiency at different Dox concentrations to variations in the parameter ω.

4.3.3 Complete Operational Space of Circuit

In order to obtain a complete representation of the circuit to capture the collective effects of plasmid copies and miRNA concentration on the output, we superimposed the data collected from all Dox concentrations then condensed the resulting data by employing a k-means clustering[43] algorithm for both the circuit and its control (Figure 4.5b, Supp. Figure 4.S9). From this perspective, we were able to visualize the circuits operational expression profile at all levels of miRNA concentration and plasmid copies. By coupling the individual pairwise relationship (Figure 4.2) along with the uncovered repression efficiency (Eq. 1), we constructed a model whose intention is to represent the complete operational output space of the circuit, π, as a function of plasmid copies and miRNA concentration, μ, given as

$$\pi(c,\mu) = \alpha \frac{c}{\eta(c)\,\mu} + \beta = \alpha \frac{c}{e^{-\gamma c}\frac{c}{\omega + c}\,\mu} + \beta$$

which simplifies to,

$$\pi(c,\mu) = \alpha \frac{(c+\omega)}{e^{-\gamma c}\,\mu} + \beta \qquad (2)$$

where α is a proportionality constant, β is a parameter which represents the basal expression of the output, and γ, c, and ω are the same parameters previously discussed (Eq. 1).

The model is composed of three experimental observations; namely, the output is proportional to plasmid copies (Figure 4.3b), inversely proportional to miRNA concentration at a given plasmid copy number (Figure 4.4b), and the effect of miRNA concentration is scaled by an exponentially decaying saturating function (Figure 4.5a). The constructed model was then fit to the clustered data in three-dimensional expression space of the superimposed, clustered data with an R-Squared value of ~0.85 (Figure 4.5b). Sensitivity analysis of equation (2) revealed that the circuit output is increasingly sensitive to changes in miRNA concentration as plasmid copies increase (Supp. Figure 4.S10) but increasingly robust to changes in plasmid copies as miRNA concentration increases (Supp. Figure 4.S11).

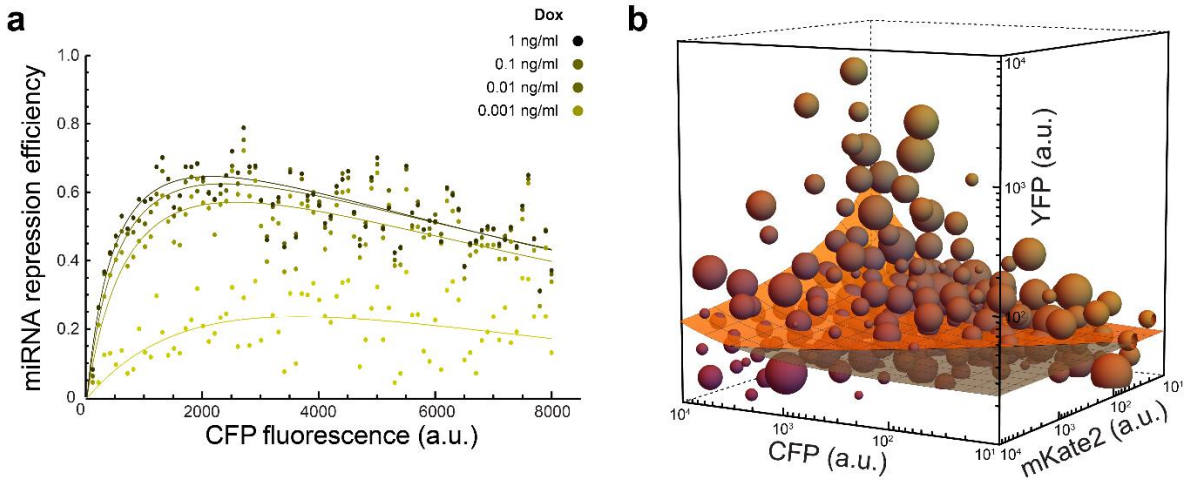


Figure 4.5: **Copy number dependence of miRNA repression efficiency and Operational output space of circuit.** a) Copy number dependence of miRNA repression efficiency. The percent change of YFP fluorescence was calculated for each bin and non-zero Dox concentrations with respect to the 0 Dox control circuit. The effective repression of YFP follows an exponentially decaying, saturating function where YFP reaches a maximum at plasmid copy numbers corresponding to CFP fluorescent values of ~2000 (a.u.) and exponentially decays as plasmid copy number increases. b) Operational output space of circuit. Data from all Dox concentrations were superimposed and clustered. The diameter of a cluster is proportional to the number of cells it contains and the color is a proportional blend of yellow, red, and blue with respect to the cluster's mean YFP, mKate2, and CFP, respectively. The orange lamina represents a proposed relationship of YFP expression as a function of mKate2 and CFP. The gray portion of the lamina along with the bounds of the plot are exclusionary regions where YFP expression was not observed.

4.4 Discussion

Our analytical framework enabled us to uncover a unique, previously unexplored effect which played an important role in shaping the regulated protein's expression behavior as its genetic template increased that was not obvious using the traditional ODE modeling approach (Supp. Figure 4.S12, Supp. Figure 4.S13). Specifically, the observation that miRNA repression drops in efficiency as plasmid copies increases suggests that the rates of the biological processes which generate the output protein's mRNA and the miRNA effector do not scale in tandem as functions of plasmid copies. By mapping the complete circuit's output, we observe that at higher plasmid

copies, the output can out produce the miRNA and eventually overcome repression. One possible explanation is that some of the processing elements uniquely responsible for miRNA biogenesis become saturated due to our circuit producing synthetic miRNA in abundances which far exceed physiological concentrations.

This nuanced effect of miRNA repression at higher genetic copy could be present either in naturally occurring copy number variation genomic regions or consequentially through genomic instability. From an engineering perspective, the consequence of our work provides insight into the subtle behavior of plasmid based, transient gene delivered treatments as the genetic load is non-uniformly distributed among the impinged cells. Taking into account the effects we present could also provide a blueprint from which miRNA-based therapeutics would be optimized or utilized the nonlinear repression efficiency for template quantity specific gene suppression.

4.5 Methods and Procedures

4.5.1 Recombinant DNA Cloning

The restriction enzymes, polymerase, and T4 ligase enzyme used for cloning and ligation were obtained from NEB. QIAGEN Plasmid isolation, gel extraction and PCR purification kits were used. For transformation, either NEB-5alpha competent *E. coli* (catalog #C2987H) or competent DH5alpha cells (originally obtained from Life Technologies) were prepared using the standard $CaCl_2$ method of competent cell preparation. Bacterial culture media and agar (BD Biosciences) were prepared according to manufacturer's instructions. Primers for the experiments were designed using A Plasmid Editor (Ape – version 1.17) and synthesized from Sigma. The primers received were diluted into stocks of 100pmol/μL. The plasmid was sequenced by Genewiz.

4.5.2 Cell Culture and Transfection

HEK293 TET-On Advanced cells (Clontech, Catalog #630931) were maintained at 37C, 100% humidity and 5% CO_2. The cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Catalog #11965-118) supplemented with 10% fetal bovine serum (FBS) (Invitrogen, Catalog #26140-079), 0.1 mM non-essential amino acids (Invitrogen, Catalog #11140-050), 0.045 ug $mL^{-1}$of penicillin and 0.045 mg $mL^{-1}$of streptomycin antibiotics (penicillin–streptomycin liquid, Invitrogen, Catalog #15140-122) and sterilized using a filter (Corning, Catalog #431097). To pass the cells, the adherent culture was trypsinized with 0.25% Trypsin with EDTAX4Na (Invitrogen, Catalog #25200-114) and diluted in fresh medium upon reaching 50–90% confluence.

For transient transfections, ~150,000 cells in 1 mL of complete medium were plated into each well of 12-well culture treated plastic plates (Griener Bio-One, catalog #665180) and grown for 20-24 hours. For JetPrime transfection, 500 ng of the plasmid and 250 ng of untranslatable "junk" DNA was added to 75 µL of JetPrime buffer and 1.75 µL JetPrime (Polyplus Transfection, catalog #114-15). Transfection solutions were mixed and incubated at room temperature for 10 minutes. The transfection mixture was then applied to the cells and mixed with the medium by gentle shaking. When applicable, doxycycline (Clontech, catalog #631311) was added immediately following transfection.

4.5.3 Flow Cytometry

Approximately 48 hours post transfection cells from each well of the 12-well plates were trypsinized with 0.2 mL 0.25% Trypsin-EDTA at 37°C for 3 minutes. Trypsin-EDTA was then neutralized by adding 0.7 mL of complete medium. The cell suspension was centrifuged at 4000

rpm for 1 min and after removal of supernatants, the cell pellets were resuspended in 0.5 mL PBS buffer (Dulbecco's Phosphate Buffered Saline; Mediatech, catalog #21-030-CM). The cells were analyzed on a BD LSRFortessa flow analyzer. TagCFP was measured with a 445-nm laser and a 470/20 band-pass filter, mKate2 with a 561-nm laser, 600 emission filter and 610/20 band-pass filter, and YFP with a 488-nm laser, a 525 emission filter and 545/35 band-pass filter. For all experiments performed 100,000 events were collected. A FSC (forward scatter)/SSC (side scatter) gate was generated using a un-transfected negative sample and applied to all cell samples. We use a compensation matrix on our flow cytometry data to remove cross-talk observed between the three fluorescent proteins (Supp. Figure 4.S14). Data processing was performed in FlowJo 7.6.5. All experiments were performed in triplicates.

4.6 Appendix

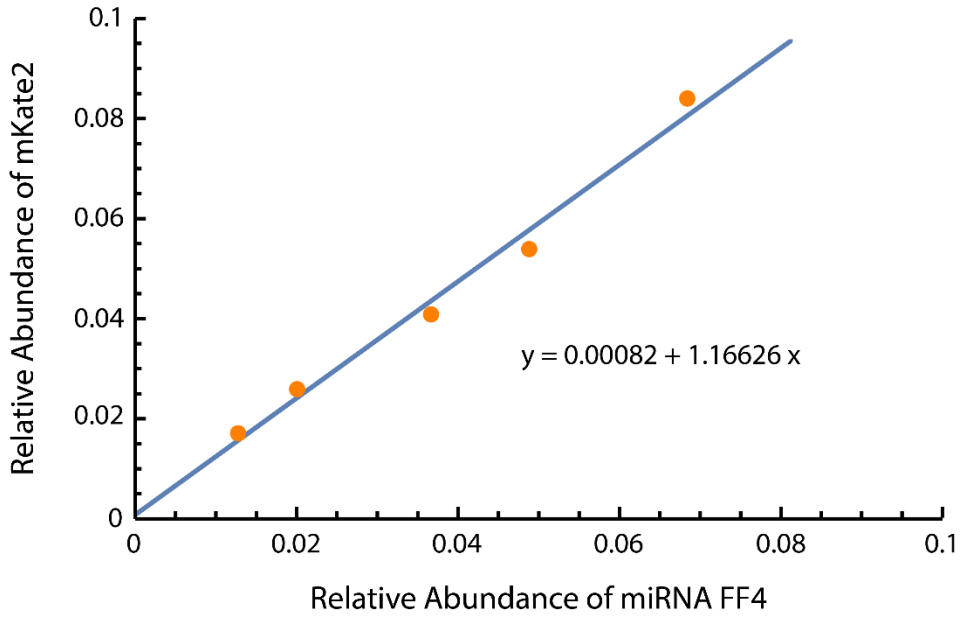


Figure 4.6: **Quantification of the relationship between mKate2 and miRNA-FF4.** Quantitative results of qPCR measurements reveal a linear relationship between mKate2 and miRNA-FF4.

