

Modeling and design of synthetic biochemical circuits for biological phenotypes

Thesis by
Pranav Bhamidipati

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended 8 December 2023

© 2024

Pranav Bhamidipati
ORCID: 0000-0002-6199-6505

All rights reserved except where otherwise noted

ACKNOWLEDGEMENTS

I recently spoke with an undergraduate who is thinking of pursuing a PhD, and to my horror I watched as the very advice I received at the beginning of my journey spewed helplessly from my mouth, only mildly solicited. I know it must have contained certain phrases – “Before all else, love your science. . . Resilience is important. . . Have hobbies outside of research. . .” – which I once dismissed as probably irrelevant to me. After all, I had survived the pre-med curriculum, the seemingly never-ending applications and interviews to MD-PhD programs, and a rigorous pre-clinical curriculum in medical school. How much harder could a PhD be?

I was wrong to be so self-assured. In reality, these few years have included many failures and challenges. This thesis does not explicitly include the many failed projects (today’s count is six) that together occupied half of my years in graduate school, if not the majority of it. I say “explicitly” because, even though the desired outcome was never realized, each failure taught me a skill that was used to produce some part of this document. I also have come to believe in the importance of a gentle, congenial relationship with failure. In the words of the Irish poet David Whyte, “a vocation always includes the specific, heartrending way we will fail at our attempt to live fully.” Looking back, I feel the vital pulse beating beneath the words of his short essay titled “Ambition” (“Consolations” 2019).

Ambition is natural to the first steps of youth, who must experience its essential falsity to know the larger reality that stands behind it. . . [But a life’s work always includes] the passing on of a sense of sheer privilege, of having found a road, a way to follow, and the having been allowed to walk it, often with others, with all its difficulties and minor triumphs; the underlying primary gift of having been both a witness to and a full participant in the conversation.

I am struck by gratitude for having been given this opportunity and for the people who have made this journey rich and joyful. First, I am grateful for my advisor Matt, who has always and foremost encouraged my growth and independence as a scientist. I arrived at Caltech with no engineering education, no experience in synthetic biology, and a very basic foundation in mathematics, hoping to fashion

myself into a computational biologist focused on synthetic cell engineering. I am tremendously thankful for his patience in training me towards my goals and his generosity with his time and resources. His willingness to pursue new research directions and uncanny sense for good ideas are inspiring to me, and I can feel how his example has, slowly but surely, shaped me into a scientist in the same mold. His patience and encouraging words during some of the rough patches of failed experiments, missed deadlines, and unexpected life events also taught me that the process of science is always more important than its outcome.

I also could not have succeeded in my work without a network of mentors and collaborators. I would like to thank my committee members. Justin Bois's courses played an important role in shaping my research interests, and they gave me the quantitative skills to call myself an engineer despite coming from a non-quantitative background. And there were a few pivotal moments during my PhD when Michael Elowitz's wise words guided my decisions, including the fortuitous decision to rotate with Matt in the first place. Al Barr brought to my committee an enthusiasm for my research direction that was nourishing and motivating. Apart from my committee, I am also proud to count the post-doctorates Dominik Schildknecht and Josquin Courte as both mentors and friends, a rare and precious combination. I also greatly appreciate my collaborators in the Morsut Lab at USC for their consistent collaborative spirit, open minds, and valuable feedback. Although it was brief, my time working with Leonardo Morsut helped spark my passion for the important role of computation in synthetic circuit design. My work was also improved from interesting conversations with many members of the Elowitz Lab, including Gal Manella, Joseph Parres-Gold, and Martin Tran.

I am also grateful for the rich and supportive community of friends and labmates I had at Caltech. Thanks to my labmates, coming to work at Caltech was a pleasure. I would like to thank Jerry, who always let me interrupt his day to talk about a random project idea or workshop a new slide deck. Those conversations were so spontaneous, and they always left me feeling more creative. I will miss talking about science (and gossiping about scientists) over lunch with Arjuna, Shichen, and Meera. Thanks to Arjuna and Shichen, I thoroughly enjoyed going to conferences and will never hear the phrase "conference room" the same again. My musical jam sessions with Aman, epic barbecue dinners with James, and many board game nights with my labmates will not soon be forgotten. I also greatly enjoyed the few months I got to spend sitting side-by-side with Dominik. To Laura, I am grateful to

have had a friend with which to share all the ups and downs, from venting over long walks to celebrating with chicken sandwiches. I am also thankful for my MD-PhD mentors and colleagues for their advice and camaraderie, and I am glad that I will get to share my clinical years with them as well.

This journey would not have been possible without the support of my parents Dr. B.V.R. Murthy and Dr. Sujatha Bhamidipati. Their constant encouragement has been so precious to me, and their daily loving presence reminds me that the most important things in life are love, health, and happiness. Amma and Nanna, you were my original teachers, and I feel blessed that I am still learning so many lessons from you. I would also like to thank my sister Harini Bhamidipati for being my cheerleader all these years and for providing me with a never-ending stream of dog memes. Speaking of which, I would like to acknowledge the important contributions of Chikoo Bhamidipati, who is a very good boy.

Finally, I would like to acknowledge those who have been a part of my personal and spiritual journey over these last few years. The incredible warmth and understanding provided by Michelle Kim have been a guiding light throughout the last two years, and I count her as a role model. I hope to spend much more time in the stillness and beauty of the Caltech campus and the rolling greens of Monterey Hills, which have grounded me in uncertain times. I also deeply appreciate the authors Thich Nhat Hanh, Alan Watts, Sam Harris, Pete Walker, Forrest Hanson, Rick Hanson, and David Goggins. Their written and recorded works have been an invaluable resource for my growth over these past few years.

ABSTRACT

Biological behaviors arise from the dynamical interactions of biochemical networks. For example, the various immune responses to damage are manifestations of signaling networks between immune cell types. A central goal in systems and synthetic biology is to elucidate the design principles of these networks, or circuits, both in the sense of dissecting how function arises from structure in the natural context and in the sense of understanding the guidelines for optimal engineering of synthetic biological systems. The study of design principles in both senses is aided by mathematical modeling and simulation, which provide a self-consistent framework for evaluating the theoretical implications of biological hypotheses as well as a testbed for the development of novel circuits for desired biological phenotypes. This thesis pertains to two related challenges in this field, namely the scaling of computational design to larger circuits and the engineering of global phenotypes that emerge nonlinearly from local interactions.

The first section of this thesis presents a novel design platform for biological circuits, called CircuiTree, that uses a game-playing paradigm to overcome the combinatorial complexity of *de novo* circuit design. This platform treats circuit design as a game of circuit assembly and traverses the tree of possible assemblies using Monte Carlo tree search (MCTS). Borrowed from artificial intelligence (AI) agents that have mastered complex games, MCTS is a reinforcement learning (RL)-based search algorithm that efficiently searches for the most effective design strategies and naturally discovers design principles in the form of network motifs, which appear as clusters of solutions in the search tree. Finally, when tasked with designing fault-tolerant oscillators with five components, CircuiTree finds a novel design strategy, which we call motif multiplexing, in which multiple sub-oscillators are interleaved so as to render the circuit highly resistant to deletions and knockdowns. This design principle, which may be responsible for the multiple oscillatory loops observed in eukaryotic circadian clocks, opens the possibility of engineering synthetic circuits at a larger scale and suggests that larger biological circuits contain yet-unknown design features that are not simply extensions of smaller circuits.

The second section describes a novel mechanosensitive property of the SynNotch synthetic chimeric receptor and uses a multicellular modeling framework to show how it can be used to control spatiotemporal patterning *in vitro*. Modified from the endogenous juxtacrine receptor Notch, SynNotch binds to an arbitrary extracellular

ligand and, in response, releases an arbitrary transcription factor, thus acting as a user-defined signal transducer. We show that, in mouse fibroblasts, a simple sender-receiver SynNotch circuit ceases to transduce a membrane-bound GFP signal at high cell densities in 2D culture. Because of this feature, a lawn of cells expressing a signal-relay circuit, which we call the transceiver circuit, can undergo spatially limited activation, where the signal propagates in a wave outward from a GFP-expressing sender cell until, due to cell division, the cell density crosses a threshold value and the signaling system shuts down. Using a multicellular lattice-based model combined with experiments, we demonstrate that perturbations of growth parameters can be used to control the size of activated spots. Finally, we achieve spatiotemporal patterns of activation by seeding the growth dish nonuniformly, creating a wave of activation at the millimeter scale that recapitulates the kinematic wave patterning phenomenon observed during vertebrate somitogenesis.

Together, this body of work represents an advance in the use of computational methods and mathematical modeling to guide the design and control of complex biological phenotypes. Advances in these methods promise to catalyze the development of more advanced cell-based therapies and engineered tissues.

PUBLISHED CONTENT AND CONTRIBUTIONS

Bhamidipati, P. S. and M. Thomson (n.d.). *Designing biochemical circuits with tree search*.

Manuscript in preparation. P.S.B. developed the theoretical and computational framework, performed computational experiments, analyzed the data, created figures, and performed the calculations. P.S.B. and M.T. conceived the study, wrote the manuscript, and contributed to the overall direction and planning.

Santorelli, M. et al. (Oct. 5, 2022). *Control of spatio-temporal patterning via cell density in a multicellular synthetic gene circuit*. DOI: 10.1101/2022.10.04.510900.

(Pre-print uploaded to *bioRxiv*.) P.S.B performed the mathematical modeling, data analysis, figure making, data management, and writing of the supplemental manuscript, and participated in the design of experiments and writing of the main manuscript. Authors M.S. and P.S.B. contributed equally.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	vi
Published Content and Contributions	viii
Table of Contents	viii
Chapter I: Introduction	1
1.1 Biological behaviors emerge from the activity of biochemical circuits	1
1.2 Combinatorial complexity is a key barrier in the forward-design of biological phenotypes	2
Chapter II: Designing biochemical circuits with tree search	10
Chapter III: Control of spatio-temporal patterning via cell density in a multi- cellular synthetic gene circuit	51
Chapter IV: Concluding remarks	118
4.1 Overcoming barriers to engineering complex behavior with <i>in silico</i> circuit design and synthetic mechano-chemical coupling	118
4.2 Future perspectives	120

Chapter 1

INTRODUCTION

1.1 Biological behaviors emerge from the activity of biochemical circuits

Since its coining in 1974, the "genotype-phenotype map" has become an authoritative model for how an organism's genetic material determines its behavior (Lewontin, 1974; Alberch, 1991). This model posits that differences in an organism's inherited genome, or genotype, manifest as differences in its behavior, including how it responds to stimuli as well as its static traits such as shape and appearance (Vienne, 2022). In the intervening decades, we have found that genetic and epigenetic factors produce biological phenotypes by altering the expression levels, spatial distributions, and interrelations between biomolecules. In many cases where the underlying network is not well understood, it is useful to study the genotype-to-phenotype (or gene-and-environment-to-phenotype) mapping alone, while treating the intervening dynamics as a black-box. However, a focus on primarily genetic causation risks excluding aspects of the underlying system that play a determining role in generating complex behavior but are not genetically encoded such as transport phenomena, cell mechanics, boundary conditions, and tissue-level material properties (Thompson, 1917; Müller and Newman, 2003; Pfeifer, Shyer, and Rodrigues, 2024). The study of these biochemical networks, or circuits, also provides a deeper insight into pathological phenotypes and their possible therapeutic interventions (Alon, 2024). To understand, engineer, and control biology, one must interface with biological systems at the level of these circuits.

1.2 Combinatorial complexity is a key barrier in the forward-design of biological phenotypes

Biochemical circuits can be represented mathematically as a time-evolving system of interacting variables, or dynamical system. Consider the system of k variables $\mathbf{x} = [x_0, \dots, x_k]$, where the time-derivatives of the variables $\dot{\mathbf{x}} = F(\mathbf{x}; s, \theta(s))$ are determined by a system of (generally nonlinear) functions F that depends both on the circuit's topology s and its parameters θ . By "topology" we mean the qualitative nature of the interactions between the biomolecules x_i , such as "Protein A binds to Protein B" or "Transcription Factor X inhibits Gene Y". The set of topologies \mathcal{S} is the set of possible combinations of these interactions. Because each of these interactions may involve reactions with different rate parameters, the size of the parameter set depends on s ($\theta \in \mathbb{R}^{n(s)}$).

The distinction between "parameters" and "topologies" is, in a sense, arbitrary. For example, the effect of a transcription factor on target gene expression can be modeled discretely as either positive or negative regulation (with a magnitude determined by the associated parameters) or continuously with a parameter that denotes either activation or inhibition, depending on its sign. However, the latter choice can obscure the importance of the qualitative difference between repression and activation, which serves a central role in determining the phenotype of a circuit (Diego et al., 2018; Dassow et al., 2000). Additionally, certain choices are inherently discrete. For example, a regulatory region may have 3 or 4 binding sites, but not $3\frac{1}{2}$.

To forward-design a biochemical circuit to exhibit a target phenotype, therefore, is to optimize its topology and reaction rate parameters with respect to some real-valued phenotype score $f(\mathbf{x}(t))$. Although both topology and parameters are important for phenotype, I will primarily discuss topology optimization in this chapter due to its unique challenges as a problem of combinatorial optimization.

A key initial step in engineering a synthetic biological circuit is to decide its topology. The first implementations of synthetic gene circuits in bacteria recapitulated circuits that had been identified in nature and extensively studied in theory such as oscillatory repressor systems (Elowitz and Leibler, 2000) and bistable switches (Gardner, Cantor, and Collins, 2000). Alternatively, one could design a topology by hand that theoretically achieves the desired goal, using human ingenuity and mathematical analysis of the governing equations. However, in the absence of natural examples and failing this creative endeavor, the problem of design can be approached using computation.

The canonical method that has emerged to discover circuit topologies *de novo* is a method we call enumeration. The steps of enumeration are:

1. Define a set of possible circuit topologies \mathcal{S} and list all the possible configurations. For instance, the space of all regulatory circuits consisting of four transcriptional repressors has $2^{4^2} = 65,536$ topologies before symmetries (4^2 pairwise interactions that can each be present or absent). Also implement a phenotype function, where each evaluation $f_i : s \mapsto q \in \{0, 1\}$ returns a Boolean result q denoting whether or not the phenotype has been achieved. Typically, each evaluation runs a simulation of the dynamics of $\mathbf{x}(t)$ with a different random parameter set θ_i and classifies the dynamics. The parameter distribution $P(\theta)$ should be predetermined based on physical limits and domain knowledge and have a finite expectation. One could also define a deterministic phenotype that can be inferred directly from the circuit topology, without requiring simulation.
2. Evaluate the phenotype of all topologies $s \in \mathcal{S}$. Typically, this involves running an exhaustive number of simulations per topology ($N = 10^4 - 10^7$) and computing the robustness (in the sense of robustness to parameter variation) of each topology $Q(s) = \frac{1}{N} \sum_i f_i(s)$.
3. Identify the best individual topologies *post hoc* as the most robust topologies.
4. Apply a threshold to Q (determined before step 3) to classify topologies as successful or unsuccessful, and run a statistical test to determine which of the successful topologies are so-called *network motifs* (Alon, 2007), common patterns that recur among circuits of the same phenotype. Because these systems are often poorly understood in theory (thus necessitating a computational approach), this step is particularly helpful.

Network motifs found by enumeration provide a window into the design principles of a phenotype. For instance, Schaerli and colleagues performed an enumeration study to design three-node circuits (2,897 possible topologies) that respond non-monotonically to a morphogen input and therefore can form stripes in the presence of a morphogen gradient (Schaerli et al., 2014) thus recapitulating the "French flag" mechanism of domain patterning hypothesized by Wolpert in 1969 (Wolpert, 1969). The 109 successful stripe-forming circuits are visualized in Figure 1 of Schaerli et al., 2014 as a *complexity atlas* (Cotterell and Sharpe, 2010), where each topology

is a node in a graph (thus making a graph-of-graphs), nodes are organized in layers based on the number of interactions present in the topology, and nodes in adjacent layers are connected by edges if they differ by the addition/removal of a single interaction. The solutions cluster into four groups of topologically similar circuits, each based on a simple incoherent feedforward loop (IFFL) motif. Thus, they find that incoherent regulation is a common and necessary principle for stripe-formation in this search space and use this discovery to engineer stripe-forming bacterial colonies. Enumeration has been used similarly to discover motifs for phenotypes of biochemical adaptation (Ma et al., 2009), single-cell polarization (Chau et al., 2012), morphogen patterning (Cotterell and Sharpe, 2010), temporal filtering (Gerardin, Reddy, and Lim, 2019), and multi-functional patterning (Jiménez et al., 2017).

Despite these advances, the drawback of enumeration lies in its poor scaling. For example, consider a network of m transcriptional repressors, where each repressor can repress (or not repress) each other one. The number of possible topologies for this network, modulo symmetries, is 2^{m^2} (the number of possible colorings for each pairwise edge to the power of the number of edges) which scales super-exponentially in m . If regulation is bifunctional (activation is possible in addition to inhibition), there are now three possible colorings for each edge, and the space scales as 3^{m^2} . For dimerization networks in which each monomer component can dimerize with another to regulate a third component, the scaling is 3^{m^3} .

Enumeration becomes computationally intractable on the order of $|\mathcal{S}| \gtrsim 10^6$ topologies, a bound exceeded by $m \geq 5$ repressors or $m \geq 4$ bifunctional regulators. Specifically, consider the quite generous case where we have access to 1,000 CPUs for a month's duration (roughly 4×10^7 CPU-mins). If each simulation and phenotype evaluation takes 100ms, we could run roughly 2.4×10^{10} simulations, or 10^4 simulations each for 2.4×10^6 . Even with a 10x improvement in both computational capacity and evaluation speed, this method does not scale to $m = 6$.

This bound presents a severe limitation from a systems biology perspective because it limits our ability to study natural systems at their natural level of complexity. For instance, a simplified model of the Wnt- β catenin signaling pathway includes six proteins, not counting phosphorylated species and protein complexes (Lee et al., 2003; Goentoro and Kirschner, 2009). A similar conundrum is looming in the field of synthetic biology, where a plethora of powerful tools for combinatorial engineering have emerged (Morsut et al., 2016; Toda et al., 2018; I. Zhu et al., 2022; R. Zhu et al., 2022; Stevens et al., 2023), begging the question of how these tools

should be combined for practical applications.

Is there a scalable paradigm for designing circuits to achieve a desired goal? Current alternatives to enumeration generally fall into two categories: single-circuit optimization methods and systematic sampling methods. Single-circuit methods such as evolution (François and Hakim, 2004; François, Hakim, and Eric D Siggia, 2007; François and Eric D. Siggia, 2008; François and Eric D. Siggia, 2010), mixed-integer optimization (Otero-Muras and Banga, 2016), and recurrent neural networks (RNNs) (Shen et al., 2021) search the combined fitness landscape (the product space of $S \times \mathbb{R}^{n(s)}$) for optimal topology-parameter combinations. While these methods (particularly RNNs and evolution) seem to scale quite well when designing single circuits, the inference of general principles from their results is nontrivial. The results of an evolutionary trial are strongly influenced by the nature of the fitness landscape, and the hyperparameters of training seem to bias RNNs towards particular classes of solutions (Shen et al., 2021), complicating the recovery of ground truth from these methods. Mixed-integer optimization (the term for joint optimization of both discrete and continuous variables), on the other hand, uncovers pareto-optimal tradeoffs in design space and has stronger convergence guarantees (Otero-Muras and Banga, 2016). However, both the RNN method and mixed-integer optimization require the solution to adopt a specific mathematical form, that of ODEs.

On the other hand, Bayesian methods, in particular approximate Bayesian computation (ABC) (Barnes et al., 2011; Woods et al., 2016), approach the scaling problem by reducing the number of simulations required to evaluate each topology. Approximate Bayesian methods use a type of sampling, based on particle filters, called sequential Monte Carlo (Sisson, Fan, and Tanaka, 2007) to sample the posterior distribution directly. The strengths of ABC are that it avoids sampling low-density regions of the posterior while also quickly building an accurate picture of the quality of each topology. ABC measures the latter as the Bayes factor, roughly equivalent to the odds ratio between the posterior odds and prior odds of a successful phenotype. In order to establish a prior over topologies, however, this method requires listing all possible topologies up to symmetry, a daunting task for design spaces of scale $m \geq 5$ and a computational nightmare for $m \geq 6$.

Overall, methods for circuit topology design are generally either comprehensive in their scope (and thus generalizable to the level of principles and motifs) or scalable to large circuits. This stark trade-off is the motivation for the circuit design platform presented in the next chapter of this thesis.

References

- Alberch, P. (May 1991). “From genes to phenotype: dynamical systems and evolvability”. In: *Genetica* 84.1, pp. 5–11. ISSN: 0016-6707, 1573-6857. DOI: [10.1007/BF00123979](https://doi.org/10.1007/BF00123979). URL: <http://link.springer.com/10.1007/BF00123979> (visited on 03/05/2024).
- Alon, Uri (June 2007). “Network motifs: theory and experimental approaches”. In: *Nature Reviews Genetics* 8.6. Number: 6 Publisher: Nature Publishing Group, pp. 450–461. ISSN: 1471-0064. DOI: [10.1038/nrg2102](https://doi.org/10.1038/nrg2102). URL: <https://www.nature.com/articles/nrg2102> (visited on 09/15/2023).
- (2024). *Systems medicine: physiological circuits and the dynamics of disease*. Boca Raton: CRC Press. ISBN: 978-1-03-241185-9.
- Barnes, Chris P. et al. (Sept. 13, 2011). “Bayesian design of synthetic biological systems”. In: *Proceedings of the National Academy of Sciences* 108.37. Publisher: Proceedings of the National Academy of Sciences, pp. 15190–15195. DOI: [10.1073/pnas.1017972108](https://doi.org/10.1073/pnas.1017972108). URL: <https://www.pnas.org/doi/full/10.1073/pnas.1017972108> (visited on 01/31/2024).
- Chau, Angela H. et al. (Oct. 12, 2012). “Designing Synthetic Regulatory Networks Capable of Self-Organizing Cell Polarization”. In: *Cell* 151.2, pp. 320–332. ISSN: 0092-8674. DOI: [10.1016/j.cell.2012.08.040](https://doi.org/10.1016/j.cell.2012.08.040). URL: <https://www.sciencedirect.com/science/article/pii/S009286741201121X> (visited on 01/31/2023).
- Cotterell, James and James Sharpe (Jan. 2010). “An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients”. In: *Molecular Systems Biology* 6.1. Publisher: John Wiley & Sons, Ltd, p. 425. ISSN: 1744-4292. DOI: [10.1038/msb.2010.74](https://doi.org/10.1038/msb.2010.74). URL: <https://www.embopress.org/doi/full/10.1038/msb.2010.74> (visited on 02/08/2023).
- Dassow, George von et al. (July 2000). “The segment polarity network is a robust developmental module”. In: *Nature* 406.6792. Number: 6792 Publisher: Nature Publishing Group, pp. 188–192. ISSN: 1476-4687. DOI: [10.1038/35018085](https://doi.org/10.1038/35018085). URL: <https://www.nature.com/articles/35018085> (visited on 02/07/2023).
- Diego, Xavier et al. (June 20, 2018). “Key Features of Turing Systems are Determined Purely by Network Topology”. In: *Physical Review X* 8.2. Publisher: American Physical Society, p. 021071. DOI: [10.1103/PhysRevX.8.021071](https://doi.org/10.1103/PhysRevX.8.021071). URL: <https://link.aps.org/doi/10.1103/PhysRevX.8.021071> (visited on 02/10/2023).
- Elowitz, Michael B. and Stanislas Leibler (Jan. 2000). “A synthetic oscillatory network of transcriptional regulators”. In: *Nature* 403.6767. Number: 6767 Publisher: Nature Publishing Group, pp. 335–338. ISSN: 1476-4687. DOI: [10.1038/35002125](https://doi.org/10.1038/35002125). URL: <https://www.nature.com/articles/35002125> (visited on 10/17/2023).

- François, Paul and Vincent Hakim (Jan. 13, 2004). “Design of genetic networks with specified functions by evolution in silico”. In: *Proceedings of the National Academy of Sciences* 101.2. Publisher: Proceedings of the National Academy of Sciences, pp. 580–585. DOI: 10.1073/pnas.0304532101. URL: <https://www.pnas.org/doi/full/10.1073/pnas.0304532101> (visited on 02/07/2023).
- François, Paul, Vincent Hakim, and Eric D Siggia (Jan. 2007). “Deriving structure from evolution: metazoan segmentation”. In: *Molecular Systems Biology* 3.1. Publisher: John Wiley & Sons, Ltd, p. 154. ISSN: 1744-4292. DOI: 10.1038/msb4100192. URL: <https://www.embopress.org/doi/full/10.1038/msb4100192> (visited on 02/07/2023).
- François, Paul and Eric D. Siggia (June 2008). “A case study of evolutionary computation of biochemical adaptation”. In: *Physical Biology* 5.2, p. 026009. ISSN: 1478-3975. DOI: 10.1088/1478-3975/5/2/026009. URL: <https://dx.doi.org/10.1088/1478-3975/5/2/026009> (visited on 02/07/2023).
- (July 15, 2010). “Predicting embryonic patterning using mutual entropy fitness and in silico evolution”. In: *Development* 137.14, pp. 2385–2395. ISSN: 0950-1991. DOI: 10.1242/dev.048033. URL: <https://doi.org/10.1242/dev.048033> (visited on 02/07/2023).
- Gardner, Timothy S., Charles R. Cantor, and James J. Collins (Jan. 20, 2000). “Construction of a genetic toggle switch in *Escherichia coli*”. In: *Nature* 403.6767. MAG ID: 2134759932, pp. 339–342. DOI: 10.1038/35002131.
- Gerardin, Jaline, Nishith R. Reddy, and Wendell A. Lim (Sept. 25, 2019). “The Design Principles of Biochemical Timers: Circuits that Discriminate between Transient and Sustained Stimulation”. In: *Cell Systems* 9.3, 297–308.e2. ISSN: 2405-4712. DOI: 10.1016/j.cels.2019.07.008. URL: <https://www.sciencedirect.com/science/article/pii/S2405471219302650> (visited on 08/31/2023).
- Goentoro, Lea and Marc W. Kirschner (Dec. 11, 2009). “Evidence that Fold-Change, and Not Absolute Level, of β -Catenin Dictates Wnt Signaling”. In: *Molecular Cell* 36.5, pp. 872–884. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2009.11.017. URL: <https://www.sciencedirect.com/science/article/pii/S1097276509008582> (visited on 03/05/2024).
- Jiménez, Alba et al. (2017). “A spectrum of modularity in multi-functional gene circuits”. In: *Molecular Systems Biology* 13.4, p. 925. ISSN: 1744-4292. DOI: 10.15252/msb.20167347.
- Lee, Ethan et al. (Oct. 13, 2003). “The Roles of APC and Axin Derived from Experimental and Theoretical Analysis of the Wnt Pathway”. In: *PLOS Biology* 1.1. Publisher: Public Library of Science, e10. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0000010. URL: <https://journals.plos.org/>

- [plosbiology/article?id=10.1371/journal.pbio.10000010](https://doi.org/10.1371/journal.pbio.10000010) (visited on 03/05/2024).
- Lewontin, Richard C. (1974). *The genetic basis of evolutionary change*. Columbia biological series 25. New York: Columbia Univ. Pr. 346 pp. ISBN: 978-0-231-08318-8.
- Ma, Wenzhe et al. (Aug. 21, 2009). “Defining Network Topologies that Can Achieve Biochemical Adaptation”. In: *Cell* 138.4, pp. 760–773. ISSN: 0092-8674. DOI: 10.1016/j.cell.2009.06.013. URL: <https://www.sciencedirect.com/science/article/pii/S0092867409007120> (visited on 02/07/2023).
- Morsut, Leonardo et al. (Feb. 11, 2016). “Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors”. In: *Cell* 164.4. Publisher: Elsevier, pp. 780–791. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.01.012. URL: [http://www.cell.com/cell/abstract/S0092-8674\(16\)00052-0](http://www.cell.com/cell/abstract/S0092-8674(16)00052-0) (visited on 07/14/2021).
- Müller, Gerd B. and Stuart A. Newman (2003). *Origination of organismal form: beyond the gene in developmental and evolutionary biology*. The Vienna series in theoretical biology. Cambridge: MIT press. ISBN: 978-0-262-13419-4.
- Otero-Muras, Irene and Julio R. Banga (Dec. 15, 2016). “Design Principles of Biological Oscillators through Optimization: Forward and Reverse Analysis”. In: *PLOS ONE* 11.12. Publisher: Public Library of Science, e0166867. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0166867. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166867> (visited on 10/19/2023).
- Pfeifer, Charlotte R., Amy E. Shyer, and Alan R. Rodrigues (Feb. 1, 2024). “Creative processes during vertebrate organ morphogenesis: Biophysical self-organization at the supracellular scale”. In: *Current Opinion in Cell Biology* 86, p. 102305. ISSN: 0955-0674. DOI: 10.1016/j.ceb.2023.102305. URL: <https://www.sciencedirect.com/science/article/pii/S0955067423001540> (visited on 03/05/2024).
- Schaerli, Yolanda et al. (Sept. 23, 2014). “A unified design space of synthetic stripe-forming networks”. In: *Nature Communications* 5, p. 4905. ISSN: 2041-1723. DOI: 10.1038/ncomms5905.
- Shen, Jingxiang et al. (May 25, 2021). “Finding gene network topologies for given biological function with recurrent neural network”. In: *Nature Communications* 12.1. Number: 1 Publisher: Nature Publishing Group, p. 3125. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23420-5. URL: <https://www.nature.com/articles/s41467-021-23420-5> (visited on 02/07/2023).
- Sisson, S. A., Y. Fan, and Mark M. Tanaka (Feb. 6, 2007). “Sequential Monte Carlo without likelihoods”. In: *Proceedings of the National Academy of Sciences* 104.6. Publisher: Proceedings of the National Academy of Sciences, pp. 1760–1765.

- DOI: 10.1073/pnas.0607208104. URL: <https://www.pnas.org/doi/10.1073/pnas.0607208104> (visited on 03/05/2024).
- Stevens, Adam J. et al. (Feb. 2023). “Programming multicellular assembly with synthetic cell adhesion molecules”. In: *Nature* 614.7946. Number: 7946 Publisher: Nature Publishing Group, pp. 144–152. ISSN: 1476-4687. DOI: 10.1038/s41586-022-05622-z. URL: <https://www.nature.com/articles/s41586-022-05622-z> (visited on 03/24/2023).
- Thompson, D’Arcy Wentworth (1917). *On growth and form*.
- Toda, Satoshi et al. (July 13, 2018). “Programming self-organizing multicellular structures with synthetic cell-cell signaling”. In: *Science (New York, N.Y.)* 361.6398, pp. 156–162. ISSN: 1095-9203. DOI: 10.1126/science.aat0271.
- Vienne, Dominique de (Aug. 1, 2022). “What is a phenotype? History and new developments of the concept”. In: *Genetica* 150.3, pp. 153–158. ISSN: 1573-6857. DOI: 10.1007/s10709-021-00134-6. URL: <https://doi.org/10.1007/s10709-021-00134-6> (visited on 03/05/2024).
- Wolpert, L. (Oct. 1, 1969). “Positional information and the spatial pattern of cellular differentiation”. In: *Journal of Theoretical Biology* 25.1, pp. 1–47. ISSN: 0022-5193. DOI: 10.1016/S0022-5193(69)80016-0. URL: <https://www.sciencedirect.com/science/article/pii/S0022519369800160> (visited on 02/18/2022).
- Woods, Mae L. et al. (June 17, 2016). “A Statistical Approach Reveals Designs for the Most Robust Stochastic Gene Oscillators”. In: *ACS Synthetic Biology* 5.6. Publisher: American Chemical Society, pp. 459–470. DOI: 10.1021/acssynbio.5b00179. URL: <https://doi.org/10.1021/acssynbio.5b00179> (visited on 10/13/2023).
- Zhu, Iowis et al. (Apr. 14, 2022). “Modular design of synthetic receptors for programmed gene regulation in cell therapies”. In: *Cell* 185.8, 1431–1443.e16. ISSN: 0092-8674. DOI: 10.1016/j.cell.2022.03.023. URL: <https://www.sciencedirect.com/science/article/pii/S0092867422003336> (visited on 03/05/2024).
- Zhu, Ronghui et al. (Jan. 21, 2022). “Synthetic multistability in mammalian cells”. In: *Science* 375.6578. Publisher: American Association for the Advancement of Science, eabg9765. DOI: 10.1126/science.abg9765. URL: <https://www.science.org/doi/10.1126/science.abg9765> (visited on 04/14/2023).

Chapter 2

DESIGNING BIOCHEMICAL CIRCUITS WITH TREE SEARCH

Abstract

Finding biochemical circuits that exhibit a desired behavior is an outstanding problem in biological engineering. The canonical approach of enumerating all possible circuit topologies becomes intractable for circuits with more than four components due to combinatorial scaling. Here, we use the game-playing algorithm Monte Carlo Tree Search (MCTS) to optimize circuit topology for a target phenotype by framing circuit design as a game of stepwise assembly. When tasked with designing a three-component oscillator, our MCTS-based design framework, which we call CircuiTree, efficiently and comprehensively identifies robust topologies by prioritizing sparse architectures. CircuiTree also mines tree search results for overrepresented assembly motifs, even identifying obscure, yet optimal solutions. Finally, we scale the search up to five-component circuits and find that, using a strategy we call multiplexing, oscillators can tolerate the failure of components by interleaving multiple oscillatory motifs in a manner reminiscent of natural circadian clocks. Overall, CircuiTree provides a scalable, open-source platform for investigating the design principles of biochemical circuits.

Introduction

Understanding how biochemical networks produce biological behavior is one of the central goals in systems and synthetic biology. Studying the design principles of these networks, or circuits, provides fundamental insights into their computational capabilities and is a critical step for applying synthetic biology to realms such as cell-based therapeutics (Lim, C. M. Lee, and Tang, 2013; Williams et al., 2020) and microbial engineering (Jones, Marken, and P. A. Silver, 2024; Purnick and Weiss, 2009). Deciding the topology, or the qualitative network architecture, is a critical first step for engineering synthetic gene circuits to achieve a functional goal such as switching in response to an input (Gardner, Cantor, and Collins, 2000), spontaneous oscillations (Elowitz and Leibler, 2000), or transcriptional multistability (Zhu et al., 2022). Currently, this aspect of design is primarily guided by reverse engineering of natural circuits, careful hand-design, or (when all else fails) computational enumeration. The canonical approach for discovering design principles *de novo*,

enumeration involves listing all the topologies in a predefined space of possibilities and simulating their behavior with an exhaustive set of random perturbations of the modeling parameters (Ma et al., 2009; Cotterell and Sharpe, 2010; Chau et al., 2012; Schaerli et al., 2014; Gerardin, Reddy, and Lim, 2019). The enumeration method yields a comprehensive picture of the combined fitness landscape of topologies and parameters and identifies both robust individual topologies (in the sense of parameter robustness) and their shared network motifs that exemplify the design principles of the phenotype (Alon, 2007; Alon, 2019).

Currently, the computational complexity of the enumeration method has limited the role of automated circuit design. While suitable for smaller circuits, an exhaustive search becomes computationally difficult for circuits with 4 components and intractable for 5 components. The number of distinct circuit topologies scales (modulo symmetries) as $O(3^{N^2})$, where N is the number of circuit components and 3 represents the possible pairwise interactions (activation, inhibition, or no interaction). Typically, each topology must be simulated with $\geq 10^4$ parameter perturbations, so enumeration requires on the order of 10^{10} - 10^{11} simulations for $N = 4$ and 10^{15} - 10^{16} simulations for $N = 5$ components. This "curse of dimensionality" constitutes the main barrier to understanding the structure-function relationship for large circuits. The search space can be narrowed by modularly combining smaller motifs (Qiao, Zhao, et al., 2019), much like building blocks, but this approach risks excluding novel solutions to the design problem that appear at higher complexity (Jiménez et al., 2017). For large circuits, a functional topology can still be found with evolutionary optimization (François, Hakim, and Eric D Siggia, 2007; François and Eric D. Siggia, 2010) or recurrent neural network-guided inference (Shen et al., 2021), but the topologies discovered with these methods can depend on the optimization parameters and the nature of the fitness landscape, a fact that complicates the inference of more general principles. The ideal design framework for circuit topologies would be: (i) *Efficient*: Using relatively few samples, it would return a set of reasonably robust solutions. (ii) *Generalizable*: It would be able to infer motifs for the design problem. (iii) *Scalable*: When searching a large space of topologies, it does not require enumeration and appropriately maintains features (i) and (ii).

Our inspiration for a circuit design approach with these features comes from algorithms developed in modern artificial intelligence to solve decision problems, which share core computational similarities with genetic circuit design. In a decision problem, or game, one starts from an initial position (the root state) and makes a

sequence of decisions ending in success or failure. Algorithms for optimal game playing must traverse the branching tree of possible decisions and find the best decision paths. Large game trees are characteristic of long and/or complex games like chess ($\sim 10^{43}$ states) and Go (10^{170} states) and can overwhelm classical exhaustive search methods (e.g., flat search, A^* , and minimax) and branch-and-bound methods (e.g., α - β pruning) due to their computation and memory requirements (Shannon, 1950; Tromp and Farnebäck, 2007). In contrast to comprehensive search methods, probabilistic search methods, such as Monte Carlo tree search (MCTS), sample paths in the game tree based on an optimistic estimate of the success probability for each decision branch (Kocsis and Szepesvári, 2006). MCTS has proven to be instrumental in solving large games (D. Silver, Huang, et al., 2016; D. Silver, Hubert, et al., 2018) due to its low memory footprint, balance of exploitation and exploration (Auer, Cesa-Bianchi, and Fischer, 2002), and myriad variations (Świechowski et al., 2022; Browne et al., 2012).

In this work, we present a circuit design platform, which we call CircuiTree, based on tree search. CircuiTree approaches circuit design as a sequence of assembly decisions and optimizes the circuit topology for a given phenotype by using MCTS to search the tree of possible circuit assemblies. We first use CircuiTree to search the space of 3,325 connected circuits with three transcription factors in search of topologies that exhibit spontaneous, sustained oscillations. As it dynamically explores the design space, CircuiTree discovers sparse oscillators before complex ones and reliably identifies the most robust topologies. Next, we present a method for identifying assembly motifs, which are successful circuit topologies that are statistically overrepresented among successful assembly paths. We show that the best assembly motifs for 3-component oscillators form a cluster in the decision tree downstream of constructing the activator-inhibitor (AI) and repressilator (Rep) motifs, consistent with the canonical design principles for 2- and 3-component oscillators. Finally, we use CircuiTree to find five-component circuits that oscillate when subjected to random deletion of a component. Fault-tolerant oscillators contain multiple AI and Rep motifs that are synergistically interleaved to provide robustness to deletions and single-gene knockdowns. Our results demonstrate that CircuiTree's assembly framework enables the efficient, generalizable, and scalable design of biochemical circuits for arbitrary phenotypes, which could be used to engineer and understand complex intracellular and multicellular biological systems.

Results

Circuit design as an assembly game

We define a circuit as a system of k dynamically interacting biochemical species $\mathbf{x} = [x_0, \dots, x_k]$ characterized by its topology s , a member of the set of possible topologies \mathcal{S} , and its parameters $\theta \in \mathbb{R}^n$, which describe the rates of biochemical reactions. For convenience, we consider systems where all topologies share the same set of reactions and parameters, but this is not a necessary assumption. We also assume the existence of a phenotype function $f(\mathbf{x}(t)) = q \in \{0, 1\}$ that delivers a yes-or-no verdict as to whether the circuit exhibits a desired phenotype. We conceptualize this function quite generally to include dynamical behaviors such as adaptation or oscillation, as well as other properties such as multistability. Like many prior computational studies of design principles (Ma et al., 2009; Chau et al., 2012; Cotterell and Sharpe, 2010; Schaeferli et al., 2014), each simulation of the system is performed with a random parameter set, and we define the robustness of a topology as the mean phenotype score from these samples $Q(s) = \mathbb{E}_\theta[q(s)] = \mathbb{E}_\theta[f(\mathbf{x}(t; s, \theta))]$. Thus, we approach circuit design as a combinatorial optimization problem in search of the most robust topology $s^* = \operatorname{argmax}_s Q(s)$. Note that Q in this context should not be confused with the value function of the same name used in reinforcement learning.

We frame circuit topology design as a multi-step decision problem, or game (Figure 1A). Beginning with an "empty" circuit s_0 with a given set of components and without interactions, each step of the game can add an activating or inhibitory (auto-)regulatory interaction. The game ends when the builder chooses to "terminate" at some topology s_i , at which point the outcome is decided by a simulation with win probability $Q(s_i)$. In the language of Markov decision processes (MDPs), each topology s represents a state of the game, and the addition of an interaction or the termination of assembly represents an action a that can be taken. We can conceptualize each trial of the game as tracing a path on a decision tree T rooted at s_0 , where each node in the tree $s \in \mathcal{S}$ represents a topology and each directed edge $(s_i, s_j) \in \mathcal{E}$ represents the assembly of s_j from s_i by, for example, adding an inhibitory interaction between two components. Because multiple decision paths can converge on the same s_j , T is technically a directed acyclic graph (DAG); nonetheless, we will call it a tree for simplicity.

For large design spaces, it is infeasible to estimate Q for every s in T . For example, a five-component regulatory circuit has 5^2 potential pairwise interactions, each of

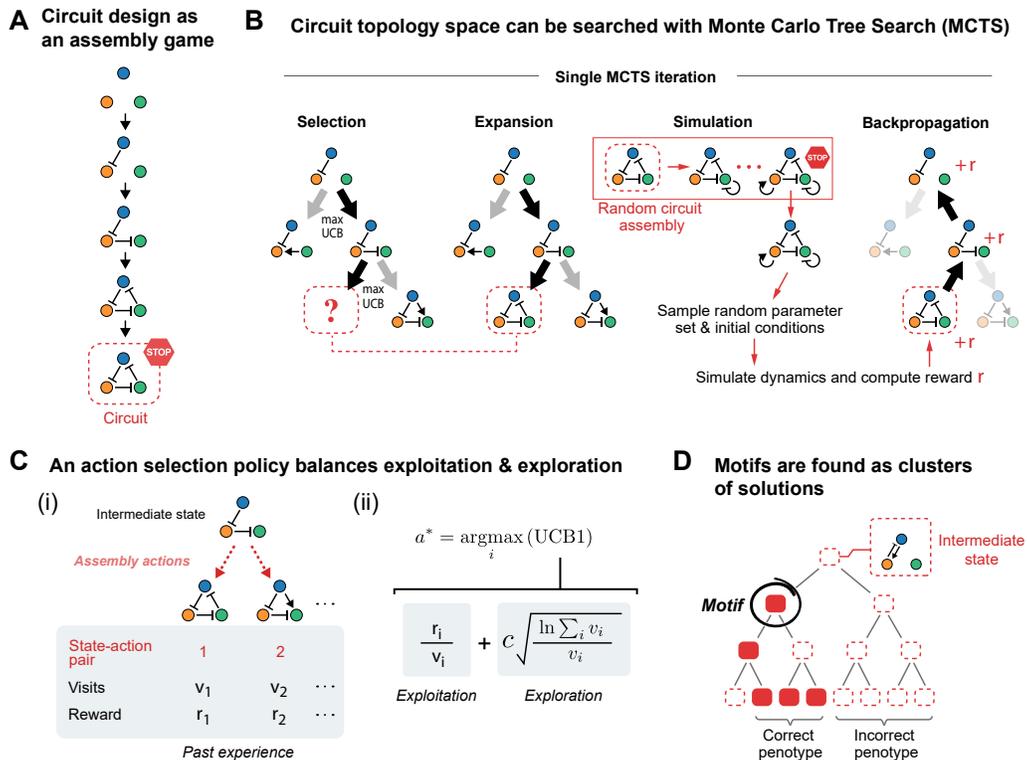


Figure 1: A stepwise assembly framework enables circuit topology optimization with tree search. (A) Circuit topologies are built step-by-step by adding interactions until the game is ended by taking the “terminate” action (the STOP sign). (B) Each MCTS iteration undergoes four phases: (1) Selection: The UCT criterion is used to recursively select the most promising action a_i^* from the current state s_i . (2) Expansion: If the edge (s_i, a_i^*) has not been sampled yet, it is added to the tree. (3) Simulation: If the circuit has not been completed yet, take random assembly moves until “terminate” is chosen. Simulate the resulting topology with random parameters and evaluate the reward q based on a target phenotype. (4) Backpropagation: update the history of rewards r_{ij} over past visits v_{ij} for the visited edges (s_i, a_i^*) . (C) The UCT policy for selecting a_i^* balances exploitation and exploration based on past trials. The exploration term, tuned by a hyperparameter c , favors actions that are under-sampled relative to the alternatives. (D) Once assembled, a motif creates an enriched subgraph in the search graph.

which can be absent, activating, or inhibitory, leading to $|\mathcal{S}| = 3^{5^2} \approx 8.5 \times 10^{11}$ topologies before accounting for symmetries. To address this gap, we use Monte Carlo tree search (MCTS), a reinforcement learning (RL) strategy that is applied heavily in MDPs and game-playing artificial intelligences (D. Silver, Huang, et al., 2016; D. Silver, Schrittwieser, et al., 2017; D. Silver, Hubert, et al., 2018) to bound the computational complexity of searching large decision trees. In these applications, tree search is performed during game-play (and in some cases also during training) to plan future moves by simulating a series of games forward in

time. Each iteration of MCTS starts at the root state and undergoes four phases, shown in Figure 1B and outlined in Algorithm 1. The key innovation of MCTS is how it allocates computational resources during the first phase (selection) to balance the competing goals of (i) exploiting actions that have yielded high rewards in the past and (ii) exploring new actions that may yield even higher rewards. At each decision branch (Figure 1C, left), from the current state s_i , there are multiple possible actions a_j that have each yielded a cumulative reward $r_{i,j}$ over $v_{i,j}$ past visits. MCTS decides the optimal action using the decision policy $a_i^* = \operatorname{argmax}_j \operatorname{UCT}(i, j)$ (Figure 1C, right), originally developed to solve the classic multi-armed bandit problem Auer, Cesa-Bianchi, and Fischer, 2002.

$$\operatorname{UCT}(i, j) = \frac{r_{i,j}}{v_{i,j}} + c \sqrt{\frac{\ln \sum_j v_{i,j}}{v_j}} \quad (2.1)$$

The first term of the UCT criterion estimates the mean reward for (i, j) , while the second term, derived from Hoeffding’s inequality (Kocsis and Szepesvári, 2006), is an upper bound estimate of the sampling error. Thus UCT is an optimistic prediction of the underlying mean reward (for unvisited edges, $\operatorname{UCT}(i, j) = +\infty$). Note that the first term encourages exploitation and the second exploration. Because the second term is larger for an action that is chosen less frequently ($\sum_j v_{ij} \gg v_j$), even seemingly sub-optimal actions will be chosen occasionally to account for sampling error. Actions are selected recursively, breaking any ties randomly, until an unsampled or terminal state is encountered, and for an unvisited state, that state is added to T . Then, the outcome q of the trial is determined by taking random assembly actions until the game ends (“terminate” is chosen) and simulating the resulting topology with a random parameter set and initial conditions. Finally, the record of rewards and visits is updated for each selected edge in T (backpropagation).