In order to establish the correlation between mKate2 and miRNA-FF4 we performed the following qPCR experiment: 31.25, 62.5, 125, 250 and 500 ng of the miRNA sensor plasmid was co-transfected with 100 ng of PCMV-RTTA into HEK293 cells. 1 µg/mL DOX was added. 48 hours later, the cells were harvested and the total RNA was harvested using the RNeasy Mini Kit (Qiagen, #74104). First strand synthesis was performed using the QuantiTect Reverse Transcription Kit (Qiagen, #205311). Quantitative PCR was performed using the KAPA SYBR FAST Universal qPCR Kit (KAPA Biosystems, #KK4601). GAPDH was used for normalization. The relative mRNA levels for mKate and miRNA were calculated using the $2^{-\Delta Ct}$ method.

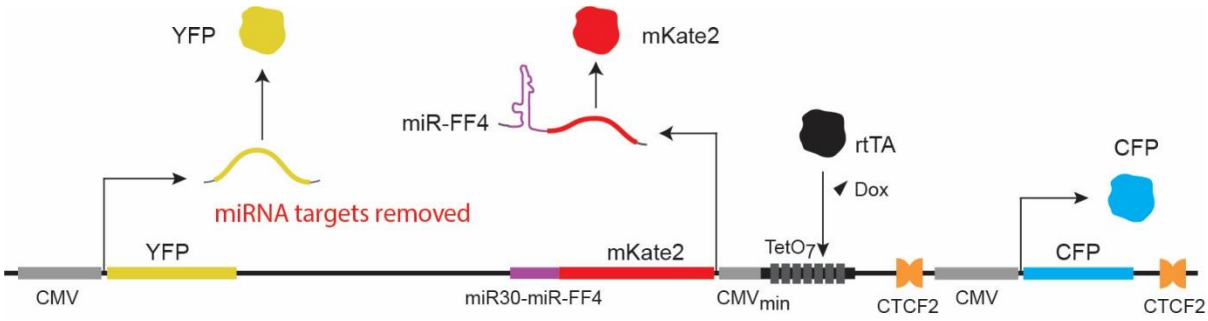


Figure 4.7: **Biological schematic of the control.** The constitutively produced CFP fluorescent protein is used to quantify plasmid copy number. The synthetic miRNA-FF4 is produced in response to Dox and is quantified by mKate2 fluorescence. The FF4 targets found within the 3'UTR of the circuit's output fluorescent protein YFP have been removed.

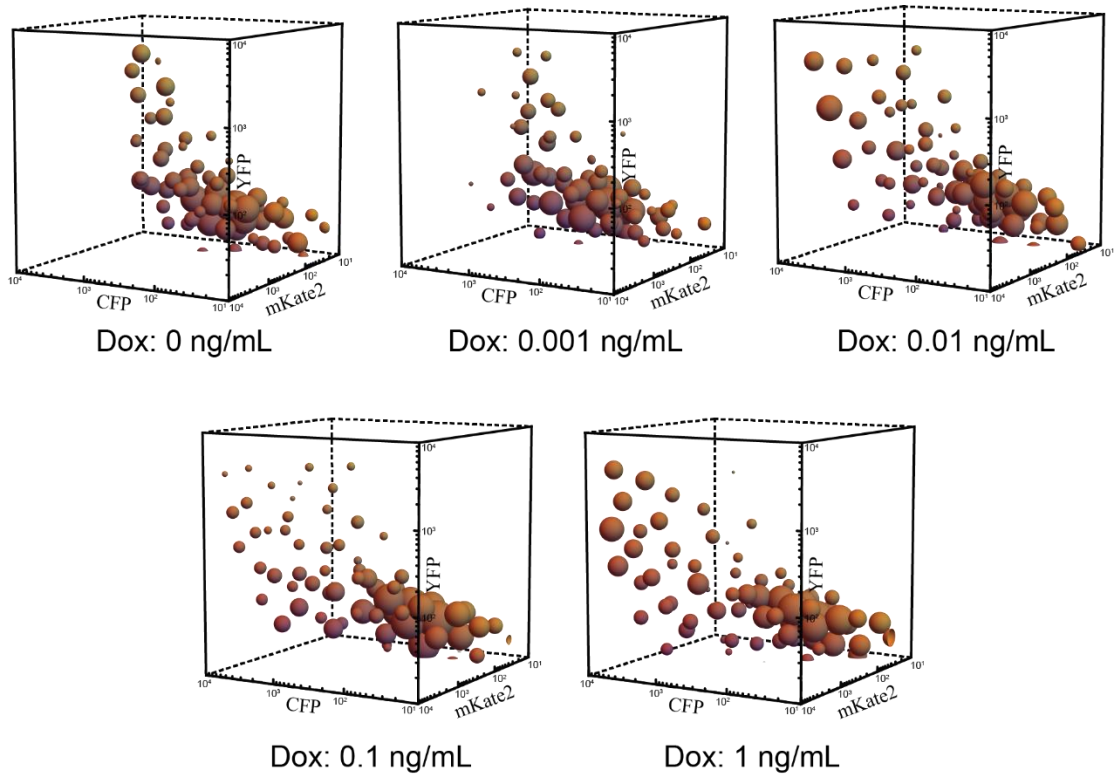


Figure 4.8: **Dox titration cluster plots of control.** Cells containing the control were clustered using k-means clustering. The color of a cluster is a blend of yellow, red, and blue in proportion to the cluster's mean values of YFP, mKate2, and CFP fluorescent arbitrary units, respectively. The diameter of a cluster is proportional to the number of cells it contains.

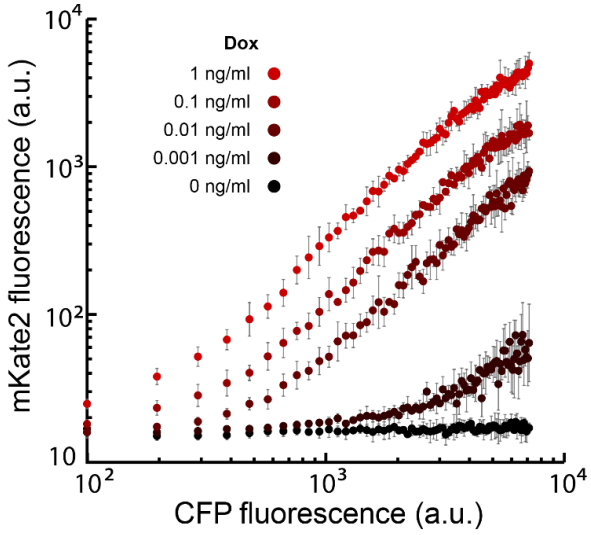


Figure 4.9: **Pairwise expression of mKate2 and CFP of control.** mKate2 fluorescence of the control linearly increases as CFP increases in a Dox dependent manner.

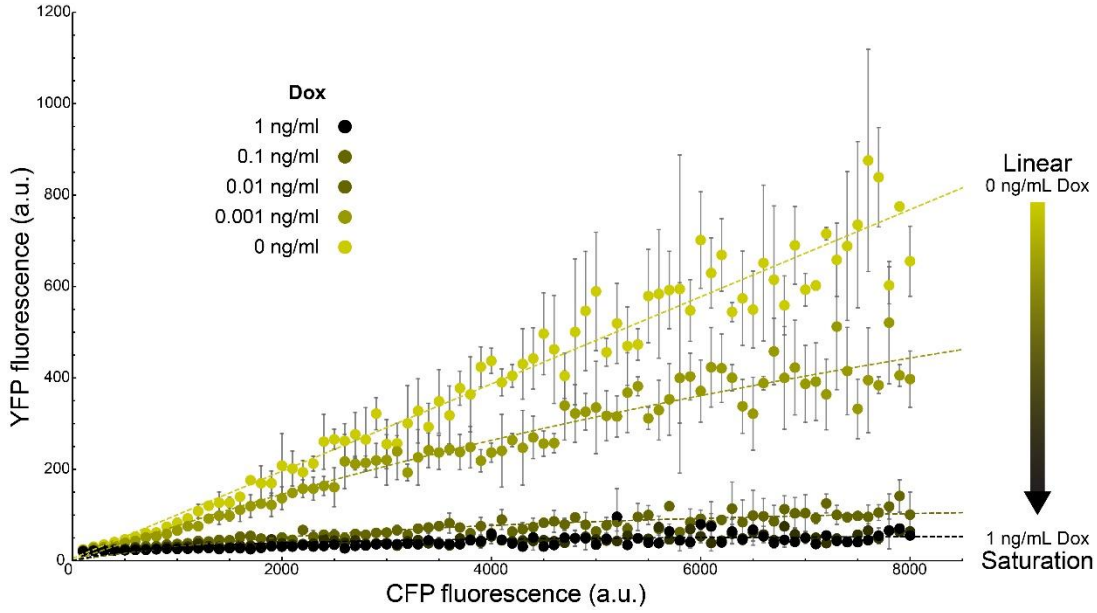


Figure 4.10: **Linear-linear plot perspective of YFP behavior as a function of CFP.** YFP fluorescence transitions from a linear response to saturating response as a function of plasmid copy numbers at increasing Dox concentrations.

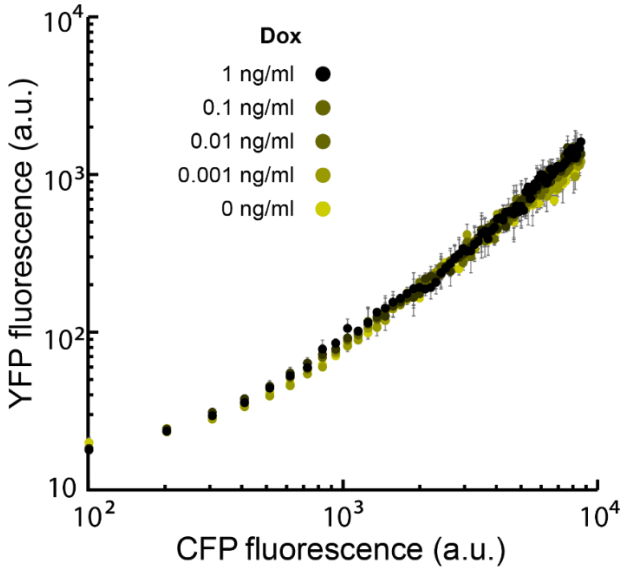


Figure 4.11: **Log-log plot perspective of pairwise expression of control YFP to CFP fluorescence.** linearly responding to increasing CFP fluorescence independent of Dox.