Tree search is also a natural paradigm for identifying motifs, or structural patterns that recur among topologies that successfully generate the target phenotype. In games, there typically exists a specific set of moves (a strategy) that lead to an entire subtree of T with largely favorable outcomes, and search algorithms like MCTS seek to efficiently identify and exploit these winning regions. In the game of circuit assembly, network motifs (Alon, 2007) play a similar role. Extensive literature has demonstrated that, for a range of circuit behaviors including adaptation (Ma et al., 2009), morphogen patterning (Cotterell and Sharpe, 2010; Schaerli et al., 2014), and single-cell polarization (Chau et al., 2012), successful topologies tend to contain

Algorithm 1 Circuit topology search using MCTS

```

procedure MONTECARLOTREESearch( $s_0, N$ )
   $T \leftarrow$  EMPTYTREE()
  ADDNODE( $T, s_0$ )            $\triangleright$  Initialize  $T$  with the root topology  $s_0$ 
  repeat
    [ $s_0 \cdots s_n$ ]  $\leftarrow$  SELECT( $T, s_0$ )  $\triangleright$  Choose assembly moves and select the
                                          last topology
    if  $s_n$  not in NODES( $T$ ) then  $\triangleright$  Expand the search graph if necessary
      ADDNODE( $T, s_n$ )
    end if
    if ( $s_{n-1}, s_n$ ) not in EDGES( $T$ ) then
      ADDEDGE( $T, (s_{n-1}, s_n)$ )
    end if
     $s_{\text{sim}} \leftarrow s_n$             $\triangleright$  Make random moves until termination
                                          occurs
    while ISNONTERMINAL( $s_{\text{sim}}$ ) do
       $s_{\text{sim}} \leftarrow$  DORANDOMMOVE( $s_{\text{sim}}$ )
    end while
     $q \leftarrow$  SIMULATE( $s_{\text{sim}}$ )      $\triangleright$  Simulate  $s_{\text{sim}}$  and obtain a reward  $q$ 
    for  $i := 0$  to  $n - 1$  do        $\triangleright$  Backpropagate the results of each deci-
                                          sion
       $v[s_i, s_{i+1}] \leftarrow v[s_i, s_{i+1}] + 1$ 
       $r[s_i, s_{i+1}] \leftarrow r[s_i, s_{i+1}] + q$ 
    end for
    until  $N$  iterations or TIMEOUT
    return  $T, v, r$ 
end procedure

procedure SELECT( $T, s_0$ )
   $i \leftarrow 0$ 
  while NONTERMINAL( $s_i$ ) do
     $S' \leftarrow$  POSSIBLEMOVES( $s_i$ )
     $s_{i+1} \leftarrow \operatorname{argmax}_{s_j \in S'} \text{UCB}(s_i, s_j)$   $\triangleright$  Choose the assembly move that maxi-
                                          mizes the UCB score
    if  $v[s_i, s_{i+1}] == 0$  then  $\triangleright$  Until a new move is encountered
      return [ $s_0 \cdots s_{i+1}$ ]
    end if
     $i \leftarrow i + 1$ 
  end while  $\triangleright$  Or termination occurs
  return [ $s_0 \cdots s_{i+1}$ ]
end procedure

```

at least one motif, creating clusters of topologically similar solutions. Thus, each motif creates an enriched region in the search graph T that may be discovered and exploited during tree search (Figure 1D). To identify these motifs from the results of a tree search, we developed a statistical test for overrepresentation that is described in detail in the Methods section. CircuiTree is available online as an open-source Python package (<https://github.com/pranav-bhamidipati/circuitree>).

Establishing a ground-truth for three-node stochastic oscillators with enumeration

Oscillations appear in diverse cellular contexts such as metabolism (Chandra, Buzi, and Doyle, 2011), DNA damage (Geva-Zatorsky, Rosenfeld, et al., 2006; Geva-Zatorsky, Dekel, et al., 2010), the cell cycle (Ferrell, Tsai, and Q. Yang, 2011), and circadian rhythms (Tyson et al., 1999), and consequently, many of their design principles have been elucidated (Novák and Tyson, 2008). We first benchmark CircuiTree on the well-studied problem of designing an oscillator circuit with three nodes (Figure 2A, part i). Specifically, we consider a system of symmetric transcription factors (TFs) modeled as a stochastic birth-and-death process of individual mRNAs, TFs, and TF-response element (TF-RE) complexes, with elementary reactions of transcription, translation, degradation, binding, and unbinding (Figure S1A; reactions and their rate parameters are summarized in Table 2.1). TFs regulate transcription by binding to one of two REs in the regulatory region of each promoter, and cooperativity is introduced by assuming that TF-RE binding is stronger when both sites are occupied by the same TF. This assumption reduces the computational cost associated with modeling every homo- and heterodimer and their dimer-RE complexes. For each stochastic simulation, the system was randomly initialized and stochastically simulated using the Gillespie method for $t_{\max} = 4 \times 10^4 \text{ sec} = 11.1 \text{ hrs}$ (see Methods for details of initialization). Oscillations were quantified by computing the normalized autocorrelation function (ACF) and identifying its lowest minimum value across all TFs ACF_{\min} , excluding the bounds, and a Boolean reward value was assigned based on a cutoff value (Figure 2A, part ii).

$$r = \begin{cases} 1 & \text{if } \text{ACF}_{\min} < -0.4 \\ 0 & \text{if } \text{ACF}_{\min} > -0.4 \end{cases} \quad (2.2)$$

To compare with enumeration, all 3,325 unique, fully connected topologies were simulated with 10^4 randomly sampled parameter sets and initial conditions. The

10 rate parameters of the model were reduced to 8 dimensionless variables, which were sampled uniformly from a range of values determined based on experimentally measured rates (Ron Milo et al., 2010) and known requirements for oscillations (Elowitz and Leibler, 2000; Novák and Tyson, 2008). (See Table 2.2 for variable definitions and sampling ranges and Methods for details of parameter sampling). The robustness to parameter perturbation Q was calculated as the fraction of parameter sets that produced oscillation, and topologies with $Q > 0.01$ were considered oscillators. This procedure uncovered 221 oscillator topologies (6.65% of the search space) (see Figures S1B and S1C for a summary of the best topologies). These results generally agree with prior studies of two- and three-node oscillators (Qiao, Zhang, et al., 2022; Elowitz and Leibler, 2000; Novák and Tyson, 2008; Woods et al., 2016; Stricker et al., 2008). For instance, almost all oscillators contain a repressilator (Rep) loop, activator-inhibitor (A-I) loop, or a combination thereof. Positive autoregulation (PAR), which has been shown to buffer extrinsic and intrinsic noise (Qiao, Zhang, et al., 2022), is also ubiquitous among these oscillators and is required for robust A-I loop oscillations (Novák and Tyson, 2008; Stricker et al., 2008). Further discussion of these motifs can be found in the section below. Interestingly, stochastic switching between stable states was occasionally mistaken for low-frequency oscillations (Figure S2), causing some toggle switch circuits to be classified as oscillators (highest $Q = 0.024$, ranked #139).

Reaction	Rate parameter	Value in Figs. 4D & 5
TF-RE binding	k_{on}	1.0 sec^{-1}
TF-RE unbinding (uncooperative)	$k_{\text{off},1}$	$99.0 \text{ molec}^{-1} \text{ sec}^{-1}$
TF-RE unbinding (cooperative)	$k_{\text{off},2}$	$9.9 \text{ molec}^{-1} \text{ sec}^{-1}$
Transcription, basal	k_{unbound}	0.05 sec^{-1}
Transcription, 1-2 activators	k_{act}	8.0 sec^{-1}
Transcription, 1-2 inhibitors	k_{inh}	$5 \times 10^{-4} \text{ sec}^{-1}$
Transcription, 1 each	k_{mixed}	0.05 sec^{-1}
Translation	k_p	$0.167 \text{ molec}^{-1} \text{ sec}^{-1}$
Degradation, mRNA	γ_m	$0.025 \text{ molec}^{-1} \text{ sec}^{-1}$
Degradation, protein	γ_p	$0.025 \text{ molec}^{-1} \text{ sec}^{-1}$

Table 2.1: Model rate parameters

Across all oscillators, oscillation was favored by tight TF-RE binding ($K_{D,1} < 10^3$), low basal transcription ($k_{m,\text{unbound}} < 0.3 \text{ sec}^{-1}$), strong repression ($k_{m,\text{rep}}/k_{m,\text{unbound}} < 10^{-1}$), and a high activated transcription rate ($k_{m,\text{act}}$, monotonic effect) (Figure

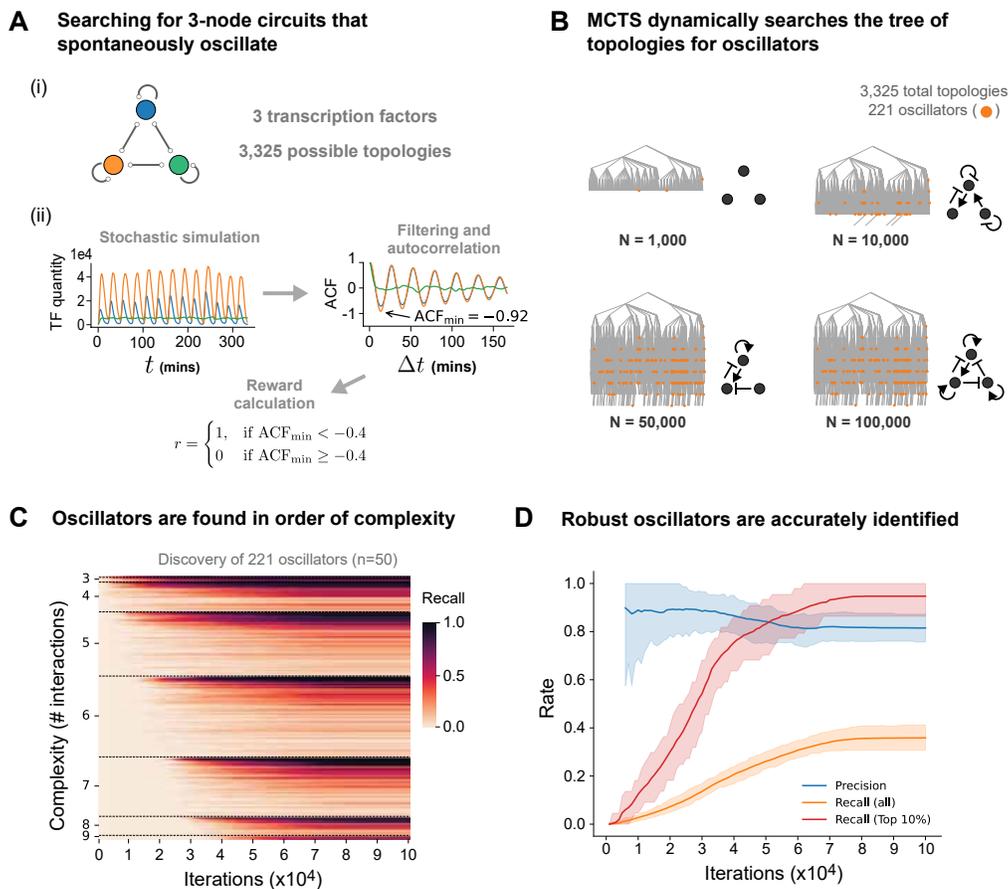


Figure 2: CircuiTree efficiently identifies simple and robust 3-component oscillators. (A) All three-component transcription factor (TF) circuits (3,325 up to symmetry) were enumerated with 10^4 random parameter sets (i) and evaluated for oscillations (ii) using an autocorrelation-based reward function. (B) A representative MCTS run. With more iterations (N), the search graph T (represented by a spanning tree for simplicity) expands to encounter more oscillators (orange circles) and improve its best predicted oscillator topology (shown in black). (C) A heatmap showing the average rate of discovery, or recall, for each oscillator (proportion of $n = 50$ replicates). Rows (oscillators) are sorted in order of complexity, or the number of interactions, and oscillators with the same complexity are sorted by descending robustness Q . Sparse oscillators are found before more complex ones, with a preference for the most robust candidates. (D) Precision (blue) and recall (orange) of oscillator classification (mean \pm 95% CI, $n = 50$). CircuiTree’s recall is particularly high for the 10% most robust oscillators (red), reaching 94.7% after 10^5 iterations. See also Figures S1, S2, S3, and S4.

S3A). Protein and mRNA degradation rates had a non-monotonic effect with a peak at $\gamma_m \approx \gamma_p \approx 0.2 \text{ molec}^{-1} \text{ sec}^{-1}$ (Figure S3A), and oscillation period depended strongly on these rates and their ratio (Figures S3B and S3C). However, the best oscillators were exceptionally robust to the choice of parameters. For example,

the most robust oscillator (the repressilator with PAR) has a Q of 0.767, meaning that, with 8 sampled variables, a randomly sampled variable a 96.7% chance of permitting oscillations on average ($0.967^8 \approx 0.767$).

Dimensionless variable	Definition	Sampling limits
κ_1	$\log_{10} [K_{D,1} \cdot 1 \text{ molec}]$	$[-2, 4]$
κ_2	$K_{D,2}/K_{D,1}$	$[0.00, 0.25]$
k'_{unbound}	$k_{\text{unbound}} \cdot 1 \text{ sec}$	$[0, 1]$
k'_{act}	$k_{\text{act}} \cdot 1 \text{ sec}$	$[1, 10]$
r_{inh}	$\log_{10} [k_{\text{inh}}/k_{\text{act}}]$	$[0, 5]$
k'_p	$k_p \cdot 1 \text{ molec} \cdot 1 \text{ sec}$	$[0.015, 0.250]$
γ'_m	$\gamma_m \cdot 1 \text{ molec} \cdot 1 \text{ sec}$	$[0.001, 0.1]$
γ'_p	$\gamma_p \cdot 1 \text{ molec} \cdot 1 \text{ sec}$	$[0.001, 0.1]$

Table 2.2: Dimensionless variables and limits imposed on random parameter sampling

CircuiTree efficiently and systematically discovers three-node oscillators

MCTS masters a game with a limited number of trials by balancing deep sampling of promising regions with broad sampling of under-explored regions. To understand how this strategy performs for circuit topology design, we use CircuiTree to search for 3-node oscillators given $N = 10^5$ iterations of MCTS ($n = 50$ replicates; see Methods for MCTS implementation specifics). In the average run, the first putative result is discovered after just 0.69 samples/topology (2,280 iterations), and by 19.6 samples/topology (65,360 iterations), a top-5 oscillator has been sampled >100 times. A topology is discovered as a “successful” oscillator if its estimated robustness $\hat{Q}_j = r_j/v_j$ exceeds the threshold value $Q_{\text{thresh}} = 0.01$, which was chosen heuristically prior to the search. Figure 2B shows a representative example of how, over sampling time, CircuiTree incrementally builds the search tree from the root state (top to bottom), encounters more oscillators (shown in orange), and improves its prediction for the most robust design (shown as a black circuit diagram). By the end of sampling time, the average run saturates the tree, sampling 99.97% of topologies at least once.

CircuiTree balances efficiency and comprehensiveness by first finding sparse solutions before exploring deeper areas of the tree. In Figure 2C, each row of the

heatmap is one of the 221 oscillators, and the rows are sorted first by the circuit’s complexity, or the number of interactions, then by decreasing robustness (Q). The color scale indicates the likelihood of discovery, or recall, of each oscillator measured as the proportion of replicates that discovered it. Because sparse topologies require fewer assembly steps, MCTS encounters the sparsest oscillators (such as the repressilator) first, and it discovers increasingly complex solutions over time until oscillators with 9 interactions (the maximum) are found at $N \approx 5 \times 10^4$. As shown by the line plot in Figure 2D, CircuiTree has a very high recall of 94.7% for the top 10% of oscillators (95% CI: 86.4% – 100.0%) and has a recall of 35.9% (95% CI: 30.8% – 41.5%) for oscillators in general. This is fairly high, considering that it gets an average of $\bar{3}0$ samples/topology, and a majority of oscillators have a Q of less than $1/30$. Additionally, CircuiTree has a precision of 81.5% (95% CI: 74.1% – 88.3%), indicating a low rate of false positives.

The competing goals of breadth and depth manifest during tree search as distinct temporal phases in which MCTS first explores a broad set of topologies before focusing on a narrow, enriched subset. Among topologies containing a combination of AI and Rep feedback loops, 27.2% (47/173) are oscillators, a very high proportion compared to 6.7% (221/3,325) for the entire design space and 9.3% (146/1,575) and 8.0% (6/75) for the AI or Rep loops alone (Figure S4A). Figure S4B shows how MCTS allocates samples to each of these categories over a total sampling time of 10^5 and 10^6 iterations (mean of $n = 50$ and $n = 12$ replicates, respectively). During an initial exploratory phase, samples are taken broadly across categories; however, at $N \approx 6 \times 10^4$, the AI-Rep combination becomes heavily favored, followed by Rep alone, for the rest of sampling time. This transition can be observed directly from the search history by measuring regret. Defined as

$$R_N = NQ^* - \sum_{n=1}^N r_n, \quad (2.3)$$

regret is the difference between the expected reward from sampling the best topology (with robustness Q^*) and the actual reward over N iterations. As MCTS moves from exploration into exploitation, the reward rate becomes higher and regret accumulates more slowly (Figures S4C and S4D). Thus, the discovery of an enriched region during search triggers a shift to more focused sampling that can be identified in real-time without *a priori* knowledge about design principles or motifs.

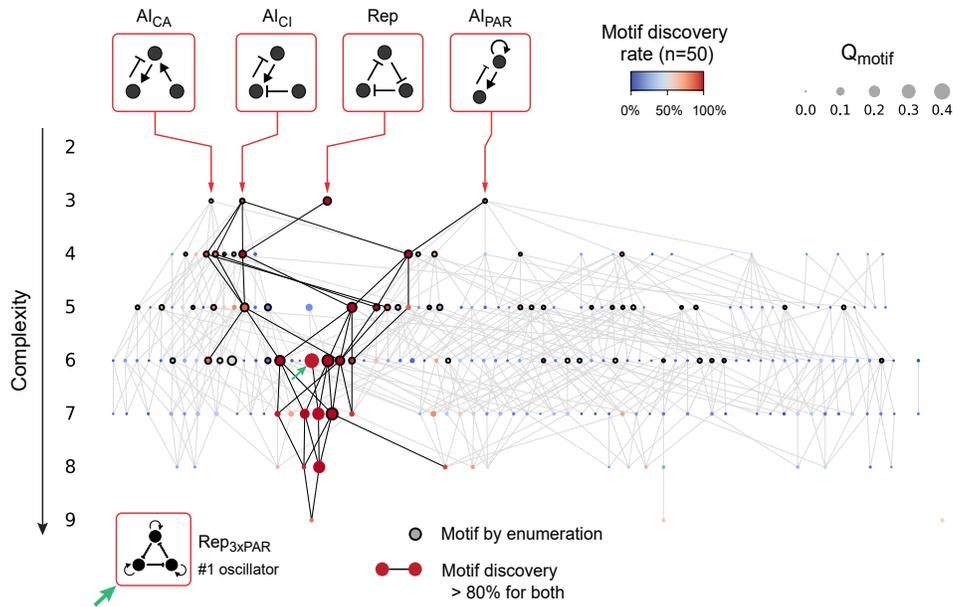


Figure 3: Motifs identified from search results form a cluster of optimal 3-node oscillators. A complexity atlas of oscillators with ≤ 3 components. Circles are oscillator topologies identified by enumeration, and edges link oscillators that differ by the addition/removal of one interaction. 97.7% of oscillators (216/221) are topologically related to one of the four motifs for 3-node oscillation, shown above the atlas in red boxes. Bold circle borders indicate oscillators found to be motifs based on enumeration. Circle color indicates the rate with which CircuiTree labels each oscillator as an assembly motif. Circle size indicates Q_{motif} , the average robustness for a circuit completed randomly starting from this state of the assembly game. The correlation between discovery rate and Q_{motif} (plotted in Figure S5B) suggests that motifs found by CircuiTree correspond to beneficial game states. The bolded edges, which connect oscillators with a discovery rate $> 80\%$, form a contiguous cluster representing optimal assembly strategies. The most robust oscillator, the repressilator with PAR of all components, is shown on the bottom-left and indicated on the atlas by a green arrow. See also Figures S4 and S5.

CircuiTree infers oscillator motifs from search results

In addition to finding individual circuit topologies, an important goal of circuit design is to identify structural features that underlie successful designs, often in the form of network motifs (R. Milo et al., 2002; Alon, 2007; Ma et al., 2009; Cotterell and Sharpe, 2010; Shah and Sarkar, 2011; Chau et al., 2012; Lim, C. M. Lee, and Tang, 2013; Schaerli et al., 2014). CircuiTree uses overrepresentation analysis (described in Methods and shown schematically in Figure S5A) to mine the results of a tree search for motifs. We define a motif, similarly to prior studies, as any successful topology (that is, $Q > Q_{\text{thresh}}$) that is significantly overrepresented as a pattern within other successful topologies.

Looking for 3-node oscillator motifs with ≤ 3 interactions, CircuiTree identifies the same four motifs as enumeration, shown at the top of Figure 3: the repressilator motif (Rep) and the activator-inhibitor (AI) loop with either PAR of the activator (AIPAR), constitutive inhibition of the inhibitor (AICI), or constitutive activation of the activator (AICA). CircuiTree finds these minimal motifs in 100% (Rep), 86% (AIPAR), 94% (AICI), and 62% (AICA) of replicates. To see how these motifs are situated in the overall design space, we plot all 221 oscillators in Figure 3 as a “complexity atlas” (Cotterell and Sharpe, 2010; Schaerli et al., 2014), a graph-of-circuits where every oscillator is a node and nodes are organized in layers according to their complexity. Edges connect nodes in adjacent layers if they differ by the addition of a single circuit interaction, analogous to a move in the assembly game. Notably, 216/221 oscillators constitute a large connected component of the graph originating from the four minimal motifs, indicating that almost all 3-node oscillators live in a subset of design space defined by Rep and/or AI motifs. The size of each node in Figure 3 denotes its motif robustness Q_{motif} , the overall oscillation probability for all circuits containing this motif (in other words, the mean reward once reaching this state in the game). The color of each node reflects its motif discovery rate, measured as the percentage of $n = 50$ replicates that labeled it as a motif.

Note that these features correlate visually (large circles tend to be red and *vice versa*) and quantitatively, as shown by the scatterplot in Figure S5B, indicating that higher quality motifs are more likely to be found. Additionally, the motifs that are discovered most reliably (discovery rate $< 80\%$, indicated by bold edges) form an optimal subset corresponding to the best motifs (the largest nodes). Surprisingly, the most robust oscillator, the repressilator with PAR on all three TFs (Rep+3xPAR, $Q = 0.767$), is discovered at a rate of 98%. Because Rep+2xPAR is a poor oscillator (ranked #187) and Rep+1xPAR does not oscillate at all, Rep+3xPAR cannot be assembled from Rep (or any other intermediate) without breaking oscillations. Nonetheless, its high motif discovery rate suggests that CircuiTree is capable of identifying special design strategies that require a specific combination of moves. Overall, we find that CircuiTree reliably infers the most robust motifs for a given phenotype, even finding obscure but optimal solutions, and the resulting motifs for 3-node oscillation form a single cluster of optimal designs. Note that, in contrast to classic network motifs which compare the frequency of a pattern in successful topologies against the entire set of topologies (or a comparable null model), we compare frequencies between circuit assemblies that were successful during search and

random circuit assemblies. Therefore our method returns motifs that are overrepresented *among search results* and thereby accounts for the bias inherent to sampling the space of topologies by assembly rather than by flat enumeration.

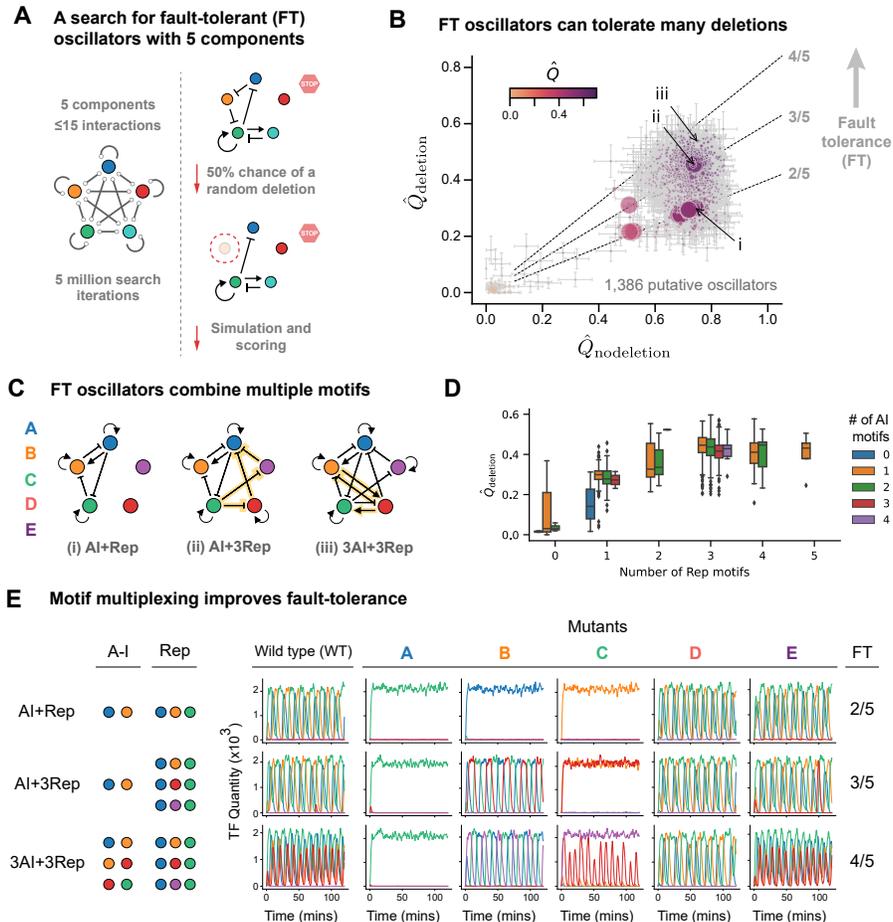


Figure 4: Parallelized CircuiTree scales to large design spaces. (A) A parallelized version of CircuiTree was used to search for five-node oscillators with ≤ 15 interactions (left) that oscillate despite a 50% chance of a single random deletion (right). (B) Search results after $5 \cdot 10^6$ iterations. Circles are putative oscillators ($\hat{Q} > 0.01$, $v_i > 100$), plotted on axes of robustness based on samples with ($\hat{Q}_{\text{deletion}}$) or without a deletion ($\hat{Q}_{\text{nodeletion}}$). Error bars indicate standard error of the mean. Circle size and color indicate the number of samples v_i and the overall robustness \hat{Q} , respectively. Dashed lines show different theoretical values of fault-tolerance (FT), or the average number of deletions a circuit can sustain. Examples of circuits with different FT (labeled i-iii) are shown in (C). (D) Box plots of $\hat{Q}_{\text{deletion}}$, grouped by the number of motifs in each putative oscillator. Multiple motifs, particularly Rep, increase robustness to deletions. (E) Simulated trajectories for the topologies in (C) for all single deletions. The 3AI+3Rep oscillator (FT $\approx 4/5$) contains 6 interleaved oscillator motifs. See also Figures S6 and S7.

Motif multiplexing allows five-node oscillators to compensate for deletions and single-gene knockdowns

While synthetic circuits generally rely on a single, minimal circuit module, naturally occurring circuits often use many functional modules. For instance, circadian oscillators across divergent taxa contain two or more oscillatory feedback loops (K. Lee, Loros, and Dunlap, 2000; Cheng, Y. Yang, and Liu, 2001; Bell-Pedersen et al., 2005; Pokhilko et al., 2012). Could these additional modules make circadian oscillators more resistant to deletions during evolution (Wagner, 2005)? To explore this possibility, we implemented a parallelized version of CircuiTree with pruning (described in details in the Methods and illustrated in Figure S6) and used it to design 5-node circuit topologies that oscillate despite a 50% chance of deletion of a random TF (Figure 4A). Please see Methods for a description of the parallel implementation and Figure S6 for a schematic of the pruning.

After 5 million search iterations (less than four days of real time using 1,000 parallel search threads and 300 CPUs), CircuiTree finds 1,386 putative oscillators with $\hat{Q}_i > 0.01$ and > 100 samples, the first result arriving in just 1.3 hours. Due to the computational cost of exact simulations of large stochastic systems, only topologies with up to 15 interactions were included. As shown by the heatmap in Figure S7A, when clustered based on their structural similarity (measured as the pairwise graph edit distance between topologies), these topologies separate into sparse and dense clusters (Figure S7B). In Figure 4B, each of these topologies is plotted as a circle based on their observed robustness with or without a random deletion (\hat{Q}_{del} and \hat{Q}_{node1} , respectively). The overall robustness \hat{Q} is indicated by the color gradient, and the circle size represents the number of samples. The circuit's fault-tolerance (FT), the proportion of components that can be deleted without losing oscillation, is estimated as $\hat{Q}_{\text{deletion}}/\hat{Q}_{\text{node1}}$, and contour lines for $\text{FT} = \frac{2}{5}$, $\frac{3}{5}$, and $\frac{4}{5}$ are shown as dashed lines. The putative oscillators included 42 topologies with 2 or 3 nodes, all of which had low robustness to deletions ($\hat{Q}_{\text{deletion}} < 0.32$). For instance, the second-best oscillator in the 3-node search, the AI+Rep circuit (labeled (i) in Figure 4B and Figure 4C), was found to have a fault tolerance of $0.404 \pm 0.009 \approx \frac{2}{5}$, consistent with oscillations that persist after deletion of the two unused TFs (D and E) but attenuating with deletion of any of the active components (A, B, or C).

In contrast, fault-tolerant oscillators ($\hat{Q}_{\text{deletion}} \gtrsim 0.4$) contain many interleaved oscillatory motifs that activate under different deletion scenarios, a design feature we call motif multiplexing. As shown in Figure 4D, the number of repressilator

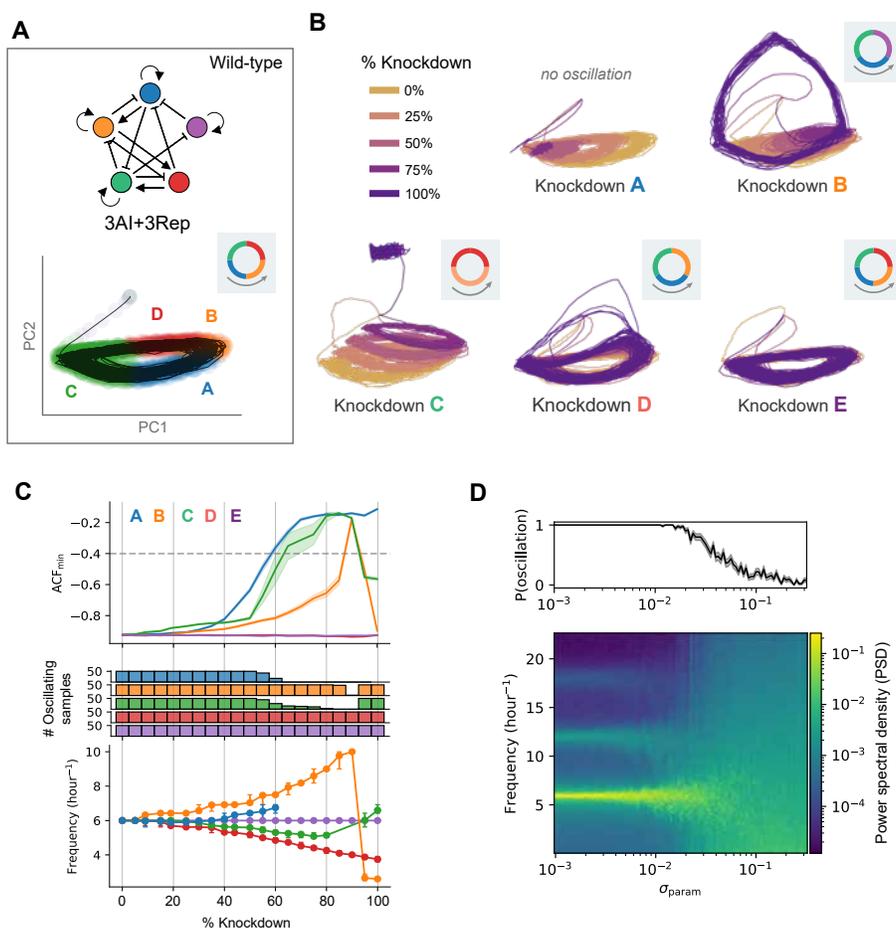


Figure 5: Motif multiplexing makes oscillators resistant to the failure of components. The 3AI+3Rep circuit (A, top) oscillates with different limit cycles after partial knockdowns of different genes. (A, bottom) An exemplary trajectory of 3AI+3Rep is shown on axes of the first two principal components (PCs) of phase space. Transparent circles indicate the dominant species at each time-point. (B) Trajectories under scenarios where transcription rate is reduced by a factor KD. The ordering of species in the limit cycle at KD = 100% is shown by the inset diagram. (C) Oscillation quality and frequency in single-gene knockdowns. Oscillations persist ($ACF_{\min} < ACF_{\text{thresh}}$) for most knockdowns of genes B, C, D, and E (middle). Oscillation frequency (bottom) is pulled from its WT value in knockdowns of A, B, C, and D. (E) Robustness to parameter variation between TFs. The power spectral density of the trajectory of TF A (bottom, mean) and the overall oscillation rate (top, mean \pm SEM) are shown for simulations in which parameters were perturbed by a Gaussian kernel of width σ_{param} ($n = 50$ replicates). A dissipation of fundamental and harmonic frequencies and corresponding loss of oscillations occurs for $\sigma_{\text{param}} > 5 \cdot 10^{-2}$.

(Rep) and activator-inhibitor (AI) motifs is a strong predictor of robustness in the presence of deletion, measured as $\hat{Q}_{\text{deletion}}$, and higher complexity in general is associated with higher \hat{Q} (Figure S7C). To see this design principle, consider the AI+3Rep circuit (labeled (ii) in Figures 4B and 4C), which is an extension of the

AI+Rep topology with two additional backup repressilator loops (A-C-D and A-C-E, highlighted in yellow). This circuit has a higher fault tolerance of $0.617 \pm 0.011 \approx \frac{3}{5}$ because, while deletion of gene B is fatal for oscillations in the AI+Rep circuit, the A-C-D motif takes over to rescue oscillations in the AI+3Rep circuit (representative simulated trajectories shown in Figure 4E, upper and middle row; see Table 2.1 for parameter values). The pattern extends to the 3AI+3PAR circuit (labeled (iii) in Figures 4B and 4C, right diagram) which is similar to the AI+3Rep except for the addition of two activator-inhibitor motifs (B-D and D-C, highlighted in yellow). This circuit similarly activates a repressilator motif (A-C-E) upon deletion of gene B. Now, however, oscillations are rescued after deletion of gene C by activating the B-D motif. Consequently, the 3AI+3PAR circuit has a high fault tolerance of $0.715 \pm 0.074 \approx \frac{4}{5}$ (Figure 4E, bottom row).

During evolution, genomic mutation may lead to a partial reduction in transcription rate rather than a complete knockout. Do multiplexed oscillators maintain their mutational robustness in these conditions? To explore this question, we simulated the 3AI+3Rep circuit under conditions where a single gene is partially knocked down by multiplying its transcription rate by a factor $(100 - \text{KD})/100$, KD being the percent of knockdown. In Figure 5A, we visualize a representative stochastic simulation of the wild-type (WT) 3AI+3Rep circuit by reducing the system of five TFs (a five-dimensional phase space) to two composite dimensions using principal components analysis and plotting a simulated trajectory on the first and second principal component axes. At each time point, the dominantly expressed TF is indicated by a transparent marker of the same color, and the inset diagram shows the order of TFs activated in the limit cycle (A-B-D-C). As gene A is knocked down from KD=0% to KD=100% (Figure 5B, upper middle panel), the limit cycle gradual drifts in phase space until the trajectory eventually flattens, consistent with the lack of oscillations after deletion of gene A (Figure 4E). During knockdown of gene B (Figure 5B, top right panel), in contrast, the limit cycle drifts before discontinuously jumping to a new limit cycle (the A-E-C repressilator motif) between 75% and 100% KD, as shown in the inset diagram. Similarly, knockdown of gene C induces a gradual drift followed by a jump between 75% and 100% KD to a limit cycle driven by the B-D activator-inhibitor motif. Unlike genes B and C, knockdown of gene D produces no obvious discontinuities as the limit cycle gradually adapts from A-B-D-C to the A-B-C repressilator. Knockdown of gene E produces no appreciable changes in the limit cycle. Overall, the multiplexed oscillator 3AI+3Rep maintains oscillatory behavior during partial knockdown of genes B, C, D, or E and transitions

between disparate limit cycles in qualitatively unique ways.

To quantitatively understand how the system responds to knockdowns, single-gene knockdowns were simulated with a range of values of KD with $n = 50$ replicates with different random seeds. The quality and frequency of oscillations were then assessed by computing the ACF_{\min} and, if oscillations were detected, the frequency of oscillation. In Figure 5C (upper panel), the ACF_{\min} (mean \pm 95% confidence interval) is plotted as a function of KD, and the dashed line indicates the value $ACF_{\min} = -0.4$ used as a threshold between oscillatory and non-oscillatory dynamics (below and above the line, respectively). Knockdown of gene A leads to a rise in ACF_{\min} and becomes lethal for oscillations above $KD = 60\%$. For gene B, ACF_{\min} increases before suddenly crossing the threshold to indicate loss of oscillations between $KD = 85\%$ and 95% , and for gene C, oscillations disappear in the range $KD = 65\%$ to 90% . For genes D and E, no detrimental effects are observed during knockdown. Thus, oscillations persist in almost all cases (91.9% of samples) when partially knocking down genes B, C, D, or E (Figure 5C, middle panel). Partial knockdowns of A, B, C, and D each have distinct effects on oscillation frequency, a phenomenon called frequency pulling (Heltberg et al., 2021). The bottom panel of Figure 5C shows how, for $KD < 60\%$, the WT resonant frequency of 6.0hr^{-1} is pulled higher when knocking down A or B and lower when knocking down C or D (no change for E). At higher values of KD, knockdowns of B and C discontinuously jump from the drifted frequency to a new resonant frequency (2.6hr^{-1} and 6.6hr^{-1} , respectively), while the D knockdown transitions smoothly to its new frequency of 3.8hr^{-1} . Thus, the 3AI+3Rep oscillator accommodates most partial knockdowns of genes B, C, D, or E by modulating, yet retaining, oscillatory dynamics.

Up to this point, our modeling has assumed that the rates of reactions such as binding, transcription, translation, and degradation are identical between circuit components. To investigate whether the oscillators we discover are fine-tuned for this symmetry, we perturbed the rate parameters for each TF individually. A Gaussian perturbation kernel with standard deviation σ_{param} was applied to each parameter, scaled to the pre-defined ranges of those parameters (see Table 2.2 for parameter ranges and Methods for details of the perturbation kernel). Stochastic simulations were performed for a range of values of σ_{param} between $1 \cdot 10^{-3}$ and $3 \cdot 10^{-2}$. To assess the effects of parameter perturbation on oscillation, the power spectral density (PSD) of the dynamics of TF A (mean of $n = 50$ replicates with different random seeds) was calculated for each value of σ_{param} and plotted as a heatmap (Figure 5D). For

$\sigma_{\text{param}} < 10^{-2}$, the PSD shows peaks at the oscillator’s fundamental frequency of 6hr^{-1} and first and second harmonics (12hr^{-1} and 18hr^{-1} , respectively). As σ_{param} increases further to $5 \cdot 10^{-2}$, the fundamental frequency peak gradually diffuses, and the oscillation rate drops from 1.0 to ~ 0.5 , as shown in the line plot in Figure 5D (envelope indicates 95% confidence interval). Above $\sigma_{\text{param}} = 5 \cdot 10^{-2}$, the oscillation rate drops to near-zero, and the mean PSD shows no visible peaks. Therefore, this oscillator can tolerate a minimal amount of heterogeneity between TFs, above which it appears somewhat sensitive to asymmetry in at least one parameter of the model.

Discussion

Natural biological networks contain a number of biochemical components that exceeds our current capabilities of engineering by orders of magnitude, underscoring the importance of scalable computational methods for synthetic circuit design and analysis of biological design principles. Inspired by state-of-the-art artificial intelligence, we approach circuit design as a game of step-by-step topology assembly where success is determined by the achievement of a target phenotype in simulation. Similar to game-playing platforms that have achieved superhuman mastery of complex games (D. Silver, Schrittwieser, et al., 2017), CircuiTree searches the space of possible circuit topologies using Monte Carlo tree search (MCTS), which balances exploitation of promising circuit assembly moves with exploration of other possibilities (Figure 1). This search strategy is comprehensive enough to infer motifs for a given phenotype (Figure 3) and efficient enough to search large spaces of designs fruitfully with limited samples (Figures 4A and 4B). Finally, we demonstrate CircuiTree’s scalability by characterizing a novel class of five-gene oscillators that use a strategy we call motif multiplexing to resist deletion and single-gene knockdown mutations by densely interleaving multiple sub-oscillators (Figures 4C-4E and 5A-5C). This design principle may grant evolutionary robustness to eukaryotic circadian clocks, which have been observed across many phyla to contain multiple oscillatory loops (Cheng, Y. Yang, and Liu, 2001; Bell-Pedersen et al., 2005; Pokhilko et al., 2012), and it leaves open the question of how multiple sub-oscillators can be coupled without triggering chaos (Heltberg et al., 2021).

CircuiTree is distinguished by its general framework. Developed for general game-playing and planning problems, MCTS can query very large spaces in search of robust topologies and assembly motifs for any measurable phenotype, without restrictions on the modeling or simulation framework, and without needing to enumerate all possible topologies. This property bridges a gap in computational circuit

design where generally, methods for finding single topologies (such as evolution (François and Hakim, 2004; François and Eric D. Siggia, 2008), mixed-integer optimization (Otero-Muras and Banga, 2016), and recurrent neural networks (Shen et al., 2021)) do not generalize easily to design principles and/or require a specific mathematical form, while more comprehensive methods (such as enumeration (Chau et al., 2012; Schaerli et al., 2014) and Bayesian sampling (Woods et al., 2016)) require an explicit list of topologies. Notably, while CircuiTree currently implements a “vanilla” version of the algorithm, MCTS can be modified in many ways to suit different problems (Browne et al., 2012; Świechowski et al., 2022).

This work also highlights the possibilities of using reinforcement learning to address difficult combinatorial design and inference problems in biology. In future work, CircuiTree could be applied to combinatorial design of therapeutics, both in the sense of optimizing chimeric receptor design (Daniels et al., 2022) or combinatorial antigen recognition (Dannenfelser et al., 2020; Williams et al., 2020), and in the sense of generating novel multicellular therapies consisting of multiple engineered cell types that collaborate *in situ*, analogous to a natural immune system. Our work could also be used to generate synthetic morphogenesis circuits, which are difficult to design due to the many possible combinations of chemical and physical components (Davies, 2017; Toda et al., 2018; Fleischer and Barr, 1997) and a high computational cost per simulation. Given a reward function that measures goodness-of-fit, CircuiTree could also be extended as an inference tool — for instance, for inferring transcriptional regulatory networks from data. More generally, this computational platform extends the study of design principles (Lim, C. M. Lee, and Tang, 2013), which has been limited to small motifs of 2 or 3 components, to a larger space of biological networks, presenting an opportunity to dissect the algorithms and strategies used by biology to assemble complex networks at scale.

Methods

Modeling and simulation

Transcription factor (TF) cooperativity was modeled as a slower unbinding rate when both response elements (REs) for the a promoter are bound by the same TF. For K genes (each with mRNA and protein species) with cumulative A activation reactions and R repression reactions, the elimination of explicit dimerization reduces the number of species from $2K + K^2 + K^3$ (mRNAs, TFs, TF-TF complexes, and TF-TF-RE complexes) to $2K + A + R$ (mRNAs, TFs, and TF-RE complexes) and the number of reactions accordingly from $4K + 2K^2 + 3K^3$ to $4K + 3A + 3R$.

Stochastic trajectories were simulated using Gillespie's exact method (Gillespie, 1977) and saved at $n_t = 2000$ time points at intervals of $dt = 20$ sec. The separation of time-scales between fast binding-unbinding kinetics and the other reactions creates a long simulation time (>2 mins) that cannot be alleviated with, for example, traditional τ -leaping (Cao, Gillespie, and Petzold, 2004). Thus, instead of using realistic but impractical binding and unbinding rates, these rates were set to virtual values determined from the equilibrium constant for first-binding $K_{D,1}$ by the solving the equations

$$K_{D,1} = k_{\text{off},1}/k_{\text{on}}$$

$$(k_{\text{off},1} \cdot 1 \text{ molec} + k_{\text{on}})(1 \text{ sec}) = Z,$$

where the value $Z = 100$ was chosen heuristically to be large enough to maintain a separation of time-scales at low quantities. For each protein species, high-frequency noise was filtered from the stochastic signal using a 9-point binomial filter (Aubury and Luk, 1996), and the quality of oscillations overall was determined by computing the normalized autocorrelation function for each TF and finding the lowest minimum among TFs (ACF_{min}), excluding the bounds. Oscillations are considered present if this quantity, related to the dissipation constant for oscillations (Otero-Muras and Banga, 2016), is below a cutoff $\text{ACF}_{\text{thresh}} = -0.4$.

For simulations with partial knockdown of a gene, all transcriptional rate parameters for that gene were multiplied by a coefficient on $[0, 1]$ (for example, $\text{KD}=80\%$ was achieved using a coefficient of 0.2.), and all species were initialized with zero quantity. For all other simulations, the initial quantity of each TF was selected from a Poisson distribution with a mean of 10 proteins, and all other species were initialized with zero quantity.

For perturbation studies, the eight sampled dimensionless variables were converted to values on $[0, 1]$ by normalization to the upper and lower values in Table 2.2. These values were stored in a $K \times 8$ matrix, where each row represented one of the $K = 5$ TFs, and each value was perturbed independently with a Gaussian kernel of standard deviation σ_{param} , truncated to the range $[0, 1]$ to prevent values outside the reasonable range. The perturbed values were then converted to rate parameters for each TF as outlined above.

Random sampling

Random sets of the 10 rate parameters listed in Table 2.1 were generated by drawing samples of the 8 dimensionless variables described in Table 2.2. Latin Hypercube

sampling was used to draw 10^4 samples from a multivariate uniform distribution, with bounds shown in Table 2.2. For the 3-node and 5-node cases, every possible initialization (with a unique set of random generator seed, initial protein quantity, and parameters) was stored in a table of 10^4 rows, and each simulation was initialized with a random row of this table.

Monte Carlo tree search

For all MCTS runs, the hyperparameter in Equation 2.1 was set to $c = 2.00$ to encourage exploration (the default value is $c = \sqrt{2}$). Replicate runs for an experiment were performed using different random seeds, which are used to break ties during the selection phase and draw a random parameter set during the simulation step. For the 3-node search, each reward value q was drawn as a Bernoulli trial with a success probability equal to the Q value found using enumeration rather than running fresh simulations.