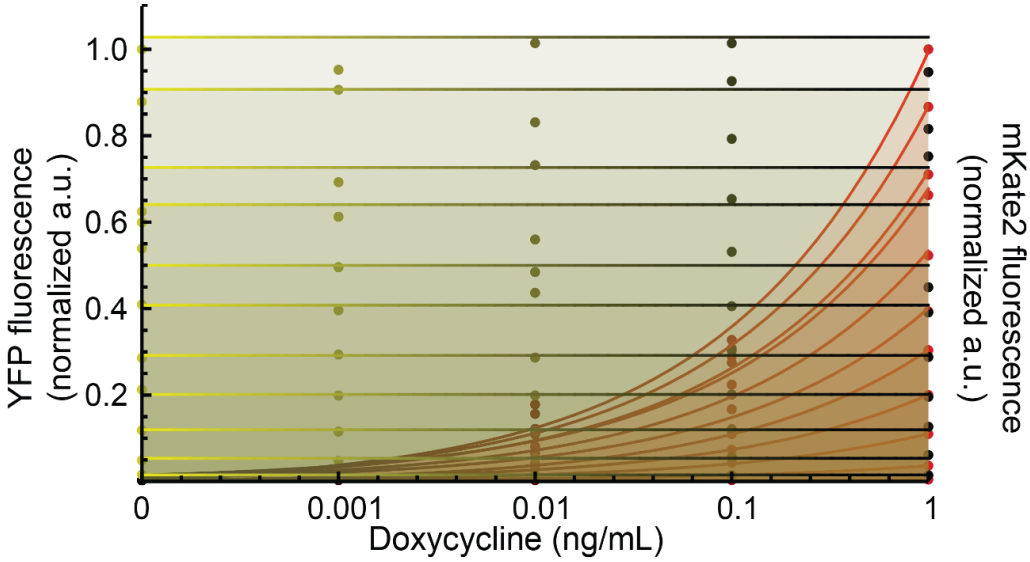


Figure 4.12: **Relative YFP expression of the control.** The relative YFP expression of the control is unaffected by Dox whereas relative expression of mKate2 exponentially increases as a function of Dox.

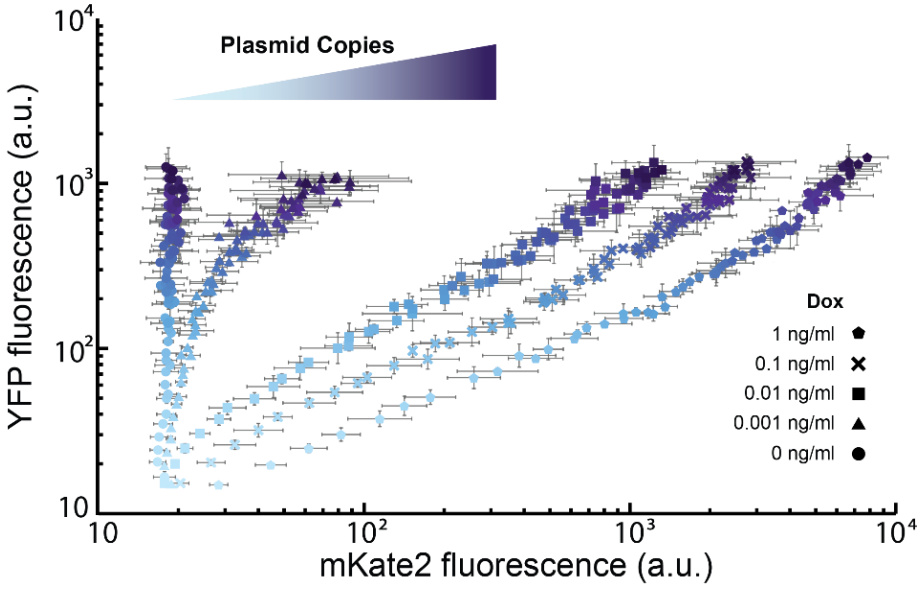


Figure 4.13: **Pairwise expression of control YFP to mKate2.** The control's YFP fluorescence is unaffected by increasing levels of Dox concentration at levels of CFP fluorescence whereas increasing Dox levels increases mKate2 fluorescence.

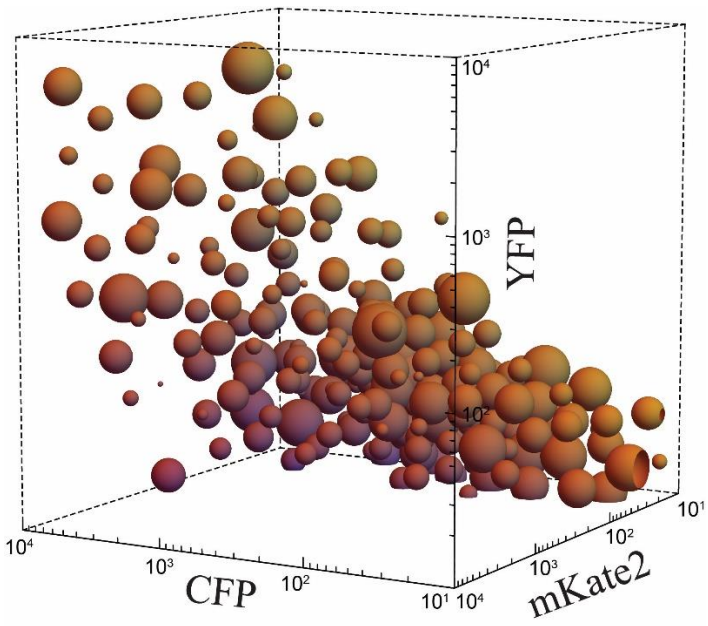


Figure 4.14: **Superposition of clustered data at each Dox concentration of cells containing the control circuit.** The color of a cluster is a blend of yellow, red, and blue in proportion to the cluster's mean values of YFP, mKate2, and CFP fluorescent arbitrary units, respectively. The diameter of a cluster is proportional to the number of cells it contains.

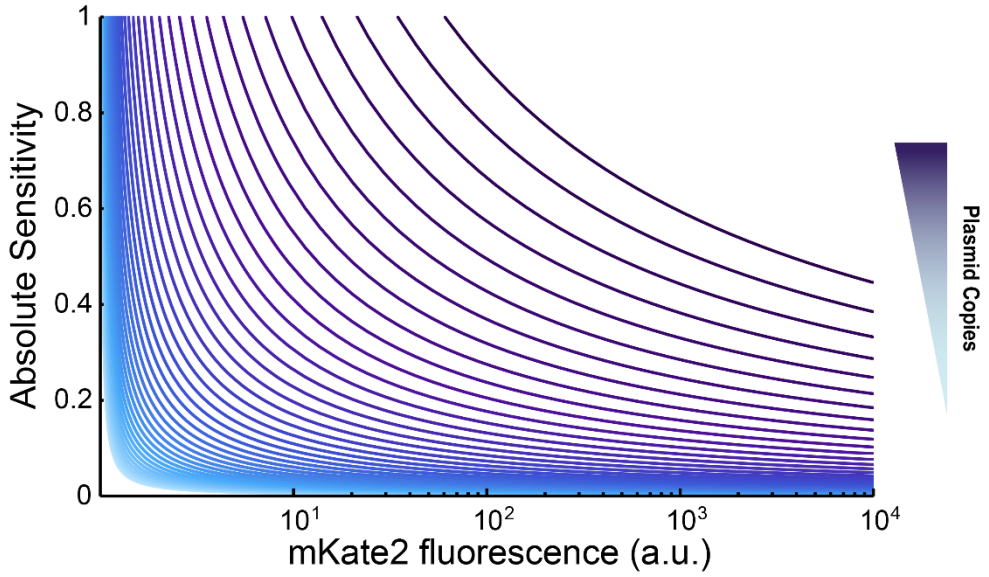


Figure 4.15: **Output sensitivity with respect to plasmid copies.** As plasmid copies increases, the circuit output becomes increasingly sensitivity to mKate2 fluorescence.

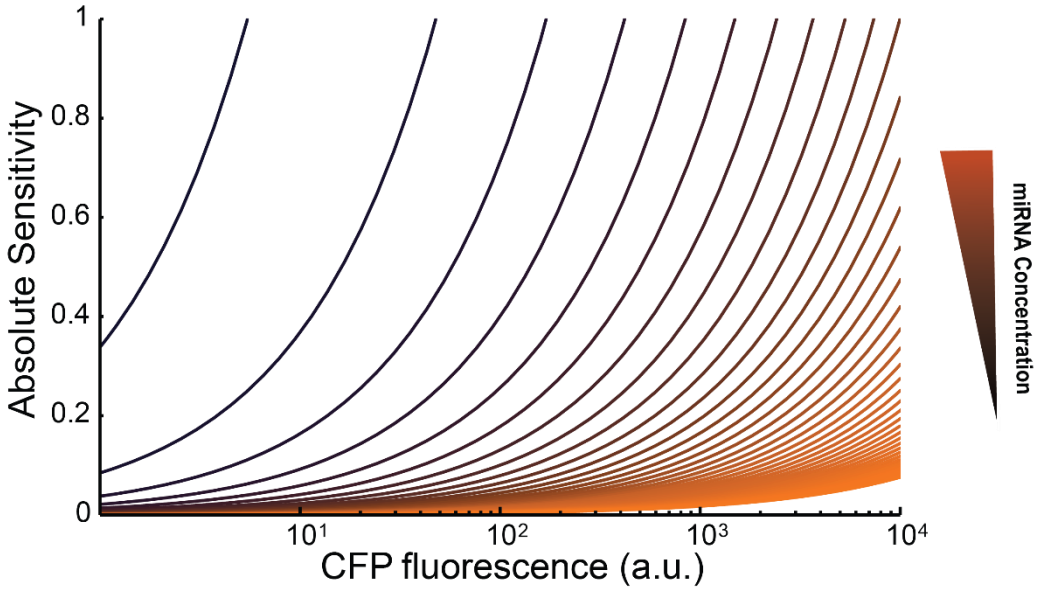


Figure 4.16: **Output sensitivity with respect to miRNA concentration.** As miRNA concentration increases, the circuit output becomes increasingly robust to changes in CFP fluorescence.