MCTS was parallelized using the lock-free method Enzenberger and Müller, 2010 and implemented with the Python utility ‘celery’ (<https://docs.celeryq.dev/en/stable/index.html>). Tree search was parallelized over 1,000 green threads which dispatched simulation jobs to run in parallel on multiple cloud computing instances totaling 300 CPUs. Reward values for each topology-parameter set pair were stored in an in-memory cache to speed up subsequent training runs (epochs). Backpropagation was executed asynchronously with virtual loss, assuming a reward of 0 until the actual reward is returned. To prevent excessive sampling of local optima during the 5-node search, we introduced a form of decision tree pruning we term "node exhaustion." Once a terminal node (a completed topology) is visited $> 10^4$ times, it is considered exhaustively sampled and pruned from the search graph, and a non-terminal node is pruned once all its successors have been pruned. As illustrated in Figure S5, the visits and reward for each in-edge to an exhausted node are subtracted from the parent node. Thus, once a node is marked exhausted, its history is "forgotten" by its predecessors.

Please see the documentation at <https://pranav-bhamidipati.github.io/circuitree/index.html> for all additional details of implementation, as well as code tutorials and descriptions of the API.

Motif identification

Before testing for motifs based on a tree search, we first determine whether each terminated topology s_j discovered during the search is a successful oscillator by

comparing its empirical robustness $\hat{Q}(s_j) = r_j/v_j$ to the predefined threshold $Q_{\text{thresh}} = 0.01$. This segregates all topologies into disjoint successful and unsuccessful sets (X and Y , respectively). We then take samples from the null distribution, drawing $n_{\text{sample}} = 10^5$ samples from the tree of topologies using random assembly, each time starting at the root of the design tree and choosing random actions until a terminal topology is reached (Figure S5A). The procedure is then repeated, this time rejecting the result unless it is a member of the successful set X (note that this can be computationally expensive if solutions are sparse). Next, a contingency table is constructed for each successful oscillator $x_i \in X$ to compare the frequency of observing x_i within samples of the successful subspace against the frequency of observation in the overall space. (see Figure S5A for an illustration of this process). Finally, statistically significant overrepresentation is determined by the χ^2 independence test with a significance threshold of $p < \alpha = 0.05$ after Bonferroni correction. To identify motifs based on the results of enumeration, we conduct the same hypothesis test on each oscillator, except using ground-truth frequencies found by enumeration.

Supplementary information

Code availability

CircuiTree, written in Python 3.10, is available on GitHub (<https://doi.org/10.5281/zenodo.11285522>). The code used to run computational experiments, perform analyses, and plot results is available separately at <https://doi.org/10.5281/zenodo.11285550>.

References

- Alon, Uri (June 2007). “Network motifs: theory and experimental approaches”. In: *Nature Reviews Genetics* 8.6. Number: 6 Publisher: Nature Publishing Group, pp. 450–461. ISSN: 1471-0064. DOI: [10.1038/nrg2102](https://doi.org/10.1038/nrg2102). URL: <https://www.nature.com/articles/nrg2102> (visited on 09/15/2023).
- (Aug. 4, 2019). *An Introduction to Systems Biology: Design Principles of Biological Circuits*. 2nd ed. Boca Raton: Chapman and Hall/CRC. 342 pp. ISBN: 978-0-429-28332-1. DOI: [10.1201/9780429283321](https://doi.org/10.1201/9780429283321).
- Aubury, Matthew and Wayne Luk (Jan. 1, 1996). “Binomial filters”. In: *Journal of VLSI signal processing systems for signal, image and video technology* 12.1, pp. 35–50. ISSN: 0922-5773. DOI: [10.1007/BF00936945](https://doi.org/10.1007/BF00936945). URL: <https://doi.org/10.1007/BF00936945> (visited on 03/04/2024).

- Auer, Peter, Nicolò Cesa-Bianchi, and Paul Fischer (May 1, 2002). “Finite-time Analysis of the Multiarmed Bandit Problem”. In: *Machine Learning* 47.2, pp. 235–256. ISSN: 1573-0565. DOI: [10.1023/A:1013689704352](https://doi.org/10.1023/A:1013689704352). URL: <https://doi.org/10.1023/A:1013689704352> (visited on 03/10/2023).
- Bell-Pedersen, Deborah et al. (July 2005). “Circadian rhythms from multiple oscillators: lessons from diverse organisms”. In: *Nature Reviews Genetics* 6.7. Number: 7 Publisher: Nature Publishing Group, pp. 544–556. ISSN: 1471-0064. DOI: [10.1038/nrg1633](https://www.nature.com/articles/nrg1633). URL: <https://www.nature.com/articles/nrg1633> (visited on 01/30/2024).
- Browne, Cameron B. et al. (Mar. 2012). “A Survey of Monte Carlo Tree Search Methods”. In: *IEEE Transactions on Computational Intelligence and AI in Games* 4.1. Conference Name: IEEE Transactions on Computational Intelligence and AI in Games, pp. 1–43. ISSN: 1943-0698. DOI: [10.1109/TCIAIG.2012.2186810](https://doi.org/10.1109/TCIAIG.2012.2186810).
- Cao, Yang, Daniel T. Gillespie, and Linda R. Petzold (Dec. 15, 2004). “The slow-scale stochastic simulation algorithm”. In: *The Journal of Chemical Physics* 122.1, p. 014116. ISSN: 0021-9606. DOI: [10.1063/1.1824902](https://doi.org/10.1063/1.1824902). URL: <https://doi.org/10.1063/1.1824902> (visited on 03/04/2024).
- Chandra, Fiona A., Gentian Buzi, and John C. Doyle (July 8, 2011). “Glycolytic Oscillations and Limits on Robust Efficiency”. In: *Science* 333.6039. Publisher: American Association for the Advancement of Science, pp. 187–192. DOI: [10.1126/science.1200705](https://www.science.org/doi/10.1126/science.1200705). URL: <https://www.science.org/doi/10.1126/science.1200705> (visited on 03/03/2024).
- Chau, Angela H. et al. (Oct. 12, 2012). “Designing Synthetic Regulatory Networks Capable of Self-Organizing Cell Polarization”. In: *Cell* 151.2, pp. 320–332. ISSN: 0092-8674. DOI: [10.1016/j.cell.2012.08.040](https://www.sciencedirect.com/science/article/pii/S009286741201121X). URL: <https://www.sciencedirect.com/science/article/pii/S009286741201121X> (visited on 01/31/2023).
- Cheng, Ping, Yuhong Yang, and Yi Liu (June 19, 2001). “Interlocked feedback loops contribute to the robustness of the *Neurospora* circadian clock”. In: *Proceedings of the National Academy of Sciences* 98.13. Publisher: Proceedings of the National Academy of Sciences, pp. 7408–7413. DOI: [10.1073/pnas.121170298](https://www.pnas.org/doi/10.1073/pnas.121170298). URL: <https://www.pnas.org/doi/10.1073/pnas.121170298> (visited on 01/30/2024).
- Cotterell, James and James Sharpe (Jan. 2010). “An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients”. In: *Molecular Systems Biology* 6.1. Publisher: John Wiley & Sons, Ltd, p. 425. ISSN: 1744-4292. DOI: [10.1038/msb.2010.74](https://www.embopress.org/doi/full/10.1038/msb.2010.74). URL: <https://www.embopress.org/doi/full/10.1038/msb.2010.74> (visited on 02/08/2023).
- Daniels, Kyle G. et al. (Dec. 16, 2022). “Decoding CAR T cell phenotype using combinatorial signaling motif libraries and machine learning”. In: *Science* 378.6625. Publisher: American Association for the Advancement of Science, pp. 1194–

1200. DOI: [10.1126/science.abq0225](https://doi.org/10.1126/science.abq0225). URL: <https://www.science.org/doi/10.1126/science.abq0225> (visited on 03/04/2024).
- Dannenfelser, Ruth et al. (Sept. 23, 2020). “Discriminatory Power of Combinatorial Antigen Recognition in Cancer T Cell Therapies”. In: *Cell Systems* 11.3, 215–228.e5. ISSN: 2405-4720. DOI: [10.1016/j.cels.2020.08.002](https://doi.org/10.1016/j.cels.2020.08.002).
- Davies, Jamie (Apr. 1, 2017). “Using synthetic biology to explore principles of development”. In: *Development* 144.7, pp. 1146–1158. ISSN: 0950-1991. DOI: [10.1242/dev.144196](https://doi.org/10.1242/dev.144196). URL: <https://doi.org/10.1242/dev.144196> (visited on 05/31/2022).
- Elowitz, Michael B. and Stanislas Leibler (Jan. 2000). “A synthetic oscillatory network of transcriptional regulators”. In: *Nature* 403.6767. Number: 6767 Publisher: Nature Publishing Group, pp. 335–338. ISSN: 1476-4687. DOI: [10.1038/35002125](https://www.nature.com/articles/35002125). URL: <https://www.nature.com/articles/35002125> (visited on 10/17/2023).
- Enzenberger, Markus and Martin Müller (2010). “A Lock-Free Multithreaded Monte-Carlo Tree Search Algorithm”. In: *Advances in Computer Games*. Ed. by H. Jaap van den Herik and Pieter Spronck. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 14–20. ISBN: 978-3-642-12993-3. DOI: [10.1007/978-3-642-12993-3_2](https://doi.org/10.1007/978-3-642-12993-3_2).
- Ferrell, James E., Tony Yu-Chen Tsai, and Qiong Yang (Mar. 18, 2011). “Modeling the Cell Cycle: Why Do Certain Circuits Oscillate?” In: *Cell* 144.6. Publisher: Elsevier, pp. 874–885. ISSN: 0092-8674, 1097-4172. DOI: [10.1016/j.cell.2011.03.006](https://www.cell.com/cell/abstract/S0092-8674(11)00243-1). URL: [https://www.cell.com/cell/abstract/S0092-8674\(11\)00243-1](https://www.cell.com/cell/abstract/S0092-8674(11)00243-1) (visited on 03/04/2024).
- Fleischer, Kurt and Alan Barr (Feb. 16, 1997). “A Simulation Testbed for the Study of Multicellular Development: The Multiple Mechanisms of Morphogenesis”. In: *Artificial Life III*.
- François, Paul and Vincent Hakim (Jan. 13, 2004). “Design of genetic networks with specified functions by evolution in silico”. In: *Proceedings of the National Academy of Sciences* 101.2. Publisher: Proceedings of the National Academy of Sciences, pp. 580–585. DOI: [10.1073/pnas.0304532101](https://www.pnas.org/doi/full/10.1073/pnas.0304532101). URL: <https://www.pnas.org/doi/full/10.1073/pnas.0304532101> (visited on 02/07/2023).
- François, Paul, Vincent Hakim, and Eric D Siggia (Jan. 2007). “Deriving structure from evolution: metazoan segmentation”. In: *Molecular Systems Biology* 3.1. Publisher: John Wiley & Sons, Ltd, p. 154. ISSN: 1744-4292. DOI: [10.1038/msb4100192](https://www.embopress.org/doi/full/10.1038/msb4100192). URL: <https://www.embopress.org/doi/full/10.1038/msb4100192> (visited on 02/07/2023).
- François, Paul and Eric D. Siggia (June 2008). “A case study of evolutionary computation of biochemical adaptation”. In: *Physical Biology* 5.2, p. 026009.

- ISSN: 1478-3975. DOI: [10.1088/1478-3975/5/2/026009](https://doi.org/10.1088/1478-3975/5/2/026009). URL: <https://dx.doi.org/10.1088/1478-3975/5/2/026009> (visited on 02/07/2023).
- François, Paul and Eric D. Siggia (July 15, 2010). “Predicting embryonic patterning using mutual entropy fitness and in silico evolution”. In: *Development* 137.14, pp. 2385–2395. ISSN: 0950-1991. DOI: [10.1242/dev.048033](https://doi.org/10.1242/dev.048033). URL: <https://doi.org/10.1242/dev.048033> (visited on 02/07/2023).
- Gardner, Timothy S., Charles R. Cantor, and James J. Collins (Jan. 20, 2000). “Construction of a genetic toggle switch in *Escherichia coli*”. In: *Nature* 403.6767. MAG ID: 2134759932, pp. 339–342. DOI: [10.1038/35002131](https://doi.org/10.1038/35002131).
- Gerardin, Jaline, Nishith R. Reddy, and Wendell A. Lim (Sept. 25, 2019). “The Design Principles of Biochemical Timers: Circuits that Discriminate between Transient and Sustained Stimulation”. In: *Cell Systems* 9.3, 297–308.e2. ISSN: 2405-4712. DOI: [10.1016/j.cels.2019.07.008](https://www.sciencedirect.com/science/article/pii/S2405471219302650). URL: <https://www.sciencedirect.com/science/article/pii/S2405471219302650> (visited on 08/31/2023).
- Geva-Zatorsky, Naama, Erez Dekel, et al. (July 27, 2010). “Fourier analysis and systems identification of the p53 feedback loop”. In: *Proceedings of the National Academy of Sciences* 107.30. Publisher: Proceedings of the National Academy of Sciences, pp. 13550–13555. DOI: [10.1073/pnas.1001107107](https://www.pnas.org/doi/10.1073/pnas.1001107107). URL: <https://www.pnas.org/doi/10.1073/pnas.1001107107> (visited on 09/20/2023).
- Geva-Zatorsky, Naama, Nitzan Rosenfeld, et al. (June 13, 2006). “Oscillations and variability in the p53 system”. In: *Molecular Systems Biology* 2, p. 2006.0033. ISSN: 1744-4292. DOI: [10.1038/msb4100068](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1681500/). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1681500/> (visited on 09/02/2021).
- Gillespie, Daniel T. (Dec. 1, 1977). “Exact stochastic simulation of coupled chemical reactions”. In: *The Journal of Physical Chemistry* 81.25. Publisher: American Chemical Society, pp. 2340–2361. ISSN: 0022-3654. DOI: [10.1021/j100540a008](https://doi.org/10.1021/j100540a008). URL: <https://doi.org/10.1021/j100540a008> (visited on 03/04/2024).
- Heltberg, Mathias L. et al. (Apr. 21, 2021). “A tale of two rhythms: Locked clocks and chaos in biology”. In: *Cell Systems* 12.4. Publisher: Elsevier, pp. 291–303. ISSN: 2405-4712. DOI: [10.1016/j.cels.2021.03.003](https://www.cell.com/cell-systems/abstract/S2405-4712(21)00108-3). URL: [https://www.cell.com/cell-systems/abstract/S2405-4712\(21\)00108-3](https://www.cell.com/cell-systems/abstract/S2405-4712(21)00108-3) (visited on 01/24/2024).
- Jiménez, Alba et al. (2017). “A spectrum of modularity in multi-functional gene circuits”. In: *Molecular Systems Biology* 13.4, p. 925. ISSN: 1744-4292. DOI: [10.15252/msb.20167347](https://doi.org/10.15252/msb.20167347).
- Jones, Ethan M., John P. Marken, and Pamela A. Silver (Jan. 22, 2024). “Synthetic microbiology in sustainability applications”. In: *Nature Reviews Microbiology*. Publisher: Nature Publishing Group, pp. 1–15. ISSN: 1740-1534. DOI: [10.1038/s41579-023-01007-9](https://www.nature.com/articles/s41579-023-01007-9). URL: <https://www.nature.com/articles/s41579-023-01007-9> (visited on 03/03/2024).

- Kocsis, Levente and Csaba Szepesvári (2006). “Bandit Based Monte-Carlo Planning”. In: *Machine Learning: ECML 2006*. Ed. by Johannes Fürnkranz, Tobias Scheffer, and Myra Spiliopoulou. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 282–293. ISBN: 978-3-540-46056-5. DOI: 10.1007/11871842_29.
- Lee, Kwangwon, Jennifer J. Loros, and Jay C. Dunlap (July 7, 2000). “Interconnected Feedback Loops in the Neurospora Circadian System”. In: *Science* 289.5476. Publisher: American Association for the Advancement of Science, pp. 107–110. DOI: 10.1126/science.289.5476.107. URL: <https://www.science.org/doi/full/10.1126/science.289.5476.107> (visited on 01/30/2024).
- Lim, Wendell A., Connie M. Lee, and Chao Tang (Jan. 24, 2013). “Design Principles of Regulatory Networks: Searching for the Molecular Algorithms of the Cell”. In: *Molecular Cell* 49.2. Publisher: Elsevier, pp. 202–212. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2012.12.020. URL: [https://www.cell.com/molecular-cell/abstract/S1097-2765\(13\)00004-X](https://www.cell.com/molecular-cell/abstract/S1097-2765(13)00004-X) (visited on 10/17/2023).
- Ma, Wenzhe et al. (Aug. 21, 2009). “Defining Network Topologies that Can Achieve Biochemical Adaptation”. In: *Cell* 138.4, pp. 760–773. ISSN: 0092-8674. DOI: 10.1016/j.cell.2009.06.013. URL: <https://www.sciencedirect.com/science/article/pii/S0092867409007120> (visited on 02/07/2023).
- Milo, R. et al. (Oct. 25, 2002). “Network Motifs: Simple Building Blocks of Complex Networks”. In: *Science* 298.5594. Publisher: American Association for the Advancement of Science, pp. 824–827. DOI: 10.1126/science.298.5594.824. URL: <https://www.science.org/doi/full/10.1126/science.298.5594.824> (visited on 09/15/2023).
- Milo, Ron et al. (Jan. 1, 2010). “BioNumbers—the database of key numbers in molecular and cell biology”. In: *Nucleic Acids Research* 38 (suppl_1), pp. D750–D753. ISSN: 0305-1048. DOI: 10.1093/nar/gkp889. URL: <https://doi.org/10.1093/nar/gkp889> (visited on 03/01/2024).
- Novák, Béla and John J. Tyson (Dec. 2008). “Design principles of biochemical oscillators”. In: *Nature Reviews Molecular Cell Biology* 9.12. Number: 12 Publisher: Nature Publishing Group, pp. 981–991. ISSN: 1471-0080. DOI: 10.1038/nrm2530. URL: <https://www.nature.com/articles/nrm2530> (visited on 10/17/2023).
- Otero-Muras, Irene and Julio R. Banga (Dec. 15, 2016). “Design Principles of Biological Oscillators through Optimization: Forward and Reverse Analysis”. In: *PLOS ONE* 11.12. Publisher: Public Library of Science, e0166867. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0166867. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0166867> (visited on 10/19/2023).

- Pokhilko, Alexandra et al. (Jan. 2012). “The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops”. In: *Molecular Systems Biology* 8.1. Publisher: John Wiley & Sons, Ltd, p. 574. ISSN: 1744-4292. DOI: 10.1038/msb.2012.6. URL: <https://www.embopress.org/doi/full/10.1038/msb.2012.6> (visited on 10/19/2023).
- Purnick, Priscilla E. M. and Ron Weiss (June 2009). “The second wave of synthetic biology: from modules to systems”. In: *Nature Reviews Molecular Cell Biology* 10.6. Publisher: Nature Publishing Group, pp. 410–422. ISSN: 1471-0080. DOI: 10.1038/nrm2698. URL: <https://www.nature.com/articles/nrm2698> (visited on 03/03/2024).
- Qiao, Lingxia, Zhi-Bo Zhang, et al. (Sept. 20, 2022). “Network design principle for robust oscillatory behaviors with respect to biological noise”. In: *eLife* 11. Ed. by Naama Barkai. Publisher: eLife Sciences Publications, Ltd, e76188. ISSN: 2050-084X. DOI: 10.7554/eLife.76188. URL: <https://doi.org/10.7554/eLife.76188> (visited on 10/17/2023).
- Qiao, Lingxia, Wei Zhao, et al. (Sept. 25, 2019). “Network Topologies That Can Achieve Dual Function of Adaptation and Noise Attenuation”. In: *Cell Systems* 9.3, 271–285.e7. ISSN: 2405-4712. DOI: 10.1016/j.cels.2019.08.006. URL: <https://www.sciencedirect.com/science/article/pii/S2405471219302753> (visited on 02/07/2023).
- Schaerli, Yolanda et al. (Sept. 23, 2014). “A unified design space of synthetic stripe-forming networks”. In: *Nature Communications* 5, p. 4905. ISSN: 2041-1723. DOI: 10.1038/ncomms5905.
- Shah, Najaf A. and Casim A. Sarkar (June 23, 2011). “Robust Network Topologies for Generating Switch-Like Cellular Responses”. In: *PLOS Computational Biology* 7.6. Publisher: Public Library of Science, e1002085. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002085. URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002085> (visited on 01/24/2024).
- Shannon, Claude (Mar. 1950). “Programming a Computer for Playing Chess”. In: *Philosophical Magazine*. 7th ser. 41.314. URL: <https://www.computerhistory.org/chess/doc-431614f453dde/> (visited on 09/18/2023).
- Shen, Jingxiang et al. (May 25, 2021). “Finding gene network topologies for given biological function with recurrent neural network”. In: *Nature Communications* 12.1. Number: 1 Publisher: Nature Publishing Group, p. 3125. ISSN: 2041-1723. DOI: 10.1038/s41467-021-23420-5. URL: <https://www.nature.com/articles/s41467-021-23420-5> (visited on 02/07/2023).
- Silver, David, Aja Huang, et al. (Jan. 2016). “Mastering the game of Go with deep neural networks and tree search”. In: *Nature* 529.7587. Number: 7587 Publisher: Nature Publishing Group, pp. 484–489. ISSN: 1476-4687. DOI: 10.1038/

- nature16961. URL: <https://www.nature.com/articles/nature16961> (visited on 05/01/2023).
- Silver, David, Thomas Hubert, et al. (Dec. 7, 2018). “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play”. In: *Science* 362.6419. Publisher: American Association for the Advancement of Science, pp. 1140–1144. DOI: 10.1126/science.aar6404. URL: <https://www.science.org/doi/10.1126/science.aar6404> (visited on 09/18/2023).
- Silver, David, Julian Schrittwieser, et al. (Oct. 2017). “Mastering the game of Go without human knowledge”. In: *Nature* 550.7676. Number: 7676 Publisher: Nature Publishing Group, pp. 354–359. ISSN: 1476-4687. DOI: 10.1038/nature24270. URL: <https://www.nature.com/articles/nature24270> (visited on 02/07/2024).
- Stricker, Jesse et al. (Nov. 2008). “A fast, robust and tunable synthetic gene oscillator”. In: *Nature* 456.7221. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7221 Primary_atype: Research Publisher: Nature Publishing Group, pp. 516–519. ISSN: 1476-4687. DOI: 10.1038/nature07389. URL: <https://www.nature.com/articles/nature07389> (visited on 08/27/2021).
- Świechowski, Maciej et al. (July 19, 2022). “Monte Carlo Tree Search: a review of recent modifications and applications”. In: *Artificial Intelligence Review*. ISSN: 1573-7462. DOI: 10.1007/s10462-022-10228-y. URL: <https://doi.org/10.1007/s10462-022-10228-y> (visited on 02/15/2023).
- Toda, Satoshi et al. (July 13, 2018). “Programming self-organizing multicellular structures with synthetic cell-cell signaling”. In: *Science (New York, N.Y.)* 361.6398, pp. 156–162. ISSN: 1095-9203. DOI: 10.1126/science.aat0271.
- Tromp, John and Gunnar Farneböck (2007). “Combinatorics of Go”. In: *Computers and Games*. Ed. by H. Jaap van den Herik, Paolo Ciancarini, and H. H. L. M. (Jeroen) Donkers. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 84–99. ISBN: 978-3-540-75538-8. DOI: 10.1007/978-3-540-75538-8_8.
- Tyson, John J. et al. (Nov. 1, 1999). “A Simple Model of Circadian Rhythms Based on Dimerization and Proteolysis of PER and TIM”. In: *Biophysical Journal* 77.5, pp. 2411–2417. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(99)77078-5. URL: <https://www.sciencedirect.com/science/article/pii/S0006349599770785> (visited on 03/04/2024).
- Wagner, Andreas (Aug. 16, 2005). “Circuit topology and the evolution of robustness in two-gene circadian oscillators”. In: *Proceedings of the National Academy of Sciences* 102.33. Publisher: Proceedings of the National Academy of Sciences, pp. 11775–11780. DOI: 10.1073/pnas.0501094102. URL: <https://www.pnas.org/doi/full/10.1073/pnas.0501094102> (visited on 10/19/2023).
- Williams, Jasper Z. et al. (Nov. 27, 2020). “Precise T cell recognition programs designed by transcriptionally linking multiple receptors”. In: *Science* 370.6520. Publisher: American Association for the Advancement of Science, pp. 1099–

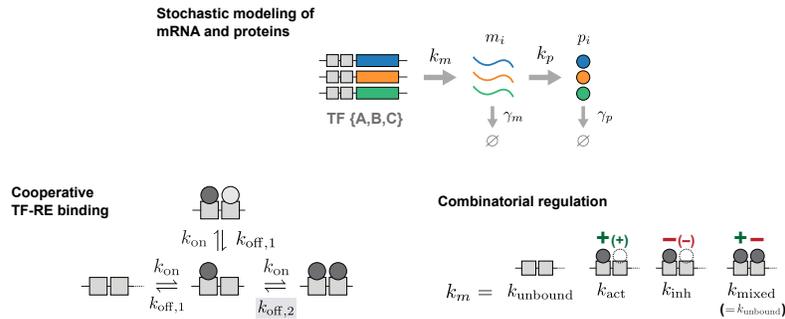
1104. DOI: [10.1126/science.abc6270](https://doi.org/10.1126/science.abc6270). URL: <https://www.science.org/doi/10.1126/science.abc6270> (visited on 03/04/2024).

Woods, Mae L. et al. (June 17, 2016). “A Statistical Approach Reveals Designs for the Most Robust Stochastic Gene Oscillators”. In: *ACS Synthetic Biology* 5.6. Publisher: American Chemical Society, pp. 459–470. DOI: [10.1021/acssynbio.5b00179](https://doi.org/10.1021/acssynbio.5b00179). URL: <https://doi.org/10.1021/acssynbio.5b00179> (visited on 10/13/2023).

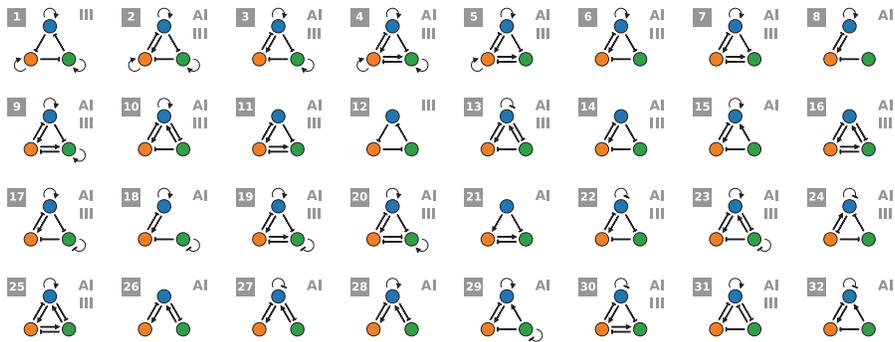
Zhu, Ronghui et al. (Jan. 21, 2022). “Synthetic multistability in mammalian cells”. In: *Science* 375.6578. Publisher: American Association for the Advancement of Science, eabg9765. DOI: [10.1126/science.abg9765](https://doi.org/10.1126/science.abg9765). URL: <https://www.science.org/doi/10.1126/science.abg9765> (visited on 04/14/2023).

Supplementary Figures

A Modeling a network of three transcription factors



B



C

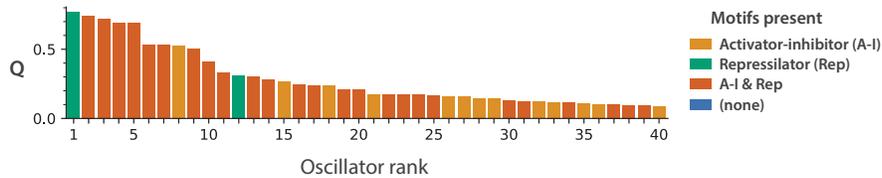


Figure S1: **Modeling and enumeration of 3-component oscillators.** (A) Stochastic modeling of a network of 3 transcription factors. See Table 1 for explanations and default values for reaction rate parameters. (B) The top-32 most robust oscillators (highest Q) based on enumeration and exhaustive simulation (10^4 parameter sets). “AI” and “III” refer to the presence of at least one activator-inhibitor or repressilator motif, respectively. (C) Robustness of the top-40 most robust oscillators. Bars are colored based on the presence or absence of an AI and/or repressilator (Rep) motif. See also Figure 2 and Table 1.

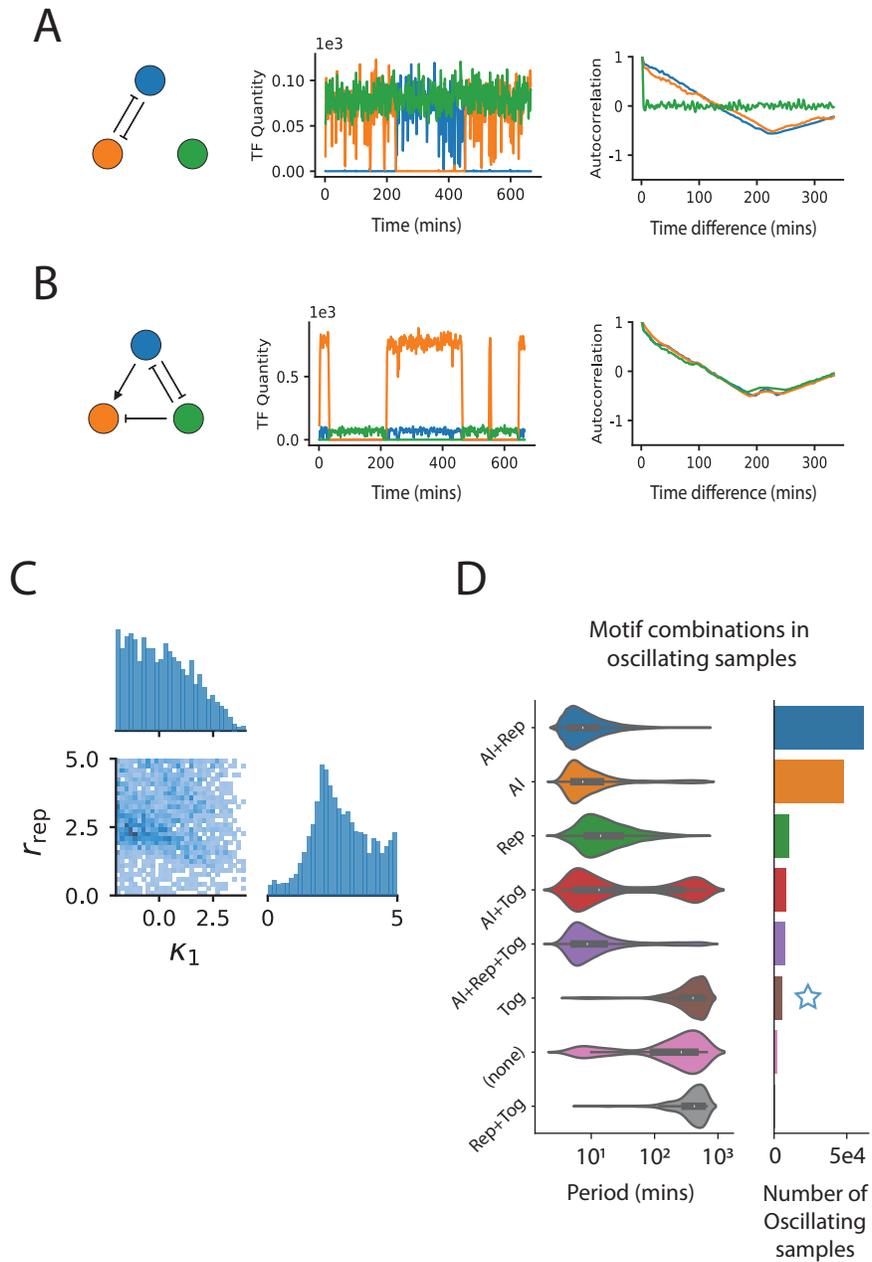


Figure S2: **Strong TF-RE binding and moderate repression enable low-frequency toggle switch “oscillations”.**

Figure S2: (A-B) Circuit topology, representative stochastic trajectory, and autocorrelation for two toggle switch topologies that were classified as oscillators, the basic toggle switch (A), ranked #148, and an amplified toggle switch (B), ranked #139. Occasional switching between stable states on a timescale comparable with the total simulation time is indistinguishable from low-frequency oscillation. (C) For oscillators containing a toggle switch and lacking either the Rep or AI oscillatory motifs, parameter sets resulting in oscillation generally showed strong TF-RE binding (low κ_1) and moderately strong repression ($r_{\text{rep}} \approx 2.5$). The latter may be a “sweet spot” that enables a persistence time on the order of the total simulation time. (D) Left: Violin plots of oscillation period for oscillators with different combinations of AI, Rep, and toggle switch (Tog) motifs. Oscillators with a toggle switch appear to have much longer periods, on the order of many hours. Right: Bar plot of the total number of oscillating samples shown in the violin plots. Samples used to generate (C) are denoted with a blue star. See also Figure 2 and Tables 1 and 2.

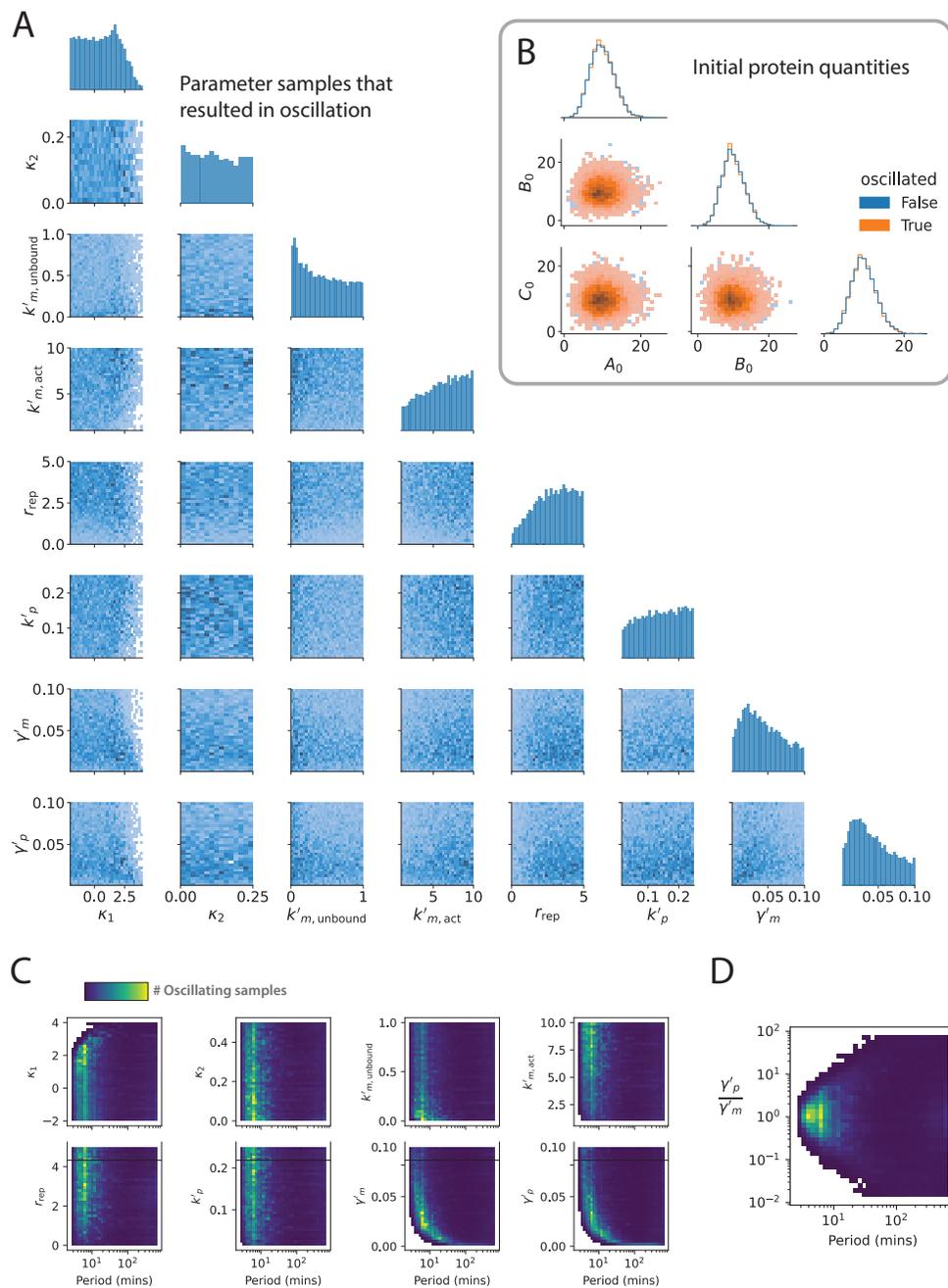
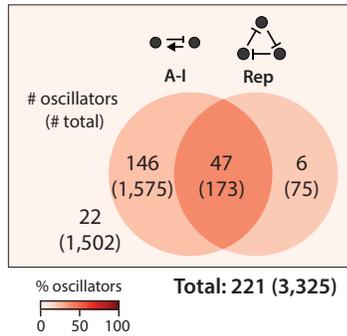


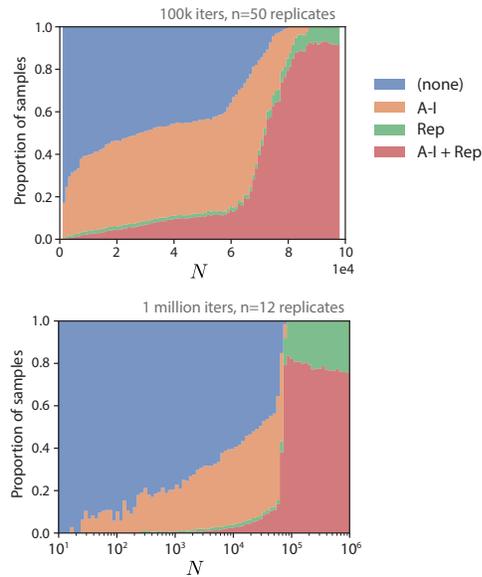
Figure S3: Dependence of oscillation likelihood and period on sampled variables.

Figure S3: (A) A corner plot of all samples taken during 3-node enumeration that resulted in oscillation, across all topologies. The histograms on the diagonal show the marginal distribution for each sampled variable, and heatmaps in the lower triangular of the grid show every pairwise dependency of these parameters. Weak binding (high κ_1) and weak repression (low r_{rep}) are the only visibly prohibitive parameter regimes for oscillation. Oscillations are favored by parameter sets with low basal expression (low $k'_{m,\text{unbound}}$), strong activation (high $k'_{m,\text{act}}$), and comparable mRNA and protein degradation rates ($\gamma'_m \approx \gamma'_p \approx 0.02$). (B) No dependence on initial protein quantities was noted. (C) 2D density plots showing the dependence of oscillation period on the variables. Only γ'_m and γ'_p appear anticorrelated with period, otherwise no dependence is observed. (D) The oscillation period, unlike the oscillation likelihood, does not seem to depend on γ'_p/γ'_m . See also Figure 2 and Tables 1 and 2.

B Topologies with AI and Rep motifs are highly enriched for oscillators



B After a phase of exploration, CircuiTree exploits (preferentially samples) A-I + Rep combinations



C Regret quantifies the cost of poor sampling

Regret

$$R_N = NQ^* - \sum_{n=1}^N r_n$$

Highest expected reward

Actual reward

D The transition from exploration to exploitation is marked by flattening regret

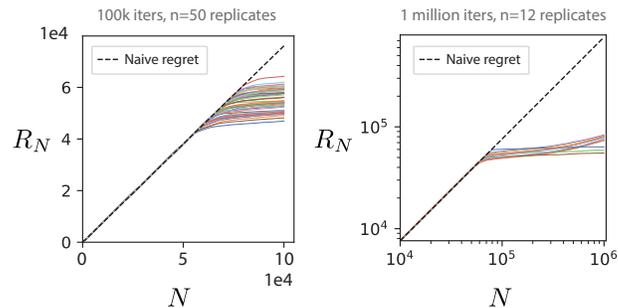
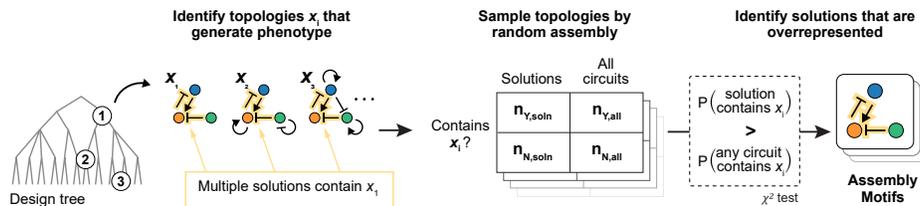


Figure S4: After a period of exploration, CircuiTree reduces sampling regret by exploiting motif combinations. (A) Venn diagram of 3-node oscillators discovered by enumeration, grouped by presence or absence of AI or Rep motifs. For each category, the number of oscillators (number of total topologies) is shown. The intensity of background color denotes the percentage of topologies in that category that are oscillators. Notably, 27.2% (47/173) of topologies with the AI-Rep combination are oscillators.

Figure S4: (B) The proportion of samples allocated to each motif combination for MCTS runs lasting 10^5 iterations (top; mean of $n = 50$ replicates; linear x-axis) and 10^6 iterations (bottom; mean of $n = 12$ replicates; logarithmic x-axis). There is a gradual shift towards sampling AI and AI-rep combinations before, at around 6×10^4 iterations, the AI-Rep motif combinations (red) and, to a lesser extent, the Rep-only category (green) are suddenly and dramatically exploited. The effect on the accumulation of reward can be seen by calculating regret. Defined in (C), regret is the opportunity cost accrued by sampling sub-optimal topologies ($Q < Q^*$). (D) At the transition between the initial phase and the exploitation phase (shown on linear axes on the left and log-log axes on the right), regret flattens because exploitation of AI-Rep combinations has increased the rate of reward. See also Figures 2 and 3.

A Overview of overrepresentation analysis



B Motif discovery correlates with motif quality, measured as average robustness

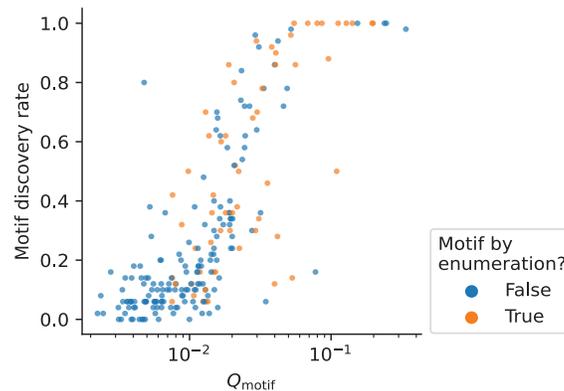


Figure S5: **Overrepresentation analysis infers assembly motifs by random sampling of the search graph.** (A) A flowchart of overrepresentation analysis, described in detail in the main Methods section. The goal is to find patterns that are overrepresented relative to the overall design space without needing to explicitly enumerate that space. This is achieved using a random sampling scheme. Unlike traditional design motifs, which are found by comparing to a flat, enumerated null distribution, these assembly motifs are (virtually by definition) good assembly strategies. This is demonstrated by the scatterplot in (B) showing the relationship between Q_{motif} and motif discovery by CircuiTree. Q_{motif} can be conceptualized as the average win probability from a given assembly state if taking random actions. Motifs with a high discovery rate by CircuiTree are likely to have high Q_{motif} and *vice versa*. Meanwhile, motifs found by traditional enumeration (orange circles), while overrepresented among solutions, are not necessarily beneficial moves in the assembly game. See also Figure 3.

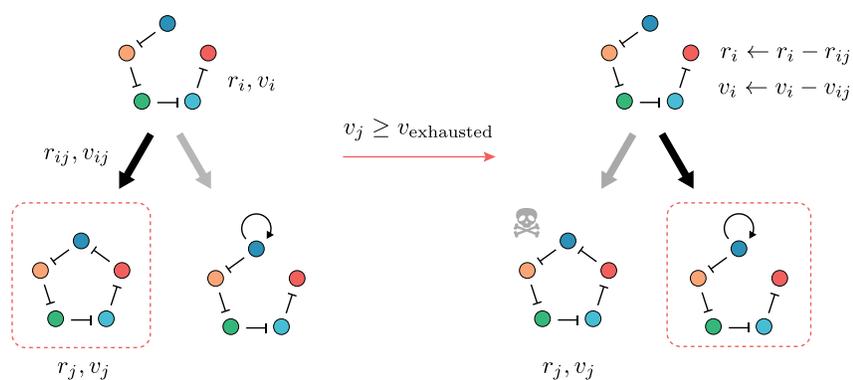


Figure S6: **States are pruned based on sampling depth to avoid over-sampling local optima.** A schematic of state pruning implemented in the parallel version of CircuiTree during the 5-node search. While asymptotically comprehensive, MCTS may persevere for many iterations in a local optimum of design space. To gently discourage this over-focusing, we implement a pruning step during selection in which a selected state that has been visited more than $v_{\text{exhausted}} = 10^4$ times (the boxed circuit on the left) is marked as “exhausted” (in the sense of exhaustively sampled), denoted by the skull and crossbones. An exhausted state s_j can no longer be selected in subsequent iterations, and each of its parent states s_i (each predecessor of s_j in T) is made to “forget” the sampling history of s_j by subtracting the visits and rewards of s_{ij} from the totals of the parent s_i .

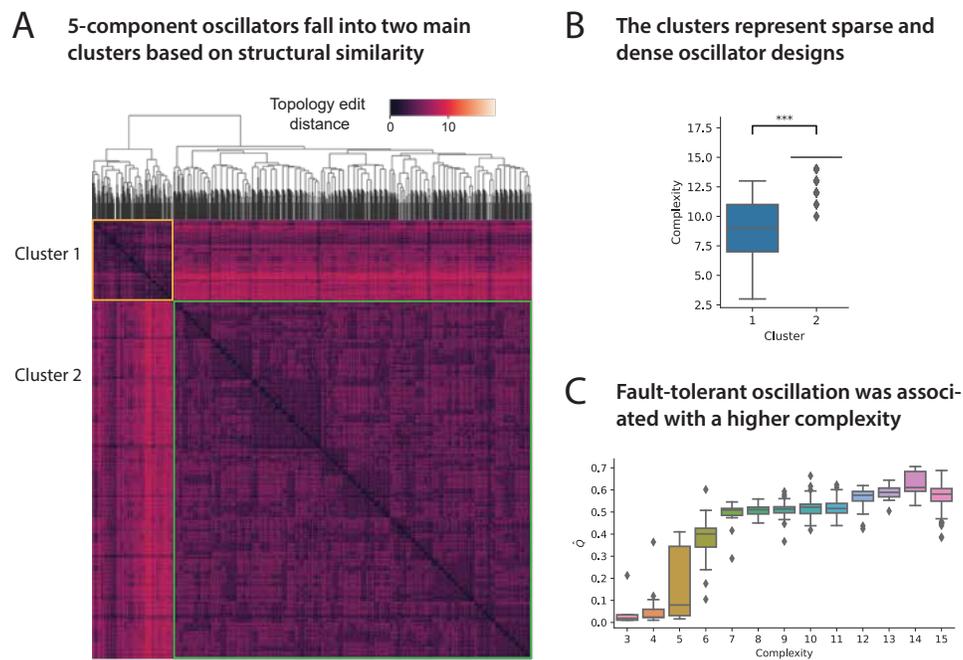


Figure S7: **Five-node fault-tolerant oscillators cluster based on topological complexity.** (A) A clustered heatmap of the 1,386 putative oscillators discovered during the search for 5-node fault-tolerant oscillators. Pairwise distance is computed as the graph edit distance between topologies (the number of edges that must be added/deleted to make two topologies equivalent). Putative FT oscillators fall into two structurally similar clusters, cluster 1 (outlined in orange) and cluster 2 (outlined in blue). (B) Box plots of topological complexity, grouped by cluster. Cluster 1 contains relatively sparse topologies, while the majority of topologies in cluster 2 have 15 interactions, the maximum allowed during search. *** $p < 0.001$ by Mann-Whitney U-test. (C) Boxplots of overall robustness \hat{Q} grouped by complexity. Regardless of cluster, higher complexity is associated with higher \hat{Q} . Specifically, there is a visible jump in robustness at 7 interactions and perhaps again at 14 interactions.