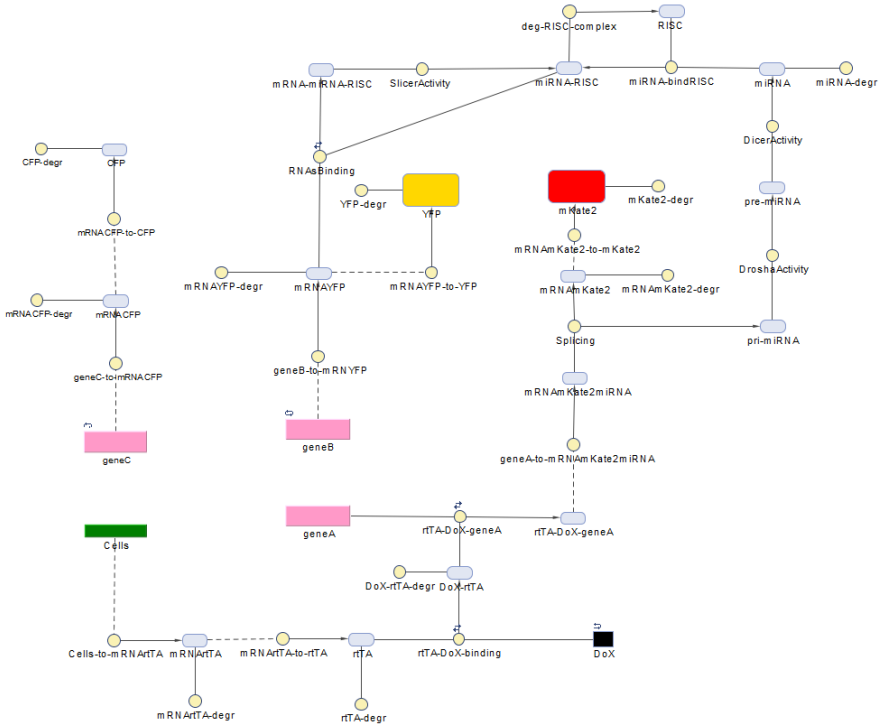


Figure 4.17: **Diagramatic representation of mechanistic SimBiology model.**

The above diagram corresponds to the following reaction array:

YFP -> null
mRNAYFP -> YFP + mRNAYFP
geneB -> mRNAYFP + geneB
[DoX-rtTA] -> null
rtTA -> null
DoX + rtTA <-> [DoX-rtTA]
[DoX-rtTA] + geneA <-> [rtTA-DoX-geneA]
mRNAYFP -> null
Cells -> Cells + mRNArtTA
mRNArtTA -> null
mRNArtTA -> mRNArtTA + rtTA
miRNA + RISC -> [miRNA-RISC]
[miRNA-RISC] -> RISC
miRNA -> null
mRNAmKate2 -> null
mKate2 -> null
[mRNA-miRNA-RISC] -> [miRNA-RISC]
[miRNA-RISC] + mRNAYFP <-> [mRNA-miRNA-RISC]

[pre-miRNA] -> miRNA
[pri-miRNA] -> [pre-miRNA]
mRNAmKate2miRNA -> [pri-miRNA] + mRNAmKate2
mRNAmKate2 -> mRNAmKate2 + mKate2
[rtTA-DoX-geneA] -> mRNAmKate2miRNA + [rtTA-DoX-geneA]
geneC -> mRNACFP + geneC
mRNACFP -> mRNACFP + CFP
CFP -> null
mRNACFP -> null

The initial values of the species geneA, geneB, and geneC were scanned from values of 10 to 100 in steps of 1 simulating different plasmid copy numbers. A further scan of Dox was at each gene quantity was performed where the initial value of Dox was scanned from values of 0 to .9 in 10 steps of .1. Each simulation was run for $3 \cdot 10^5$ seconds upon which the proteins YFP, mKate2 and CFP reach steady-state. The steady-state values of each were then recorded and log transformed to correspond to our experimental data.

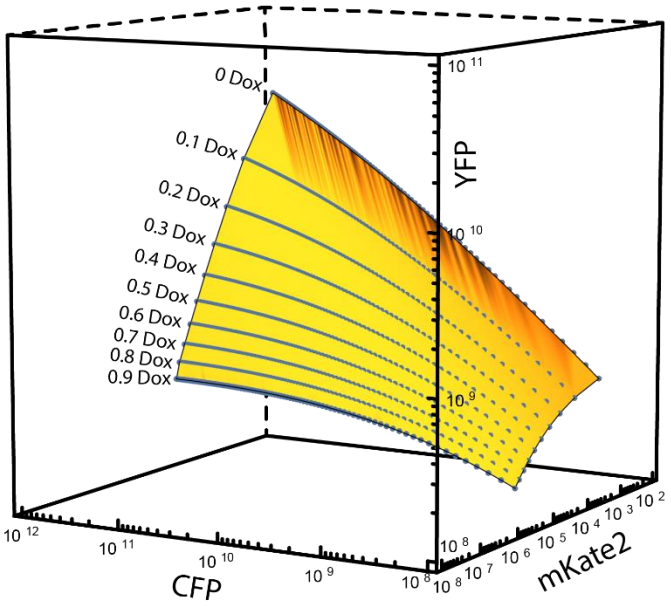


Figure 4.18: **Operational output space of circuit in silico.** Simulated steady state values of each fluorescent protein were extracted from the above ordinary differential equation system at various Doxycycline concentrations.

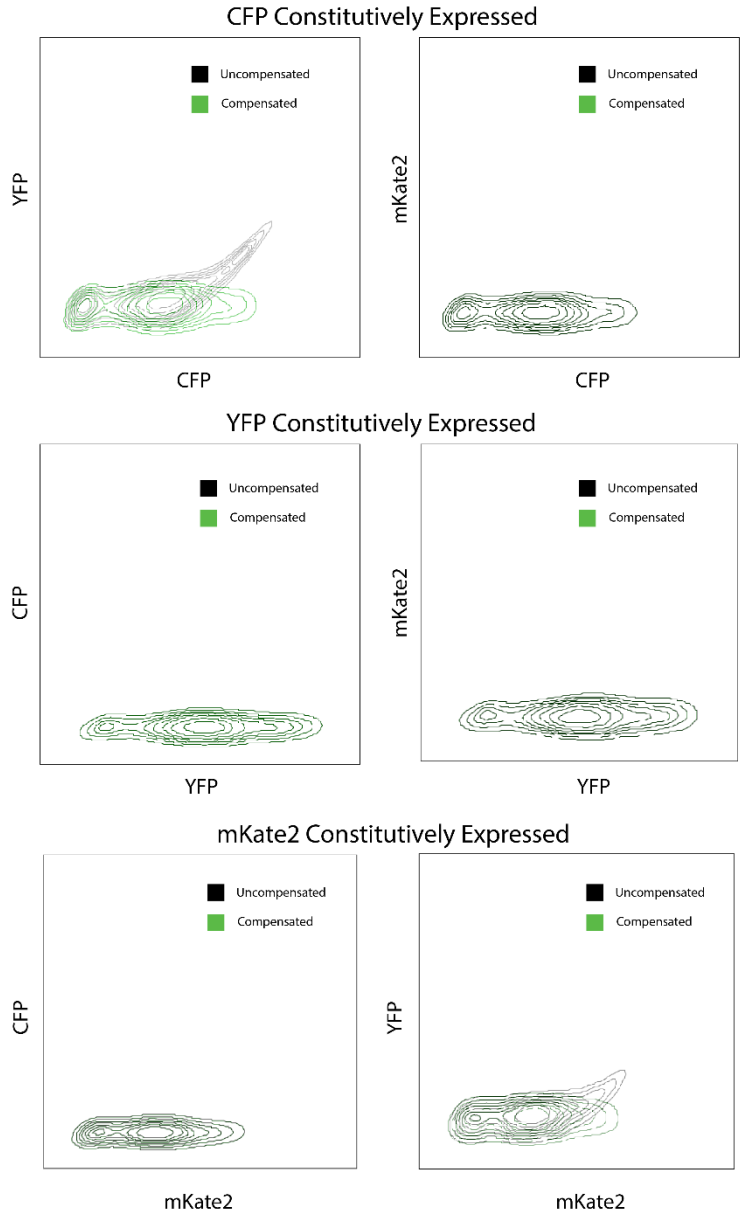


Figure 4.19: **Fluorescent emission cross-talk.** Plasmids which express CFP, mKate2, and YFP were individually transfected in HEK-293 cells.

The resulting data was used to calculate the compensation matrix applied to our data.

	YFP	CFP	mKate2
YFP	100	0	0
CFP	1.8	100	0
mKate2	0.7	0	100

4.7 References

1. Bartel, D. P. & Chen, C. Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* **5,** 396–400 (2004).

2. Bartel, D. P. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136,** 215–233 (2009).

3. Pillai, R. S., Bhattacharyya, S. N. & Filipowicz, W. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends in Cell Biology* **17,** 118–126 (2007).

4. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136,** 642–655 (2009).

5. Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **11,** 228–234 (2009).

6. Mello, C. C. & Conte, D. Revealing the world of RNA interference. *Nature* **431,** 338–342 (2004).

7. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391,** 806–811 (1998).

8. Ameres, S. L. & Zamore, P. D. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14,** 475–488 (2013).

9. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64–71 (2008).

10. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. (2010). doi:10.1038/nature09267

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

12. Ebert, M. S. *et al.* Roles for microRNAs in conferring robustness to biological processes. *Cell* **149,** 515–24 (2012).

13. Osella, M., Bosia, C., Corá, D. & Caselle, M. The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Comput. Biol.* **7,** e1001101 (2011).

14. del Rosario, R. C. H., Damasco, J. R. C. G. & Aguda, B. D. MicroRNA inhibition fine-tunes and provides robustness to the restriction point switch of the cell cycle. *Sci. Rep.* **6,** 32823 (2016).

15. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* **43,** 854–9 (2011).

16. Schmiedel, J. M. *et al.* MicroRNA control of protein expression noise. *Science* **348,** 128–132 (2015).

17. Tsang, J., Zhu, J. & van Oudenaarden, A. MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Mol. Cell* **26,** 753–767 (2007).

18. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol. Syst. Biol.* **7,** 519–519 (2011).

19. Acar, M., Pando, B. F., Arnold, F. H., Elowitz, M. B. & van Oudenaarden, A. A General Mechanism for Network-Dosage Compensation in Gene Circuits. *Science* **329,** 1656–1660 (2010).

20. Strovas, T. J., Rosenberg, A. B., Kuypers, B. E., Muscat, R. A. & Seelig, G. MicroRNA-based single-gene circuits buffer protein synthesis rates against perturbations. *ACS Synth. Biol.* **3,** 324–331 (2014).

21. Li, X. *et al.* A MicroRNA Imparts Robustness against Environmental Fluctuation during Development. *Cell* **137,** 273–282 (2009).

22. Moore, R., Ooi, H. K., Kang, T., Bleris, L. & Ma, L. MiR-192-Mediated Positive Feedback Loop Controls the Robustness of Stress-Induced p53 Oscillations in Breast Cancer Cells. *PLoS Comput. Biol.* **11,** (2015).

23. Lee, J. *et al.* Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc. Natl. Acad. Sci.* **111,** E364–E373 (2014).

24. Yuan, Y. *et al.* Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 3158–63 (2015).

25. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic biology moving into the clinic. *Science* **333,** 1248–1252 (2011).

26. Weber, W. & Fussenegger, M. Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.* **13,** 21–35 (2012).

27. Leisner, M., Bleris, L., Lohmueller, J., Xie, Z. & Benenson, Y. MicroRNA Circuits for Transcriptional Logic. *Methods Mol. Biol. (Clifton, NJ)* **813,** 169 (2012).

28. Benenson, Y. Synthetic biology with RNA: progress report. *Curr. Opin. Chem. Biol.* **16,** 278–284 (2012).

29. Guinn, M. & Bleris, L. Biological 2-input decoder circuit in human cells. *ACS Synth. Biol.* **3,** 627–33 (2014).

30. Kashyap, N., Pham, B., Xie, Z. & Bleris, L. Transcripts for combined synthetic microRNA and gene delivery. *Mol. Biosyst.* **9,** 1919 (2013).