*Chapter 3***CONTROL OF SPATIO-TEMPORAL PATTERNING VIA CELL DENSITY IN A MULTICELLULAR SYNTHETIC GENE CIRCUIT**

Pre-print uploaded to biorxiv as:

Santorelli, M. et al. (Oct. 5, 2022). *Control of spatio-temporal patterning via cell density in a multicellular synthetic gene circuit*. DOI: 10.1101/2022.10.04.510900.

Abstract

A major goal in synthetic development is to design and construct gene regulatory circuits that control the patterning and morphogenesis of synthetic multicellular structures. In natural development, an interplay between growth and chemical communication shapes the dynamics of gene regulatory circuits that underlie patterning and morphogenesis. However, for synthetic gene circuits, how the non-genetic properties of the growth environment impact circuit behavior remains poorly understood. Here, we describe an occurrence of mechano-chemical coupling in synthetic contact-dependent synNotch patterning circuits. We show that the density of cells in culture modulates synNotch-ligand signal transduction. We construct a synNotch-based multicellular signal propagation circuit that is regulated by cell density. We thus then exploit this property to control the velocity of the propagating wave and the size of the activated region using small-molecule modulators of cell proliferation. Finally, we achieve millimeter-scale spatiotemporal patterning by establishing spatial gradients of cell density. Our work demonstrates that synthetic gene circuits can be critically impacted by their context, providing an alternate means for programming multicellular circuit patterning outcomes.

Introduction

Morphogenesis emerges through the interplay between chemical and mechanical processes that occur simultaneously in development to generate the architecture of the embryo (Thompson, 1917). At least in part, this is due to the tremendous growth and cell proliferation that occurs to bring a single cell to generate a multicellular organism; this cell proliferation needs to be patterned in space and time to give

rise to the diversity of forms seen in body plans. In general, during development, different kinds of chemical and mechanical processes can proceed in sequence with patterning providing a template for mechanical regulation. For example, in developing embryos, gene regulatory networks interact with intracellular signaling events to pattern domains of gene expression (e.g. *Drosophila* early development (Fukaya, 2021)). These domains can be characterized by the induction of motor proteins or differential adhesivity that deform cell shape and change embryonic geometry, driving morphogenetic events like gastrulation (Ko and Martin, 2020) or germ-band extension (Feroze et al., 2015; Zallen and Goldstein, 2017). However, recent work demonstrates that mechanical and chemical events can also be coupled in the other direction. Signaling pathways like the Yap/Taz axis can sense mechanical cues and convert these directly into changes in gene expression (Cai, Wang, and Meng, 2021). Similarly mechanical changes in the fluidity of a tissue can yield changes in gene expression dynamics (Chanet and Martin, 2014; Tschumperlin, 2011; Heisenberg and Bellaïche, 2013). Finally, the two aspects (mechanics and chemical) can be intertwined in so-called mechano-chemical systems, that seem to abound in developmental transitions (Veerman, Mercker, and Marciniak-Czochra, 2021; Urdy, 2012; Hannezo and Heisenberg, 2019; Chan, Heisenberg, and Hiragi, 2017; Schiffhauer and Robinson, 2017; Scott, Weinberg, and Lemmon, 2019; Goehring and Grill, 2013; Mo R Ebrahimkhani and Ebisuya, 2019; Le Roux et al., 2019; Mao and Lecuit, 2016; Howard, Grill, and Bois, 2011). Despite emerging examples, many principles remain obscure regarding how information flows between mechanical processes and chemical circuits in general, and how this contributes to expand, constrain or regulate patterning and morphogenetic outcomes.

The complexity of the embryo presents inherent challenges to studying and extracting general principles both of developmental transitions in general, and of interactions between mechanical properties and gene circuit signaling dynamics in particular. A major emerging theme in the field of synthetic development is the construction of gene circuits that enable controlled morphogenesis of synthetic embryos and organoids. Highly simplified engineered systems have been generated in this fashion that provide controlled and defined experimental systems in which to analyze gene circuits within the context of a multicellular structure (Santorelli, Lam, and Morsut, 2019; Toda, Frankel, and Lim, 2019; Mo R Ebrahimkhani and Ebisuya, 2019; Velazquez et al., 2018; Ho and Morsut, 2021; Zarkesh et al., 2022; Mo R. Ebrahimkhani and Levin, 2021; Schlissel and Li, 2020; Davies, 2017; Teague, Guye, and Weiss, 2016). Engineered systems provide an ideal setting to study

mechano-chemical coupling where signaling and mechanical phenomena can be isolated, measured and modulated.

Synthetic circuits based on Notch signaling, or so-called synNotch circuits, have emerged as a modular and flexible strategy for engineering multicellular mammalian systems (Morsut et al., 2016; Toda, Blauch, et al., 2018; Toda, McKeithan, et al., 2020). The synNotch system uses engineered receptors modeled after the endogenous developmental signaling pathway Notch/Delta. This endogenous pathway is contact-dependent and is used extensively during development to generate cell-scale patterns. In the synthetic version, synNotch, both the input and output of the pathway have been rendered user-definable and, as such, are orthogonal to the endogenous Notch pathway. Using this system, developmental circuits have been developed in 2D culture as well as in 3D fibroblast aggregates where a synthetic signal affects multicellular signaling and mechanics, for example by driving expression of key adhesion proteins in the cadherin family (Morsut et al., 2016; Toda, Frankel, and Lim, 2019; Toda, McKeithan, et al., 2020). By changing the nature of the adhesion proteins or the architecture of cell-cell communication, circuits for a variety of different developmental trajectories have been implemented.

In the synNotch adhesion example, information flows from engineered signaling proteins to downstream effects on mechanical properties of the cell through changes in cell-cell adhesion. To achieve a complete synthetic mechano-chemical system with reciprocal information flow between both modalities, mechanical inputs to signaling must also be characterized. Insights that synNotch could be a good candidate to develop such a system are emerging. First, the various proposed mechanisms of activation for Notch and synNotch signaling involve a mechanical "pulling force" that exposes the protease cleavage site for further signal transduction (Kopan and Ilagan, 2009; Gordon, Arnett, and Blacklow, 2008; Sprinzak and Blacklow, 2021; Lovendahl, Blacklow, and Gordon, 2018; Stassen, Ristori, and Sahlgren, 2020), although the specific mechanisms may differ based on the cellular context and the endogenous or synthetic nature of the receptors (Khamaisi et al., 2022). Second, cellular mechanical tension, shear stress, and ECM stiffness have been shown to play a role in Notch signaling in certain contexts on the sender (ligand-expressing) cell side or the receiver (receptor-expressing) cell side (Mack et al., 2017; Hunter et al., 2019; Matsuo et al., 2021; Weijts et al., 2018; Meloty-Kapella et al., 2012; Theodoris et al., 2015). Third, synNotch has been engineered to respond to different degrees of pulling force, highlighting the mechanosensory

potential of this signaling modality (Sloas and Ngo, 2022). Whether these inputs can be used to create a system where mechanics or cell proliferation affects not only signaling, but also patterning outcomes has not been explored.

Mathematical models have been an important tool for understanding morphogenesis in natural systems (Turing, 1952; Gierer and Meinhardt, 1972; Murray, 2002) and thus provide a potential strategy for the design and analysis of synthetic systems that incorporate mechanical-chemical coupling. Cell-based models of Notch-mediated signaling (Binshtok and Sprinzak, 2018) have uncovered key insights into the self-organization of regular spatial patterns (Collier et al., 1996), the regulation of cell fate bifurcation by receptor-ligand interactions and cell geometry (Sprinzak, Lakhanpal, LeBon, et al., 2010; Sprinzak, Lakhanpal, LeBon, et al., 2011; Shaya et al., 2017), and the important roles of ligand expression levels and competition in robust patterning (Sprinzak, Lakhanpal, LeBon, et al., 2011; Formosa-Jordan, Ibañes, et al., 2012; Petrovic et al., 2014). In addition to endogenous signaling, mathematical modeling of synthetic signaling networks has been a key tool for engineering defined, controllable biological circuitry. Recently, cell-based modeling has been used similarly to catalyze the discovery and design of novel circuits for morphogenesis (Lam et al., 2022; Mulberry and Edelstein-Keshet, 2020). Such models have been used to study natural cases of mechano-chemical coupling (Shaya et al., 2017; Cohen and Sprinzak, 2021; Hufnagel et al., 2007; Pan et al., 2016) but have not yet been applied to synthetic cell systems.

A deeper analysis of mechanical-chemical coupling, especially the ones connected with cell proliferation, in synthetic gene circuits could provide new insights into the logic of these circuits and uncover novel strategies for engineering multicellular system patterning and morphogenesis. Here, we first identify cell density as a non-genetic parameter of cell culture that affects synNotch signaling through a screening of mechanical inputs. We then apply the synNotch system to study the impact of cell proliferation and cell density on a paradigmatic example of signaling-mediated patterning, i.e. lateral propagation via contact-dependent signaling. We construct multicellular tissues containing a local relay circuit consisting of a sender (signal-originating) cell type and a novel transceiver (signal-propagating) cell type that both receives and sends fluorescent signal. Using this simplified genetic circuit, we investigate the impact of non-genetic inputs on patterning and signal propagation at the multicellular level. With a combination of *in silico* and *in vitro* experiments, we show that density modulates patterning in a parameter space that is biologically

relevant and can be exploited to construct distinct macroscopic spatial and temporal patterns.

Results

Cell density impacts SynNotch signal transduction

The impact of non-genetic factors including tissue mechanics and cell density on synNotch signal propagation has not been explored. To quantify the impact of individual perturbations to the physical environment on synNotch signaling, we used a previously reported in vitro assay for synNotch activation based on a sender-receiver cell signaling paradigm. Briefly, two L929 mouse fibroblast cell lines, a sender cell line and a receiver cell line, are engineered such that contact between a sender cell and receiver cell can be assessed by the presence of a red fluorescent reporter in receiver cells. Sender cells constitutively express membrane-bound green fluorescent protein (GFP), which acts as the ligand for an anti-GFP synNotch receptor on receiver cells (anti-GFP synNotch, Figure 1A). The intracellular portion of the anti-GFP synNotch receptor contains a tetracycline-controlled transactivator (tTA) which is freed from the membrane upon contact-dependent activation and translocates to the nucleus where it activates expression of cytosolic mCherry. To assay synNotch activity, sender and receiver fibroblasts are co-cultured in a 1:1 ratio for 24 hours, by which time activated receiver cells produce mCherry (Figure 1B). Expression of mCherry is then quantified by fluorescence-activated cell sorting (FACS) at 24 hrs. This assay is performed in two control conditions as well as conditions of individually varied extracellular matrix (ECM) composition, substrate stiffness, cytoskeletal tension, and cell density. The control activation is performed on tissue-culture treated plastic dishes, at a 100% confluent cell density, in the absence of drug treatment. In the positive and negative controls (labeled "ON" and "OFF") sender cells are present and absent, respectively. We report and refer to the density of the cell culture as a multiple of the density at 100% confluence (cells cover 100% of the surface). We estimated by visual inspection that the confluent density is 1250 cells/mm² and refer to it as 1× confluence or simply "1x".

We found that increased cell density, but not perturbations to other aspects of the physical cell environment, led to attenuation (downregulation) of synNotch signalling (Figures 1C-F). Cells grown on different substrates (mimicking different ECM compositions) or at varying stiffnesses exhibited similar mCherry activation to the sender-activated "ON" reference condition (Figure 1C; for details see Methods).

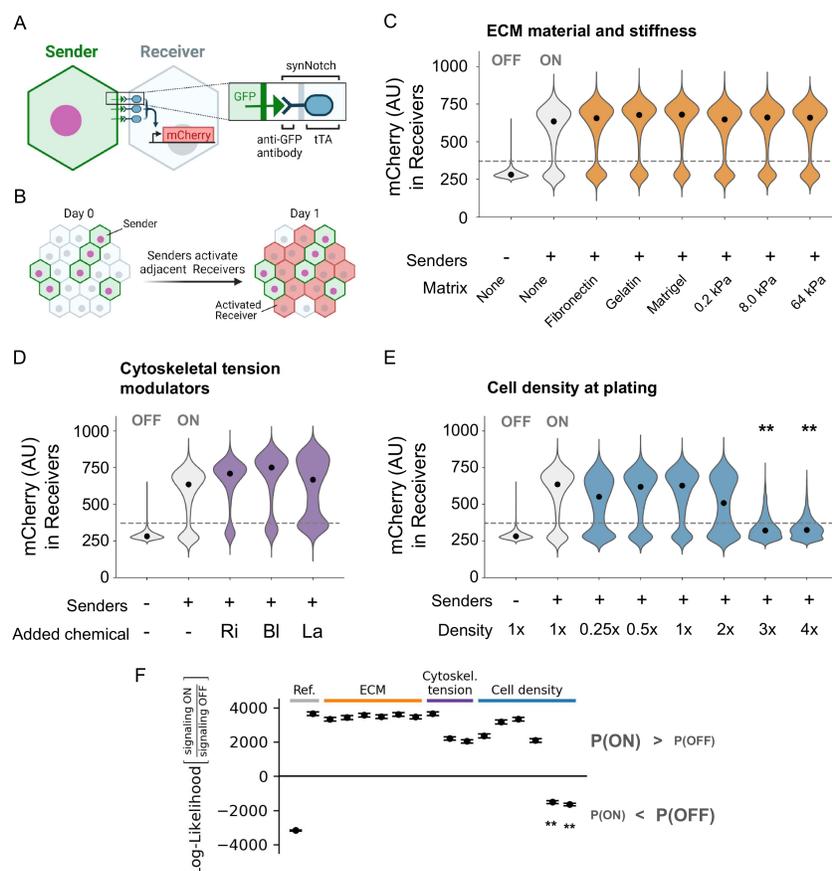


Figure 1: SynNotch activity depends on cell density, not ECM or cytoskeletal tension. (A) Conceptual scheme of Sender-Receiver synNotch signaling. Membrane-bound GFP (green triangle) in Sender cells binds synNotch in Receiver cells on the right. Cleavage of synNotch frees the intracellular domain (tTA, blue ellipsoid) to translocate to the nucleus and activate mCherry reporter expression. (B) Schematic depiction of synNotch signaling response assay. Senders and Receivers are co-cultured at 1:1 ratio, and synNotch signaling occurs at Sender-Receiver contacts (activated Receivers in red). Receiver activation (mCherry fluorescence) is measured at 24 hours by FACS under different culture conditions (C-E). Each violin in (C-E) depicts a distribution of mCherry fluorescence in Receivers. Gray violins are the reference samples for OFF and ON Receiver states (without and with Senders, respectively). Black dots indicate median. Dotted line indicates the cell-wise threshold between OFF and ON states. ** indicates the sample is more likely OFF than ON, as determined by the log-likelihood (LR) statistical comparison. Distributions are shown across (C) various growth substrate materials and stiffnesses, (D) various chemical modulators of cytoskeletal tension[†], and (E) various initial cell densities. Gray dashed line is the fluorescence cutoff between OFF and ON. (F) Scatter plot of the LLR calculated for each sample. Points above zero indicate the sample resembles the ON state more than the OFF state. Error bars represent 95% confidence intervals (calculated by bootstrapping). For details please refer to Methods. [†]Ri: 100 μ M ROCK-inhibitor (Y-27632); BL: 25 μ g/mL blebbistatin; LA: 200 μ M latrunculin-A.

Cytoskeletal tension was modulated by the addition of three drugs known to affect cytoskeletal contractility and actin polymerization - Y-27632 (ROCK-inhibitor) (Amano, Nakayama, and Kaibuchi, 2010), blebbistatin Straight et al., 2003, and latrunculin-A (Spector et al., 1983; Yarmola et al., 2000). These treatments did not suppress signaling activity. Notably, they may slightly increase synNotch activation in the presence (Figure 1D) and absence (Figure S1D) of senders. In contrast, when cells were grown across a range of cell densities from 0.25x - 4x confluent density ($312.5 - 5000 \text{ cells/mm}^2$), densities of $\geq 3x$ exhibited significantly inhibited signaling responses in receivers (Figure 1E, ** indicates statistical significance). A sample was considered activated overall based on its ON:OFF log-likelihood ratio (LLR), plotted in Figure 1F. See Methods for details on statistical calculations. Though not statistically significant, signaling activity also appears slightly inhibited at densities of 0.25x and 2x. These results suggested that synNotch signal transduction is optimally efficient at cell densities of 0.5x - 1.0x ($625 - 1250 \text{ cells/mm}^2$), whereas much lower or higher densities result in blunted signal transduction.

The observed decrease of synNotch signaling at cell densities above 1x could be due to a number of molecular mechanisms. To start addressing this question, Figure S3 summarizes several relevant experiments comparing relevant variables between 1x and 4x confluent densities. As density increases past 1x we find: (i) Cells change shape, becoming smaller surface-wise (Fig. S3A) and volume-wise (Fig. S3B), and becoming more rounded/circular (Fig. S3C). (ii) Membrane-targeted GFP ligand expression decreases in a way that corresponds with density-dependent attenuation of signaling (Fig. S3D-F). (iii) Cell size and the quantity of GFP ligand are positively correlated (Fig. S3G-H). (iv) Cell motility may decrease (Fig. S3I). However, other parameters are not affected by density: (i) YAP localization does not change (Fig. S3J), (ii) conditioned media does not affect signaling (Fig. S3K), and (iii) nuclear reporter protein accumulation levels do not change (Fig. S3L). For further details, see Methods.

Our measurements were most consistent with the hypothesis that density modulates membrane trafficking and thus membrane protein levels. Given that YAP localization does not change, and cytoskeletal drugs do not affect signaling (Figure 1D), it seems unlikely that the decrease in signaling at high densities is due to classic cytoskeletal-tension-mediated mechanotransduction. Given that growth media conditioned from high-density (4x) cultures does not alter signaling at 1x, soluble chemical factors also seem unlikely culprits. We instead hypothesize lo-

cal, curvature-based changes in membrane trafficking resulting in faster membrane protein turnover correlating with higher membrane curvature, as observed for other signaling molecules (Alfonzo-Méndez et al., 2022; De Belly et al., 2021; Chiasson-MacKenzie et al., 2018). This etiology is supported by the fact that the membrane reporter but not the nuclear reporter is affected. Thus we consider the nature of the mechanism seems to be local and mechanical in nature rather than long-range/soluble and chemical.

Signal propagation circuit exhibits density-dependency patterning outcome

We next sought to investigate the impact of density-dependent signal attenuation on the behavior of a multicellular synNotch patterning circuit. We focus on the "lateral propagation" circuit, a paradigmatic example of emergent patterning wherein a signal is relayed from cell to cell via contact-dependent signaling between neighboring cells (Sjöqvist and Andersson, 2019). This system has not been used for synthetic patterning, though it has been engineered before in a semi-synthetic manner (Matsuda et al., 2012). The circuit is shown schematically in Figure 2A and relies on cells that can both receive and send a cell-surface signal, or "Transceiver" cells. When activated by contact with the ligand (for instance, presented by an initiating sender cell) transceiver cells become activated and can relay the signal to neighboring Transceivers, triggering a propagating wave of signaling by relay.

We implemented this circuit *in vitro* in a mouse L929 fibroblast cell line. Transceiver cells were generated that express an anti-GFP receptor activating transcription of the GFP ligand. To do so, three transgenes were stably integrated (Figure 2B): one constitutively expresses a synNotch receptor with an anti-GFP nanobody (α GFP) as the extracellular domain and the transcription factor tetracycline transactivator (tTA) as the intracellular domain. A second expresses membrane bound GFP (synNotch cognate ligand) under control of the tetracycline responsive element (TRE) promoter. A third expresses cytosolic mCherry driven by TRE as a reporter of synNotch activation and constitutively expresses the blue fluorescent protein tagBFP to mark transceivers. We generated several clones of L929 cells where these transgenes were all integrated into the genome. Transceiver clone function was evaluated by co-culturing with sparse sender cells and performing high-magnification time-lapse imaging centered on individual sender cell foci.

The area of GFP fluorescence was calculated using a semi-automated image analysis workflow (see Methods). At a cell density of 1250 cells/mm² (100% confluent, or

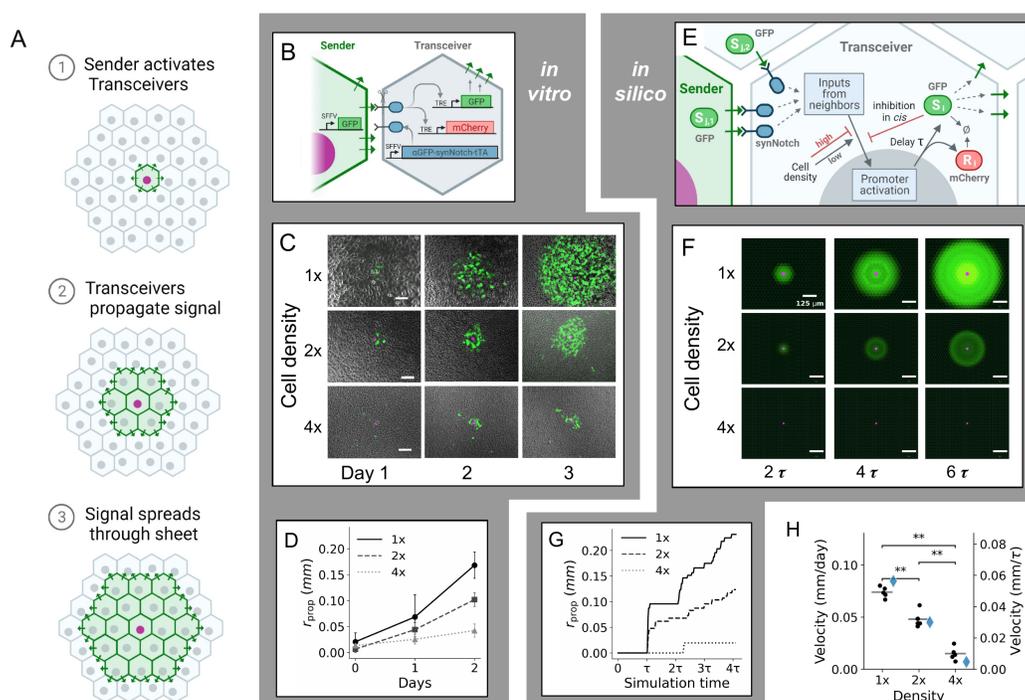


Figure 2: Cell density tunes the velocity of signal propagation. (A) Schematic of signaling wave propagation (green) in a monolayer of transceiver cells (gray) initiated by a sender cell (purple nucleus). (B) Schematic of the *in vitro* Transceiver circuit. The Sender cell (green) contains a transgene for constitutive expression of membrane-bound GFP ligand (green triangle) and a nuclear infrared fluorescent marker (not shown) under control of spleen focus-forming virus (SFFV) promoter. The Transceiver cell (gray) has a transgene for constitutive expression of GFP-sensing synNotch receptor with a tetracycline-controlled transactivator (tTA) intracellular domain and two transgenes with tTA-responsive elements (TRE) activating GFP ligand and mCherry (see Methods). (C) Micrographs demonstrating *in vitro* signal propagation over time at the indicated cell densities. Bright field (grayscale) is overlaid with GFP signal (green) and nuclear infrared fluorescent marker expressed in sender cells (purple). A single propagation focus is shown, isolated from a culture well with multiple foci. Scale bar 100 μm . See Supplemental Movies 1-3 for time-lapse movies. (D) Propagation radius (r_{prop}) over time for three cell densities ($n = 5$, mean \pm s.d.). (E) Cell-level mathematical model of Transceiver signaling. A sender cell (left) presents GFP ligand s (green ellipsoid and triangles) to a Transceiver cell (center, " i "). Ligand from cell i 's neighbors (" j ") activates SynNotch receptors (blue). Activated SynNotch stimulates production of ligand and a reporter r (red ellipsoid) after a time delay τ . Production is regulated by cell density. The ligand s also inhibits its production ("cis-inhibition"). \emptyset indicates degradation. (F) *In silico* simulation of transceiver signaling demonstrating GFP ligand (green) propagation from a sender cell (purple) at cell densities of 1x, 2x, and 4x. Scale bar 125 μm . See Supplemental Movie 4 for time-lapse. (G) Propagation velocity (r_{prop}) *in vitro* for densities of 1x, 2x, and 4x. (H) Strip plot of propagation velocity *in vitro* (black dots; horizontal line indicates mean) and *in silico* (blue diamonds) at indicated cell densities. ** $p < 0.01$, two-sided Mann-Whitney-Wilcoxon test.