31. Bloom, R. J. *et al.* Synthetic feedback control using an RNAi-based gene-regulatory device. *J. Biol. Eng. 2015 91* **12,** 279–84 (2015).

32. Deans, T. L., Cantor, C. R. & Collins, J. J. A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. *Cell* **130,** 363–372 (2007).

33. Beisel, C. L., Bayer, T. S., Hoff, K. G. & Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol. Syst. Biol.* **4,** 224 (2008).

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

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

36. Ehrhardt, K., Guinn, M. T., Quarton, T., Zhang, M. Q. & Bleris, L. Reconfigurable hybrid interface for molecular marker diagnostics and in-situ reporting. *Biosens. Bioelectron.* **74,** 744–750 (2015).

37. Nevozhay, D., Zal, T. & Balázsi, G. Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat. Commun.* **4,** 1451 (2013).

38. Egbert, R. G. & Klavins, E. Fine-tuning gene networks using simple sequence repeats. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 16817–22 (2012).

39. Beisel, C. L., Bayer, T. S., Hoff, K. G. & Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol. Syst. Biol.* **4,** (2008).

40. Schreiber, J., Arter, M., Lapique, N., Haefliger, B. & Benenson, Y. Model-guided combinatorial optimization of complex synthetic gene networks. *Mol. Syst. Biol.* **12,** (2016).

41. Mohammadi, P. *et al.* Automated Design of Synthetic Cell Classifier Circuits Using a Two-Step Optimization Strategy. *Cell Syst.* **4,** 207–218.e14 (2017).

42. Leisner, M., Bleris, L., Lohmueller, J., Xie, Z. & Benenson, Y. Rationally designed logic integration of regulatory signals in mammalian cells. *Nat. Nanotechnol.* **5,** 666–670 (2010).

43. Bishop, C. M. Neural networks for pattern recognition. *J. Am. Stat. Assoc.* **92,** 482 (1995).

CHAPTER 5

ROBUST FILTERING IN INTRAGENIC MIRNA-MEDIATED HOST REGULATION[†]

Authors – Tyler Quarton[1,2*], Taek Kang[1,2,*], Leonidas Bleris[1,2,3]

Department of Biomedical Engineering

The University of Texas at Dallas

800 W Campbell Road

Richardson, Texas 75080-3021

---

[1]Bioengineering Department, University of Texas at Dallas, Richardson, Texas, USA

[2]Center for Systems Biology, University of Texas at Dallas, Richardson, Texas, USA

[3] Department of Biological Sciences, University of Texas at Dallas, Richardson, Texas, USA

[*]Authors contributed equally.

5.1 Abstract

MicroRNAs (miRNAs) are short non-coding RNA molecules that regulate gene expression post-transcriptionally by binding to target messenger RNAs (mRNAs). The majority of human miRNAs are intragenic, located within exons or introns of protein-coding sequence (host). Intriguingly, a percentage of intragenic miRNAs negatively regulate the host transcript forming an incoherent feedforward motif topology. Here, we study intragenic miRNA-mediated host gene regulation using a synthetic gene circuit stably integrated in human embryonic kidney cells within a genomic safe harbor locus. We engineered an artificial intron by separating the coding sequence of a reporter gene using splice sites and incorporated a synthetic miRNA within an artificial intron. When this intragenic miRNA is directed to inhibit the host gene transcript via engineered targets in its 3'UTR we observe a reduction in fluorescent reporter expression accompanied by output filtering. Specifically, the system operates as a filter with respect to promoter strength, allowing output above and attenuating below certain promoter strength, with the threshold being robust to promoter strength and measurement time.

5.2 Introduction

MicroRNAs (miRNAs) are a class of short, 17-22 nucleotide, non-coding RNA, which regulate gene expression through sequence complementarity[1,2]. Approximately half of miRNA genes can be found in intergenic regions (between genes) (Supplementary Table 4.S1), whereas the intragenic miRNAs (inside genes) are predominantly located inside host gene introns and usually oriented on the same DNA strand of the "host" pre-mRNA which gives rise to both exon and intragenic miRNA[3–5]. Intergenic miRNA genes present their own promoter region[3,6,7], while

same-strand intragenic miRNAs are co-transcribed with their host gene[4,5,8], and then processed to become mature, functional miRNAs. While each intergenic miRNA is considered for its own unique transcriptional unit and expression pattern, it is believed that mechanisms which regulate the production of intragenic miRNAs follow that of its host gene[9]. One hypothesis is that intragenic miRNAs support the host gene by silencing genes antagonistic to itself or by coordinating the expression of genes related to the host gene.

Intragenic miRNAs can regulate their host genes via complementary targets present in their 3'untranslated region (3' UTR)[10,11], forming an incoherent feedforward loop (IFFL)[12–16] regulatory mechanism, as mRNA splicing and miRNA processing occur before miRNA-based repression (Figure 5.1A). More than half of known miRNAs have been predicted to be intragenic, and are implicated in a wide array of biological processes from development to cancer biogenesis[17,18]. In particular, two cases of intragenic miRNA-mediated host gene regulation that have been experimentally studied are: miR-26b and miR-128, both of which are involved in critical processes in cancer proliferation and neural development, respectively[19,20].

We previously characterized transcriptional and posttranscriptional IFFL motifs in transient transfection experiments in cultured human cells and have shown that the IFFL motif leads to adaptation in genetic template fluctuations[15]. Herein, we investigate the properties of intragenic miRNA-mediated host gene regulation using genetic circuits stably integrated in human kidney cells.

5.3 Results

5.3.1 Construction and characterization of synthetic circuit

To engineer a system that implements intragenic miRNA-mediated host gene control we adopted a two-reporter circuit[21,22] that is based on a single inducible bidirectional promoter with TET Response Element (TRE). The left fluorescent reporter coding sequence (DsRed-express) was modified with two consensus splice sites such that the transcript produces two exons surrounding a single intron before splicing (Figure 4.1B). Within the artificial intron we cloned a synthetic miRNA (namely miRNA FF3[15,23]). Notably, the FF3 target does not have crosstalk within the cell[24]. The corresponding FF3 miRNA target was cloned to the 3' UTR of the edited DsRed-express coding sequence to inhibit translation. The resulting transcript serves both as a pre-mRNA and a pri-miRNA; after splicing, the exon is ready to be translated while the intron will undergo endogenous processing to become a mature miRNA. The other reporter gene (AmCyan) is unaltered and, thus, its nascent mRNA transcript is translated normally without interruption. To increase sensitivity of the reporter gene to the miRNA-mediated control, a PEST tag was added to the C-terminal end of both reporters to accelerate protein degradation (Figure 5.1B).

The synthetic circuit was then stably integrated in HEK293 cells that constitutively produce rtTA, which enables the circuit to be activated by doxycycline (Dox). We integrated the cassette into the AAVS1 safe-harbor locus to minimize epigenetic interference and crosstalk with other genes or regulatory elements[25,26]. The integration of the transgene cargo was performed using CRISPR/SpCas9[27,28] with a single guide RNA specific to the AAVS1 locus (Figure 5.1C and Supplementary Table 5.S2).

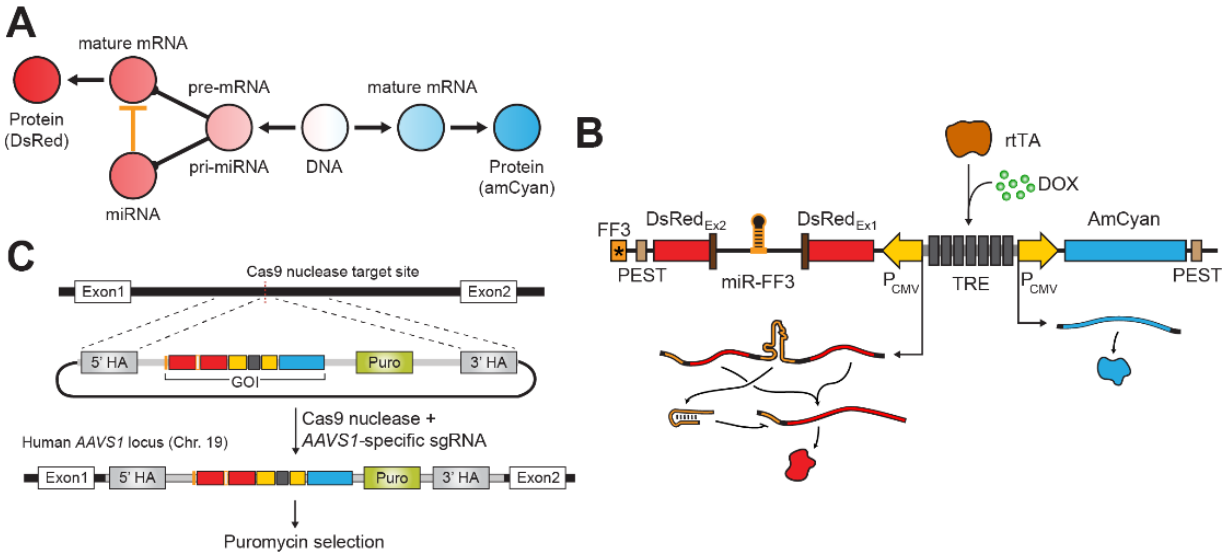


Figure 5.1: **Modeling the host-transcriptional regulation by intragenic miRNA.** (A) Graphical representation of the synthetic gene circuit, (B) schematic of the corresponding plasmid and (C) the CRISPR-Cas9-mediated genomic integration strategy. (A) The synthetic circuit models an intragenic miRNA-mediated transcriptional regulation; the nascent transcript for DsRed fluorescent protein functions as a pre-mRNA as well as a pri-miRNA. (B) The synthetic gene circuit consists of a Tetracycline-controlled promoter (Tet-On) flanked by two different fluorescent reporter genes (AmCyan and DsRed). To introduce an artificial intron in DsRed, a fragment containing synthetic miRNA FF3 flanked by splice sites are inserted in the middle of the coding sequence. The 3' end of the DsRed gene is added with PEST tag to reduce protein half-life and the FF3 miRNA target. (C) The synthetic construct is stably integrated in HEK293 genome in AAVS1 locus via sgRNA/SpCas9-targeted DNA cleavage and subsequent homologous recombination. Puromycin resistance was included in the synthetic gene circuit to aid in selection for isogenic clones.

In addition to the circuit, the donor plasmid contains the Puromycin resistance gene as a selection marker. After integration, we isolated single cells via serial dilution under selection to generate monoclonal populations with the intragenic miRNA cassette (Supplementary Material Methods subsection 4.6.1). Subsequently, we used the circuit monoclonal cell line to engineer a genetically identical control cell line that lacks the functional miRNA 3' UTR targets (thus rendering the intragenic miRNA inactive). The modification was once again carried out using the CRISPR genome editing methodology, where SpCas9 was used to cleave DsRed 3' UTR precisely on the

miRNA target (Supplementary Figure 5.4). Disruption of the miRNA target site relies on NHEJR to incorporate insertion/deletion (indels) at the site of DNA cleavage. The expectation is that AmCyan will be equivalent between the intragenic miRNA and control clones. Meanwhile, the removal of FF3 target on 3' UTR of DsRed transcript should result in higher DsRed expression in the control cell line. We indeed observe recovery of DsRed signal after the removal of intragenic miRNA activity.

5.3.2 Intragenic miRNA-mediated host repression is a robust filter for promoter activity

The observation comparing the intragenic clones to control is that when the promoter activity (mapped by AmCyan fluorescence) is below a certain threshold, the DsRed activity is abolished (Figure 5.2B). Beyond this threshold, DsRed protein is expressed and is able to accumulate, but the maximum expression is lower than that of the control circuit. To model this thresholding relationship between the intragenic miRNA-regulated DsRed and the unregulated AmCyan, we devised a three-parameter piecewise function of the form

$$\mu(x) = \begin{cases} c\,x + d & x \leq x_{thresh} \\ a\,(x - x_{thresh}) + (c\,x_{thresh} + d) & x > x_{thresh} \end{cases}, \quad (1)$$

where μ represents the miRNA repressed DsRed expression, x represents the unregulated AmCyan expression, b represents the basal expression of DsRed, x_{thresh} captures the critical point in which active DsRed expression begins and a represents the proportionality between DsRed and AmCyan above the AmCyan threshold. We fitted this model to flow cytometry data collected from our circuit after induction at 0.1, 0.2, 0.5, and 1.0 μg/mL Dox and at 24, 48, and 72 hours. Using the critical AmCyan values from our model fits, we defined a AmCyan region above background expression and below the threshold in which cells uniquely experience a DsRed suppression akin

to filtering process, in which only cells expressing above the AmCyan threshold also express DsRed in the presence of miRNA regulation (Figure 5.2B, Supplementary Figures 5.S3 and 5.S4). We found that the critical CFP threshold value is independent of both time and Doxycycline induction (Figure 5.2C), presumably because the thresholding effect is an intrinsic property of the intragenic miRNA action. We also provide the fraction of DsRed positive cells within the CFP filtering region in both the intragenic miRNA and control cases (Figure 5.2D), again highlighting the filtering properties and their impact particularly at 48 and 72 hour measurements.

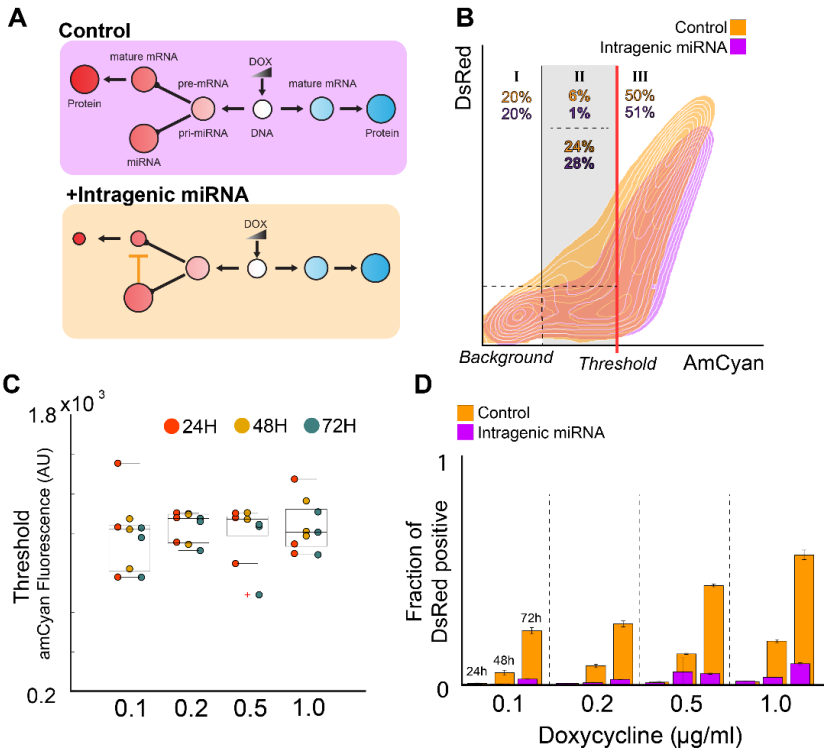


Figure 5.2: **Filtering of DsRed fluorescent activity by intragenic miRNA.** (A) Schematic representation of the control gene circuit with no transcriptional repression to either fluorescent reporters (top) and the circuit with active intragenic miRNA-controlled DsRed (bottom). (B) The threshold for DsRed expression in our genetic circuit is defined as the promoter activity level (as measured by AmCyan fluorescence) at which the DsRed expression is above the background. The numbers within each subsection of the plot indicates the fraction of the parent population. (C) The fluorescence threshold for the synthetic circuit calculated for each combination of promoter activation strength and time. (D) The fraction of cells expressing DsRed in cells below DsRed threshold (column II in (B)) for each combination of promoter activation strength and time.

5.3.3 Stochastic population simulations

Although thresholding has been associated with miRNA-based gene expression[29], we uniquely observe that the thresholding behavior is robust to promoter induction level and quasi-steady state time, potentially due to the intragenic nature of our miRNA. For a mechanistic understanding of the thresholding behavior, we developed a model of our intragenic microRNA IFFL circuit and its control based on a system of ordinary differential equations with key processing steps from gene transcription to protein translation that incorporates intragenic microRNA splicing and target suppression (Figure 5.2, Table 5.1)**.** As our genetic constructs exhibit variable co-expression in quasi-steady state populations, we introduce noise into our system following the error in variable (EIV) methodology where each reporter experiences both local and global noise (Supplementary section, Introduction of Stochasticity to Models).

We numerically simulate stochastic populations of 5,000 cells at variable induction levels and quasi-steady state measurement times. We find that a thresholding behavior emerges in our model containing the miRNA IFFL topology and is absent in the control model in which the miRNA action is removed (Figure 5.3A). Using a generalized version of the threshold model as above (1), where the expression of DsRed within the filtering region has been expanded to a linear function that allows for incomplete suppression, we characterize our stochastically simulated population's thresholding behavior (Figure 5.3B). We find that there is an induction domain in which two co-expression behaviors about a critical AmCyan threshold value becomes observable in our simulations ($0.03 < \beta < 0.07\ s^{-1}$), prominently characterized by the markedly different slopes of the proportional relationships between DsRed and AmCyan expression in both behavior regions.

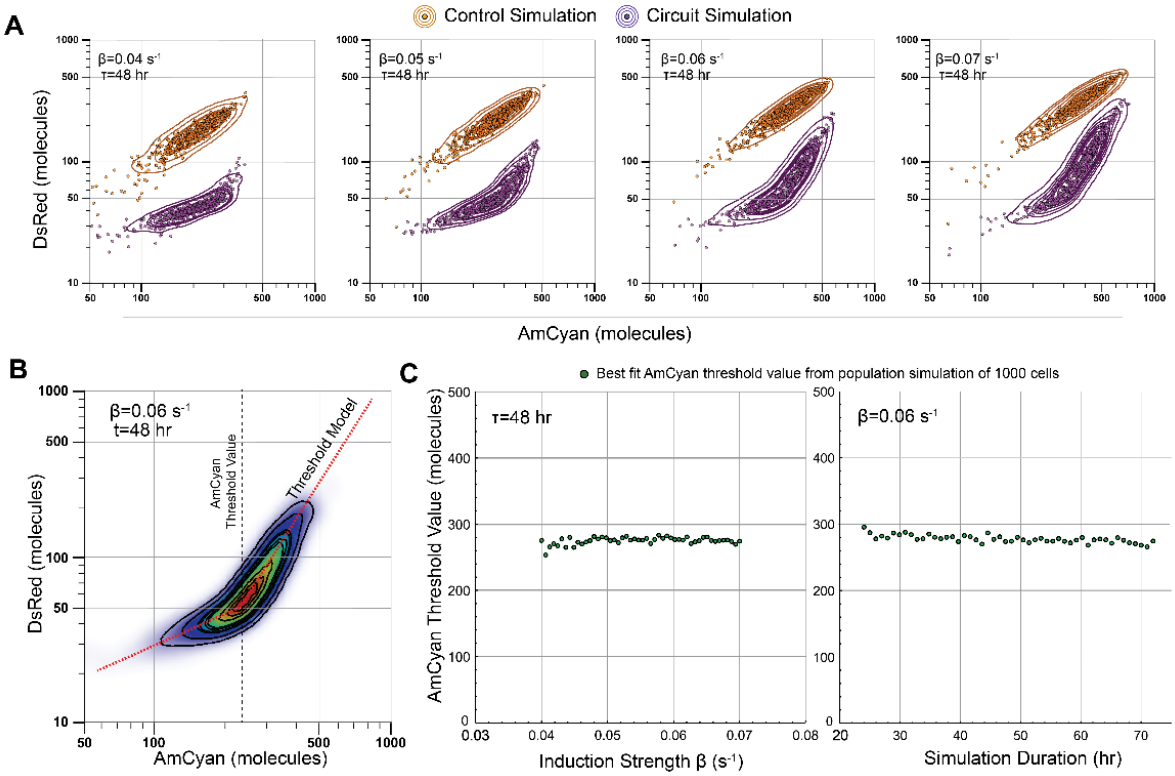


Figure 5.3: **Stochastic population simulations of circuit and control.** (A) Resulting DsRed and AmCyan expression of circuit (purple) and control (orange) using stochastic population simulations of 5000 cells at four different induction strengths, β, collected at a quasi-steady simulation time of 48 hours. (B) Example threshold behavior characterization of a simulated population. (C) Critical AmCyan threshold values extracted from stochastic population simulations by scanning different induction strengths (left) and simulation times (right).

Induction levels below this domain solely exhibit DsRed suppression as expected by miRNA action, where induction levels above this domain show a disproportional accumulation of DsRed with respect to AmCyan. We then stochastically simulate cell populations containing our circuit architecture at different induction levels and quasi-steady state measurement times. We find that within the observable threshold induction domain, the critical AmCyan threshold value is robust to both induction strength and quasi-steady state measurement times (Figure 5.3C). We found that

our models underrepresented miRNA action in the subthreshold region, hinting at the presence of biological mechanisms that bolster filtering strictness not accounted for in our models.

5.4 Methods

5.4.1 Cell Culture

Tet-On cells and the derived stable clones were maintained at 37ºC, 5% CO_2 and 100% humidity. The cells were grown in Dulbecco's modified Eagle's medium (DMEM, Life Technologies, #11965-118) supplemented with 10% Fetal Bovine Serum (FBS, Atlanta Biologicals, #S11550), 0.1 mM MEM non-essential amino acids (Life Technologies, #11140-050), 0.045 units/mL of Penicillin and 0.045 units/mL of Streptomycin (Penicillin-Streptomycin liquid, Life Technologies, #15140-122). To pass the cells, the adherent culture was washed with DPBS (Dulbecco's Phosphate-Buffered Saline, 1X with calcium and magnesium, Corning, #21-030-CM) and then trypsinized with Trypsin-EDTA (Trypsin-EDTA (0.25%), phenol red, Life Technologies, #25200-114) and finally diluted in fresh medium upon reaching 50–90% confluence. The derived stable cells were maintained in 2ug/mL puromycin.

5.4.2 Generation of stable cell line

To generate the monoclonal stable cell lines, ~10 million of the Tet-On cells were seeded onto a 10 cm petri dish. 16 hours later, the cells were transiently transfected with 5 μg of the donor plasmid, 5 μg of the U6-gRNA-PCMV-Cas9-t2A-mKate plasmid using the JetPRIME reagent (Polyplus Transfection). 48 hours later, puromycin (Life Technologies, catalog number: A1113803) was added at the final concentration of 2 μg/mL. The selection lasted ~2 weeks, after

which the surviving clones were pooled to generate the polyclonal stable cells, from which monoclonal stable cells were isolated using the serial dilution method.

5.4.3 Single Cell Serial Dilution

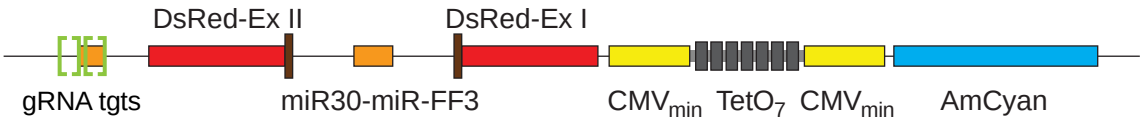
To generate clonal population, serial dilution method used with a 96-well plate. First, a confluent cell culture was detached using trypsin and resuspended using conditioned media. The conditioned media prepared by harvesting at approximately 75% confluency, then clarified by centrifugation at 1000 x g for 10 minutes then passed through a sterile filter. Resuspension was diluted to a final concentration of 2 x 10 4 cells/mL, then added inoculated to a single well in a 96-well plate.

5.4.4 Fluorescent Microscopy

Microscopy was performed 24–72 hours post induction with doxycycline. The live cells were grown on 12-well plates (Greiner Bio-One) in the complete medium. Cells were imaged using an Olympus IX81 microscope in a Precision Control environmental chamber. The images were captured using a Hamamatsu ORCA-03 Cooled monochrome digital camera. The filter sets (Chroma) are as follows: ET436/20x (excitation) and ET480/40 m (emission) for AmCyan, ET545/30x (excitation) and ET620/60 m (emission) for DsRed. Data collection and processing was performed in the software package Slidebook 5.0. All images within a given experimental set were collected with the same exposure times and underwent identical processing.

5.4.5 Flow Cytometry

Flow cytometry was performed 24-72 hours post induction with doxycycline. Each well of the 12-well plates were washed with 1.0 mL of DPBS and then trypsinized with 0.3 mL 0.25% Trypsin-EDTA at 37°C for 5 minutes. Trypsin-EDTA was then neutralized by adding 0.7 mL of complete medium. The cell suspension was centrifuged at 1,000 rpm for 5 minutes and after removal of supernatants, the cell pellets were resuspended in 0.5 mL DPBS. The cells were analyzed on a BD LSRFortessa flow analyzer. AmCyan was measured with a 445-nm laser, a 505 long-pass filter, and 525/50 band-pass filter, and DsRed with a 561-nm laser, and 582/15 band-pass filter. For all experiments ~200,000 events were collected. All wells within a given experimental set had the same number of events collected with the same voltages and underwent identical processing. The flow cytometry data was processed using FlowJo. A FSC (forward scatter)/SSC (side scatter) gate was generated applied to all cell samples. The gated cells were then further gated into AmCyan positive cells by gating above the AmCyan levels of uninduced cells. For some analysis a DsRed positive gate was then applied, again by gating above the DsRed levels of uninduced cells. All data underwent identical processing for each analysis.



Reference	Sequence
sgRNA 1 (5')	**AGCTTGTCGACGATATCTCC**
sgRNA 2 (3')	**CAAACGATATGGGCTGAATAC**

Figure 5.4: **Guide RNA sequences for removing FF3 miRNA targets.** In order to remove the FF3 miRNA target in the genomically-integrated synthetic circuit, the nuclease was targeted directly to the target site with 2 different sgRNAs. The sgRNA targets are shown in green brackets.

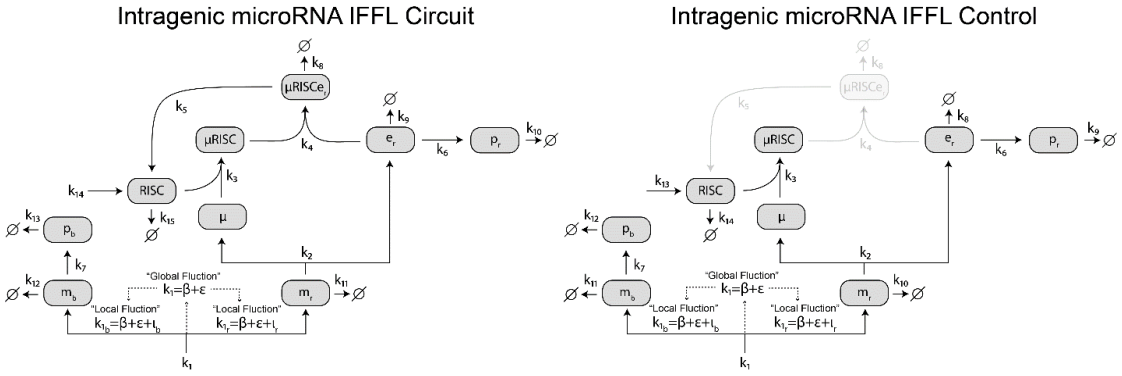


Figure 5.5: **Diagrammatic representation of the intragenic microRNA IFFL circuit (left) and the control (right) for a single cell.** To introduce stochasticity, the co-transcription rate k_1 is expanded to include noise terms and the control is obtained by nullifying the interaction parameter between the target mRNA and the RISC complex.

Table 5.1: **Parameters and species of stochastic models.** Each parameter value of the ODE models are renference to previous experimental and theoretical publications.

Specie or Parameter	Initial Value (Specie) Value (Parameter)	Units	Description	Reference
m_r	0	molecule(s)	DsRed premRNA	N/A
m_b	0	molecule(s)	AmCyan mRNA	N/A
e_r	0	molecule(s)	DsRed mRNA exon	N/A
p_r	0	molecule(s)	DsRed protein	N/A
p_b	0	molecule(s)	AmCyan protein	N/A
μ	0	molecule(s)	FF3 microRNA	N/A
$RISC$	1000	molecule(s)	RISC complex	N/A
$\mu RISC$	0	molecule(s)	RISC complex bound to FF3 microRNA	N/A
$\mu RISCe_r$	0	molecule(s)	RISC complex + FF3 microRNA + DsRed mRNA exon complex	N/A
k_2	0.002	s^{-1}	DsRed premRNA splicing and processing rate	[15]
k_3	0.00001	$(mol \cdot s)^{-1}$	RISC and FF3 microRNA binding rate	[15]

k_4	0.00000184	$(mol \cdot s)^{-1}$	RISC-FF3 microRNA complex and DsRed exon binding rate	2
k_5	0.0000216	s^{-1}	RISC complex recycling rate	[15]
k_6	0.00033333	s^{-1}	DsRed mRNA exon translation rate	[49,50]
k_7	0.00033333	s^{-1}	AmCyan mRNA translation rate	[49,50]
k_8	0.007	$(mol \cdot s)^{-1}$	RISC-microRNA-exon degradation rate	[15]
k_9	0.00034	s^{-1}	DsRed exon degradation rate	[50]
k_{10}	0.000096667	s^{-1}	DsRed protein degradation rate	[50]
k_{11}	0.00034	s^{-1}	DsRed premRNA degradation rate	[50]
k_{12}	0.00034	s^{-1}	AmCyan mRNA degradation rate	[50]
k_{13}	0.000096667	s^{-1}	AmCyan protein degradation rate	[50]
k_{14}	0.0467	s^{-1}	RISC complex basal production rate	[50]
k_{15}	0.000096667	s^{-1}	RISC complex degradation rate	[50]

5.4.6 Introduction of Stochasticity to ODE Models

To capture the experimentally observed co-expression variability in our steady state data, we incorporate error into our ODE system following an errors in variables (EIV) methodology where we rewrite the promoter activity as

$$k_1 = \beta + \varepsilon$$

where β is a tunable parameter that captures the systems doxycycline induction input and ε is an error term sampled from a normal distribution used to represent noise. Following the EIV approach, we assume the mean and standard deviation of the error term $E(\varepsilon) = 0,\ std(\varepsilon) = \sigma_\varepsilon$. Under the biological assumption that extrinsic noise is independent of promoter induction and (reference), we choose $\sigma = \frac{n}{m}\beta$ with $m > n$, for any $n, m \in \mathbf{Z}$ to be some fraction of β so that

the noise proportionally scales at different induction levels such that the noise of the promoter as measured by the coefficient of variation remains constant. As constructed, for each simulated cell, the error parameter ε will proportionally affect the expressions of both AmCyan and DsRed and represents an extrinsic, or global, source of noise. To outfit our model to include an intrinsic, or local, noise source, we add a second error term that will be independently affect both fluorescent reporters, where k_1 further expands to

$$k_{1_{b,r}} = \beta + \varepsilon + \iota_{b,r}$$

where $\iota_{b,r}$ is the local error source that independently affects the expression of both fluorescent proteins, indicated with the subscripts (*b*)lue and (*r*)ed. Procedurally, this independence is attained by generating two values from the distribution of $\iota_{b,r}$ during the initialization of the simulation. Again, following the EIV approach, we set $E(\varepsilon) = 0,\ std(\iota_{b,r}) = \sigma_{b,r}$. With this treatment of noise, our system of ODE's becomes

$$\frac{dm_r}{dt} = \beta + \varepsilon + \iota_r - k_2 m_r - k_{11} m_r$$

$$\frac{d\mu}{dt} = k_2 m_r - k_3\ RISC \cdot \mu$$

$$\frac{dRISC}{dt} = k_{14} + k_5 \mu RISC e_r - k_3 RISC \cdot \mu - k_{15} RISC$$

$$\frac{d\mu RISC}{dt} = k_3 RISC \cdot \mu - k_4\ \mu RISC \cdot e_r$$

$$\frac{d\mu RISC e_r}{dt} = k_4\ \mu RISC \cdot e_r - k_5 \mu RISC e_r - k_8 \mu RISC e_r$$

$$\frac{de_r}{dt} = k_2 m_r - k_4\ \mu RISC \cdot e_r - k_6 p_r - k_9 e_r$$

$$\frac{dp_r}{dt} = k_6 e_r - k_{10} p_r$$

$$\frac{dm_b}{dt} = \beta + \varepsilon + \iota_b - k_7 m_b - k_{12} m_b$$

$$\frac{dp_b}{dt} = k_7 m_b - k_{13} p_r$$

For our simulations, we variably chose $\sigma_\varepsilon = .3\beta$, 30% variation that scales with our induction strengths and $\sigma_{b,r} = 0.003$, a 5% variation with respect to a baseline induction of $\beta = 0.06\ s^{-1}$.

5.5 Discussion

While microRNAs have been studied thoroughly over the years and have been extensively used in synthetic biology applications, the effect that a miRNA's genetic origin has on its target mRNA's expression has yet to be distinguished. Here, we experimentally studied intragenic miRNA-mediated host transcript regulation, where the miRNA and its target are processed from the same gene transcript. We demonstrate that the system behaves as a filter with a robust cutoff threshold.

Advancements in the field of synthetic biology have produced circuits with increasingly sophisticated functions and a continually expanding library of components[46–48]. As the field of synthetic biology moves toward clinical applications, precision and robustness become paramount. Herein, inspired by a naturally-occurring architecture, we describe a compact and modular genetic circuit and demonstrate generally applicable property.

5.6 References

1. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281–297 (2004).

2. Kim, Y.-K. & Kim, V. N. Processing of intronic microRNAs. *EMBO J.* **26**, 775–783 (2007).

3. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).

4. He, C. *et al.* Young intragenic miRNAs are less coexpressed with host genes than old ones: implications of miRNA–host gene coevolution. *Nucleic Acids Res.* (2012).

5. Hinske, L., Galante, P., Kuo, W. & Ohno-Machado, L. A potential role for intragenic miRNAs on their hosts' interactome. *BMC Genomics* **11**, 533 (2010).

6. Ying, S.-Y. & Lin, S.-L. Current perspectives in intronic micro RNAs (miRNAs). *J. Biomed. Sci.* **13**, 5–15 (2006).

7. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Res.* **14**, 1902 (2004).

8. Ma, N. *et al.* Coexpression of an intronic microRNA and its host gene reveals a potential role for miR-483-5p as an IGF2 partner. *Mol. Cell. Endocrinol.* **333**, 96–101 (2011).

9. Baskerville, S. & Bartel, D. P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* **11**, 241–7 (2005).

10. Flynt, A. S. & Lai, E. C. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat. Rev. Genet.* **9**, 831–842 (2008).

11. Li, S.-C., Tang, P. & Lin, W.-C. Intronic MicroRNA: Discovery and Biological Implications. *DNA Cell Biol.* **26**, 195–207 (2007).

12. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman & Hall/Crc Mathematical and Computational Biology Series)*. ({Chapman & Hall/CRC}, 2006).

13. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* **100**, 11980–11985 (2003).

14. Bosia, C., Osella, M., Baroudi, M., Cor?, D. & Caselle, M. Gene autoregulation via intronic microRNAs and its functions. *BMC Syst. Biol.* **6**, 131 (2012).

15. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol. Syst. Biol.* (2011). doi:10.1038/msb.2011.49

16. Osella, M., Bosia, C., Corá, D. & Caselle, M. The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Comput. Biol.* **7**, e1001101 (2011).

17. Hinske, L. C. *et al.* miRIAD--integrating microRNA inter- and intragenic data. *Database* **2014**, bau099–bau099 (2014).

18. Lutter, D., Marr, C., Krumsiek, J., Lang, E. W. & Theis, F. J. Intronic microRNAs support their host genes by mediating synergistic and antagonistic regulatory effects. *BMC Genomics* **11**, 224 (2010).

19. Dill, H., Linder, B., Fehr, A. & Fischer, U. Intronic miR-26b controls neuronal differentiation by repressing its host transcript, ctdsp2. *Genes Dev.* **26**, 25–30 (2012).

20. Li, M. *et al.* miR-128 and its target genes in tumorigenesis and metastasis. *Exp. Cell Res.* **319**, 3059–3064 (2013).

21. Shimoga, V., White, J. T., Li, Y., Sontag, E. & Bleris, L. Synthetic mammalian transgene negative autoregulation. *Mol. Syst. Biol.* **9**, 670 (2013).

22. Kashyap, N., Pham, B., Xie, Z. & Bleris, L. Transcripts for combined synthetic microRNA and gene delivery. *Mol. Biosyst.* **9**, 1919 (2013).

23. Leisner, M., Bleris, L., Lohmueller, J., Xie, Z. & Benenson, Y. Rationally designed logic integration of regulatory signals in mammalian cells. *Nat. Nanotechnol.* **5**, 666–70 (2010).

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

25. Satoh, W., Hirai, Y., Tamayose, K. & Shimada, T. Site-Specific Integration of an Adeno-Associated Virus Vector Plasmid Mediated by Regulated Expression of Rep Based on Cre-loxP Recombination. **74**, 10631–10638 (2000).

26. Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012).

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

28. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).

29. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* **43**, 854–9 (2011).

30. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).

31. Blake, W. J., KÆrn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).

32. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (80-.).* (2002). doi:10.1126/science.1070919

33. Pandya-Jones, A. *et al.* Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* **19**, 811–27 (2013).

34. Swinburne, I. A. & Silver, P. A. Intron delays and transcriptional timing during development. *Dev. Cell* **14**, 324–30 (2008).

35. Takashima, Y., Ohtsuka, T., Gonzalez, A., Miyachi, H. & Kageyama, R. Intronic delay is essential for oscillatory expression in the segmentation clock. *Proc. Natl. Acad. Sci.* **108**, 3300–3305 (2011).

36. Komorowski, M., Miękisz, J. & Kierzek, A. M. Translational Repression Contributes Greater Noise to Gene Expression than Transcriptional Repression. *Biophys. J.* **96**, 372–384 (2009).

37. Herranz, H. & Cohen, S. M. MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev.* **24**, 1339–1344 (2010).

38. Paulsson, J. Summing up the noise in gene networks. *Nature* **427**, 415–418 (2004).

39. Munsky, B., Neuert, G. & van Oudenaarden, A. Using Gene Expression Noise to Understand Gene Regulation. *Science (80-.).* **336**, 183–187 (2012).

40. Waks, Z., Klein, A. M. & Silver, P. A. Cell-to-cell variability of alternative RNA splicing. *Mol. Syst. Biol.* **7**, 506 (2011).

41. Marinov, G. K. *et al.* From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510 (2014).

42. Hornung, G. & Barkai, N. Noise Propagation and Signaling Sensitivity in Biological Networks: A Role for Positive Feedback. *PLoS Comput. Biol.* **4**, e8 (2008).

43. Lee, T. I. *et al.* Transcriptional Regulatory Networks in Saccharomyces cerevisiae. *Science (80-.).* **298**, 799–804 (2002).

44. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–7 (2002).

45. Quarton, T. *et al.* Mapping the operational landscape of microRNAs in synthetic gene circuits. *npj Syst. Biol. Appl.* **4**, 6 (2018).

46. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic biology moving into the clinic. *Science* **333**, 1248–52 (2011).

47. Kitada, T., DiAndreth, B., Teague, B. & Weiss, R. Programming gene and engineered-cell therapies with synthetic biology. *Science* **359**, eaad1067 (2018).

48. Lillacci, G., Benenson, Y. & Khammash, M. Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic Acids Res.* **46**, 9855–9863 (2018).

49. Tigges, M., Dénervaud, N., Greber, D., Stelling, J. & Fussenegger, M. A synthetic low-frequency mammalian oscillator. *Nucleic Acids Res.* **38**, 2702–2711 (2010).

50. Tigges, M., Marquez-Lago, T. T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).

BIOGRAPHICAL SKETCH

Tyler Quarton was born in Albuquerque, New Mexico and raised in Tyler, Texas by his two loving parents Anna Quarton and Martin Olson. After completing his schoolwork at All Saints Episcopal School in 2006, Tyler entered Trinity University in San Antonio, Texas where he began his studies in both physics and mathematics. In 2010, Tyler transferred to The University of Texas at Dallas where he met his wife Patricia Martinez. He went on to receive a Bachelor of Science in physics from The University of Texas at Dallas in December 2013. In January 2014, he entered the biomedical engineering program as a PhD candidate at The University of Texas at Dallas. After completing his degree in December 2019, Tyler will pursue his entrepreneurial goals of revolutionizing the scientific publishing industry.

CURRICULUM VITAE

Tyler George Quarton

EDUCATION

Doctor of Philosophy, Biomedical Engineering **Dec 2019**
The University of Texas at Dallas, Richardson, TX.

Bachelor of Science, Physics **Dec 2013**
The University of Texas at Dallas, Richardson, TX.

POSITIONS

Graduate Research Assistant with Prof. Leonidas Bleris **Dec 2013–present**
Department of Biomedical Engineering
The University of Texas at Dallas, Richardson, TX.

Teacher Assistant **Dec 2013–present**
Department of Biomedical Engineering
The University of Texas at Dallas, Richardson, TX.

PUBLICATIONS

Ehrhardt, K., Guinn, M. T., **Quarton T.**, Zhang, M. Q. & Bleris, L. Reconfigurable hybrid interface for molecular marker diagnostics and in-situ reporting. *Biosens. Bioelectron.* (2015). doi:10.1016/j.bios.2015.07.035

Quarton, T., Kang, T., Sontag, E. D. & Bleris, L. Exploring the impact of resource limitations on gene network reconstruction. in *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (2016). doi:10.1109/CDC.2016.7798773

Quarton, T., Ehrhardt K., Lee, J., Kannan, S., Li, Y., Ma, L., Bleris, L. Mapping the operational landscape of microRNAs in synthetic gene circuits. *npj Syst. Biol. Appl.* (2018). doi:10.1038/s41540-017-0043-y

WORKS IN PROGRESS

Kang, T.*, **Quarton, T.*,** Ehrhardt, K., Nowak, C., Singh, A., Li, Y., Bleris, L. Robust filtering and noise suppression in intragenic miRNA-mediated host regulation *Nature Communications* (under revision, *authors contributed equally)

Quarton, T.*, Nowak, C.*, Nguyen, K., Bleris, L. Measuring Population Cell Cycle Period Variance through Desynchronization Rates (*authors contributed equally)

CONFERENCES

Quarton, T., Sontag, E. D. & Bleris, L. The Impact of Resource Constraints on the Reverse Engineering of Biological Pathways American Institute of Chemical Engineering Annual Meeting 2016 (AIChE 2016)

Quarton, T., Kang, T., Sontag, E. D. & Bleris, L. Exploring the impact of resource limitations on gene network reconstruction. in *2016 IEEE 55th Conference on Decision and Control, CDC 2016* (2016)

Quarton, T., Ehrhardt K., Lee, J., Kannan, S., Li, Y., Ma, L., Bleris, L. Mapping the operational landscape of microRNAs in synthetic gene circuits. Mapping the Operational Landscape of microRNAs in Synthetic Gene Circuits Fifth International Mammalian Synthetic Biology Workshop 2018 (MSBW 5.0)

Quarton, T., Ehrhardt K., Lee, J., Kannan, S., Li, Y., Ma, L., Bleris, L. Mapping the operational landscape of microRNAs in synthetic gene circuits. Mapping the Operational Landscape of microRNAs in Synthetic Gene Circuits American Institute of Chemical Engineering Annual Meeting 2018 (AIChE 2018)