"1x"), transceiver fibroblasts in contact with a sender cell express the GFP ligand, triggering a propagating wave of transceiver signaling that travels outward through the cell monolayer (images in Figure 2C; see Supplementary Videos 1-3). The sizes of $n = 5$ signaling discs were quantified by calculating the radius of propagation r_{prop} (Figure 2D), resulting in a mean velocity of $\Delta r_{\text{prop}}/\Delta t = 0.131 \pm 0.009$ mm/day at 1x confluent density. When initial cell density is increased, wave velocity slows to 0.085 ± 0.014 mm/day at an initial density of 2x (2500 cells/mm²) and 0.026 ± 0.011 mm/day at an initial density of 4x (5000 cells/mm²). Similar trends were observed for all clones generated (data not shown). Thus, this synthetic lateral propagation circuit can generate signaling waves whose velocity is responsive to cell density.

To gain additional insight, we devised a computational model of synNotch signaling with density-dependent attenuation, and used it to study the impact of density-dependent signal transduction on signal propagation in a multicellular sheet. To model the multicellular sheet, we use a fixed lattice of cells, a framework used extensively to study Notch-mediated patterning (Collier et al., 1996; Sprinzak, Lakhanpal, Lebon, et al., 2010; Sprinzak, Lakhanpal, LeBon, et al., 2011; Shaya et al., 2017; Formosa-Jordan, Ibañes, et al., 2012; Formosa-Jordan and Ibañes, 2014; Binshtok and Sprinzak, 2018). In our model, each cell occupies a region on a hexagonal lattice. For a given density, the area of each hexagon on the lattice is set equal to the average area of cells in a confluent monolayer *in vitro*. As depicted in Figure 2E, each cell contains a system of chemical reactions that model synNotch signal transduction (see Methods, Equation 3.4). A given cell (i) in direct contact with its neighbors (j) can express both the signaling ligand (s) and the reporter protein (r). Cells either express the ligand constitutively if they are senders or upon synNotch stimulation if they are transceivers. To model cytoplasmic projections, which are known to affect Notch-mediated patterning systems (Binshtok and Sprinzak, 2018; Vasilopoulos and Painter, 2016; Hadjivasiliou, Hunter, and Baum, 2016), contact strength is weighted by cell-cell distance on the lattice (weights visualized in Figure S5A).

Following ligand-induced receptor activation, the recipient cell responds after a time delay (τ) corresponding to transcription, translation, and membrane trafficking of the ligand. Finally, ligand expression in *cis* decreases a transceiver cell's capability to sense ligand expressed by neighbors in *trans*, a phenomenon termed *cis*-inhibition that is observed to regulate endogenous and synthetic Notch signaling (Sprinzak,

Lakhanpal, Lebon, et al., 2010). Please see Figure S5B for an example of sender, receiver, and transceiver activation dynamics *in silico*. Changes in cell density are modeled by changing the size of all cells equally. At a higher density, for example, cell size is reduced to occupy less area while preserving the hexagonal lattice (Figure S5C, inset images). The density-dependence of signaling is modeled by multiplying the amount of ligand involved in signaling by a coefficient that encodes the efficiency of signaling. This coefficient decays exponentially as cell density increases above 1x density (Figure S5C, blue curve) or decreases below 1x and was parameterized by comparison with the propagation data in Figure 2.

With this *in silico* model of density-dependent, signal-propagation circuit, we sought to simulate propagation on a monolayer lattice at different densities. To do so, we initialized a 50×50 hexagonal lattice of transceiver cells with a single sender cell. At time $t = 0$, the sender cell begins expressing the signal and Equation 3.4 are integrated forward in time (see Methods and Supplementary Text for details of mathematical modeling). We perform the simulation at cell densities of 1X, 2X, and 4X confluence (1250, 2500, and 5000 cells/mm²). In the simulations, a wave of activation begins propagating outwards from the Sender at a speed that depends on density (selected time-points rendered in Figure 2F; see Supplementary Video 4). The area of propagation was quantified as the amount of lattice area occupied by Transceivers expressing ligand at a level greater than the promoter threshold k (see 3.1 for values of all parameters used in simulation). Wave velocity was calculated similarly to the *in vitro* case. As shown in Figure 2G, in the 1x condition propagation area begins to increase after a time delay $\tau = 0.3$ for signal production and then continually rises. We show simulation time in units of τ as a characteristic time-scale. At 1x density, *in silico* wave velocity (calculated similarly to *in vitro*) is $0.103 \text{ mm}/\tau$ and drops to $0.055 \text{ mm}/\tau$ and $0.009 \text{ mm}/\tau$ at 2x and 4x densities, demonstrating that the attenuation of signaling at higher cell densities is sufficient to explain the slower speed of Transceiver propagation. Figure 2H compares the mean velocities of the experimental system and computational model. In both cases, increasing cell density above 100% confluence has a substantial suppressive effect on the velocity of the wavefront. See Supplemental Movie 4 for time-lapse lattice simulations at 1x, 2x, and 4x densities.

Collectively, these results show that the signaling dynamics of a synthetic circuit can be modulated via cell density. They further suggest that cell density could be used as a mechano-chemical control mechanism to dynamically pattern synthetic tissues

without requiring the engineering of additional biochemical circuit components.

Signal propagation reaches self-limiting regimes due to cell population growth

Cells in culture grow and divide, thereby increasing cell density over time. How does cell growth impact signaling behavior over longer time courses? Figure 3A shows an example of a seven-day time course in which transceivers were co-cultured with senders at a plating density of 1250 cells/mm² (1x). As shown in the GFP channel images, transceivers begin propagating signal by Day 1 of growth, and the signaling wave propagates outwards over the first three days. By Day 4-5, however, propagation speed and overall signal intensity start to decline, and by Day 7 GFP expression is almost fully suppressed. This resulted in self-limiting propagation with a characteristic diameter of 0.5 mm. (see also Supplemental Movies 5 and 6 for time-lapse movies). Interestingly, transceivers that express and then down-regulate GFP signal continue to express the mCherry reporter at Day 7 (mCherry channel), suggesting slower degradation kinetics. Importantly, cells re-plated at 1x density after a seven-day time course are still capable of propagation after re-plating, suggesting that GFP down regulation is not a result of decreased cell health or death during long-term culture (Figure S6).

Given the negative correlation between cell density and signaling, we hypothesized that the decrease in GFP signaling at later times is due to an increase in cell density secondary to cell proliferation. To test this hypothesis, we counted cell numbers over the time course in culture. Transceivers were plated at initial densities of 1250, 2500, and 5000 cells/mm² (1x, 2x, and 4x), and $n = 3$ cell counts were performed daily for seven days using an automated cell counter. Results for the 1x condition shown in Figure 3B demonstrate how cell density (black, mean \pm s.d.) increases over time in a sigmoid fashion and begins to plateau by 4-5 days of culture. By day 3 — around the same time signaling begins to shut off in Figure 3A — the culture achieves a density of ≥ 3750 cells/mm² (3x) which was found to be inhibitory to transceiver signaling (Figure 1E, 3x and 4x density). These results supported the hypothesis that in long-term culture, cell proliferation can shut off signaling in previously activated transceivers if the culture achieves cell densities that are not conducive to cell-cell signaling.

We then used our computational model to determine if density-induced repression is sufficient to replicate signaling attenuation in long-term cultures. We model population growth as a logistic growth process. The logistic growth equation is a common

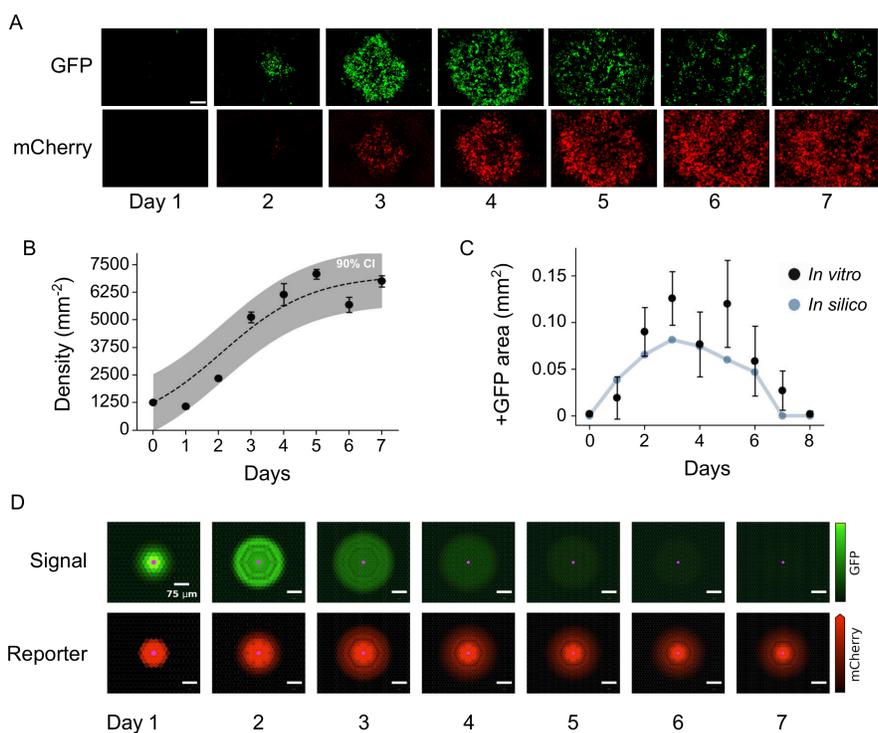


Figure 3: Cell population growth over time leads to self-limiting activation. (A) Propagation and attenuation of signal over a seven-day time-course. Fluorescence micrographs were taken of an isolated propagation focus from a sender:transceiver co-culture plated at a density of 1250 cells/mm^2 (1x). GFP produced by senders and activated transceivers is shown in green, and mCherry (reporter for synNotch activation in transceivers) is shown in red. Signal begins propagating in a disc until around day 4. From then, GFP levels decrease and are mostly decayed by day 7. The reporter remains as a record of past activation, possibly due to slower degradation kinetics. Scale bar $100 \mu\text{m}$. See Supplemental Movies 5 and 6 for time-lapse movies. (B) Graph of cell density over time. Black dots are the mean and standard deviation ($n = 3$) of cell density from sender-transceiver co-cultures, measured by automated cell counting. The dashed line shows the logistic growth equation with estimated parameters (90% confidence interval in gray). (C) Quantification of the signaling disc area over time. *In vitro*, 1:100 co-culture of senders:transceivers was plated at 1x density, and the areas of $n = 5$ individual foci were imaged daily. *In silico*, an 80×80 lattice of transceivers with one sender was simulated from an initial density of 1x. Activated area was measured as in Figure 2 for *in vitro* (black dots, mean \pm s.d.) and *in silico* (connected blue dots). Cell density is increased over time by reducing cell size to account for population growth as estimated in (B). Simulation time-scale was parameterized using the estimated generation time (see Results and Methods). Population growth thus appears sufficient to explain the experimentally observed dynamics of ligand activation followed by attenuation. (D) Rendering of model simulation. The model consists of a hexagonal lattice with one sender cell (purple) surrounded by transceivers at 1x density. GFP ligand (first row, green) and mCherry reporter (second row, red) concentrations in transceivers are rendered at daily time-points. See Supplemental Movie 7 for time-lapse.

sigmoidal growth model in which a species will proliferate exponentially until it approaches the maximum population density its environment can support (termed the "carrying capacity" ρ_{\max}), at which point density asymptotically approaches ρ_{\max} . Like exponential growth, logistic growth starts at an initial population density ρ_0 and has an intrinsic growth rate g . We applied a maximum-likelihood estimation (MLE) procedure to our cell count data and estimate that $g = 0.616 \text{ days}^{-1}$ (90% CI: $0.539\text{--}0.711 \text{ days}^{-1}$) and $\rho_{\max} = 7337 \text{ cells/mm}^2$ (90% CI: $6954\text{--}7771 \text{ cells/mm}^2$). In Figure 3B, we plot the fitted logistic growth equation (black dashed curve, 90% CI in gray), which shows good correspondence with the observed growth data. Please refer to Methods and Figure S7 for detailed MLE procedure and results. Given this parameterization, the generation time of cells is $g^{-1} \ln 2 = 1.125 \text{ days}$ (27.0 hrs). Under the simplifying assumption of a constant protein removal rate, we approximate the ligand half-life as equal to the generation time, allowing us to convert between simulation time and real units (see Methods for a detailed discussion).

Transceiver signal propagation was then simulated with logistic growth of the population density over time. We generated an 80×80 lattice of transceivers with one sender at the center at an initial density of 1x and ran the simulation for 8 days. Changing population density in our system is modeled by reducing the area of cells on the lattice rather than by explicit cell division, mitigating computational complexity (see Methods for details). The propagation area was then measured as in Figure 2 and plotted over time in Figure 3C (blue) alongside the radii of $n = 5$ *in vitro* propagation foci (black, mean \pm s.d.) under the same experimental conditions as Figure 3A. The simulated dynamics of signaling are similar to that of the *in vitro* results. In particular, the *in silico* model also exhibits an early period of 3 days of activation characterized by GFP and mCherry production and expansion of the propagation disc, followed by a period of signal attenuation during which GFP expression falls and eventually reaches levels similar to day 0. Renderings of the simulation at daily time-points are shown in Figure 3D. To model the lingering nature of the mCherry reporter, mCherry was modeled under the same activation kinetics as GFP but with 10 times slower degradation kinetics (see Methods). See Supplemental Movie 7 for a time-lapse rendering of the simulation showing GFP and mCherry levels.

These results demonstrate that cell density changes are sufficient to explain the observed attenuation of transceiver activation and signal propagation. Overall, we show using experimentation and mathematical modeling that signal propagation

through a proliferating transceiver population can have a transient, self-limiting nature consistent with a density-induced attenuation of synNotch activity.

***In silico* exploration of growth parameters reveals distinct phases of activation explained by a critical density**

Using the computational model, we then investigated the generative possibilities of this mechano-chemically coupled circuit in the confluent to super-confluent regime (densities greater than 1280 cells/mm²). So far, in the system, we have observed that the speed of propagation can be controlled via initial cell density (Fig. 2), and also over time due to cell proliferation (Fig. 3). This observation lead to identify initial cell density, and cell proliferation as two variables affecting patterning outcomes. In order to define the space of achievable qualitative and quantitative phenotypes for the density-modulated Transceiver signaling circuit, we simulated circuit behavior for different values of the initial density ρ_0 and proliferation rate g and generated a phase diagram of signaling phenotypes. For each parameter combination, a 50×50 lattice of transceiver cells and one Sender were simulated for 8.0 days. We observe signaling behavior that falls into three categories. Either signal begins to propagate or not, and if propagation begins, Transceivers either all become deactivated by the end of the simulation or remain active to some extent. Parameter sets thus fall into three categories, or "phases": attenuated, self-limited, or unlimited propagation. For classification, signaling dynamics at early and late time points are used. Parameter sets are labeled "attenuated" if the initial signal production rate $v_{\text{init}} = \max_i \dot{s}_i(t = \tau)$ is below a chosen threshold $v_{\text{thresh}} = 0.25$ (Figure S8B). If the initial production rate is above this value, they were labeled "self-limited" if all Transceivers become inactivated by the end of the time-course and "unlimited" otherwise. As above, a transceiver i was considered activated based on the amount of expressed ligand ($s_i > k$; see Table 3.1).

Each of the three behavioral phases corresponds to a discrete region of the phase diagram. In the dark blue region of Figure 4A, Transceiver activation persists throughout the time-course and propagation is unlimited. In the blue region of limited propagation, the wave of activation initiates but becomes fully attenuated at some point during the time-course. In this regime, Transceiver activation is limited in time and space, and the maximum area achieved during the time-course depends on both ρ_0 and g (Figure 4B). Finally, in the gray region, no activation occurs and signaling has been fully turned off above a critical density $\rho_c^{\text{high}} = 3.3$ that

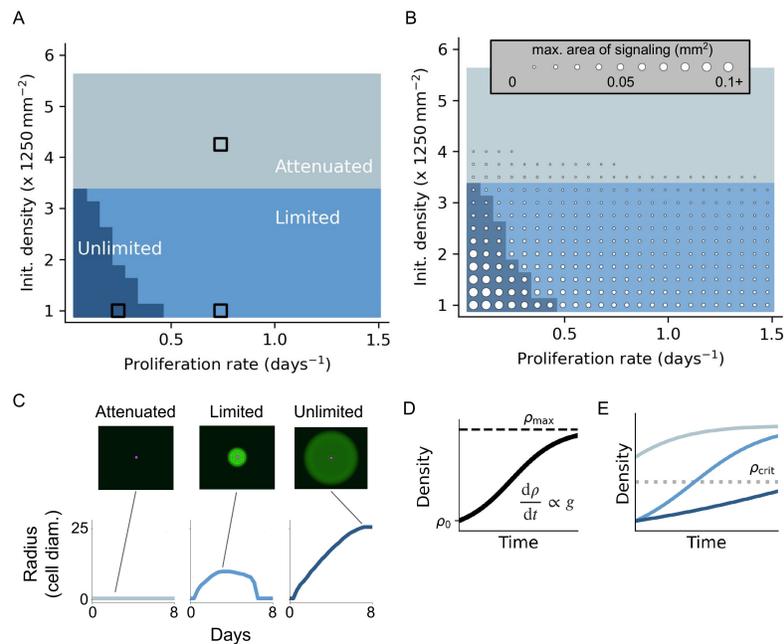


Figure 4: Population growth determines the signaling behavior of transceivers. (A) Phase diagram showing the transceiver propagation phase as a function of initial cell density and proliferation rate. For each simulation, a 50×50 lattice of Transceiver cells and one Sender were simulated with different values of the growth parameters g (intrinsic proliferation rate) and ρ_0 (initial density) for 9.5 days. Parameter sets were then classified into distinct phases based on signal propagation behavior at early and late time-points: unlimited (dark blue), self-limited (light blue), or attenuated (gray) propagation. In the unlimited phase, the ligand production rate is too low at early time-points to trigger activation. In the self-limited phase, activation occurs but all activated Transceivers become deactivated by the end of the simulation, and in the unlimited phase, some activated cells remain. Black squares are example parameters for which dynamics are shown in (C). (B) Phase diagram annotated with circles showing the maximum area of the propagation disc over the whole time-course. The disc reaches negligible sizes in the attenuated phase and persists at the end of simulation time in the unlimited phase. However, in the self-limited phase, propagation reaches a finite area. Due to the reporter's slow degradation kinetics, transceivers "remember" prior activation. Thus, the maximum area of propagation could be tuned by modifying the parameters of growth and subsequently maintained in molecular memory. (C) Example time-courses and simulation renderings for each phase. Parameters correspond to outlined squares in (A). The activation radius is shown over time, and above each graph is shown a simulation rendering at the time of maximum disc area. Time-course was sub-sampled for clarity. See Supplemental Movie 8 for time-lapse. (D) An example of logistic growth. Population density begins at an initial density ρ_0 at $t = 0$ and increases sigmoidally at an intrinsic rate g , approaching the asymptotic carrying capacity ρ_{max} . (E) Cell density over time for the three behavioral phases. The three phases can be understood in relation to a threshold density ρ_c^{high} (dotted line) above which signaling becomes attenuated. Attenuated tissues are more dense than ρ_c^{high} , unlimited tissues are less dense than ρ_c^{high} , and self-limited tissues cross the threshold during the time-course.

does not depend on other growth parameters (Figure S5; see Methods for details of calculation). Therefore, these phases represent different collective signaling dynamics (Figure 4C and Supplemental Movie 8). Furthermore, the monotonic nature of the logistic equation (plotted in Figure 4D) helps explain the three observed phases as outcomes of three types of growth curves: density stays below ρ_c^{high} , density crosses ρ_c^{high} during the time-course, or density stays above ρ_c^{high} (Figure 4E). In particular, the attenuated propagation phenotype can be accessed above a specific initial cell density, regardless of proliferation; for the other 2 phenotypes where propagation is observed, whether the systems displays a self-limiting behavior or propagates in an unlimited fashion can be controlled via controlling proliferation even for cells plated at the same initial density. Note that a similar critical density $\rho_c^{\text{low}} = 0.30$ determines the density at which signaling turns on. In sum, we identified three dynamical behaviors *in silico* that can emerge in cells harboring the same genetic circuit, and these behaviors can be accessed by manipulating parameters of cell proliferation such as intrinsic proliferation rate g and initial cell density ρ_0 .

Growth rate-modulating drugs push Transceivers into different phase regimes

According to the computational model, decreasing the intrinsic rate of proliferation (g) should greatly extend the amount of time spent at densities permissive to signaling and therefore shift our system from a regime of self-limited propagation into a regime where propagation is virtually unlimited (from the light blue region to the dark blue region in the Figure 4A phase diagram). Conversely, increasing the proliferation rate within the self-limited region should decrease the radii of propagation foci (light blue region in Figure 4B). Finally, increasing the initial plating density (ρ_0) of the culture greater than 1250 cells/mm² (1x) should reduce the size of propagation foci.

To test these predictions *in vitro*, we performed propagation experiments in conditions that perturb cell proliferation. Cells were cultured in the presence of Y-27632 (Uehata et al., 1997), a small molecule ROCK inhibitor (RI) that limits proliferation (Kosako et al., 2000; Pitha et al., 2018; Santos et al., 2019), or fibroblast growth factor 2 (FGF2), a growth factor that stimulates proliferation (Detillieux et al., 2003; Raballo et al., 2000; Dupree et al., 2006). Sender:transceiver co-cultures were plated at 1x, 2x, and 4x densities, cell density was counted daily ($n = 3$) for seven days, and the effect on the intrinsic proliferation rate g was estimated similarly to Figure 3. Figure 5A shows growth curve data from the 1x plating density, color-coded by

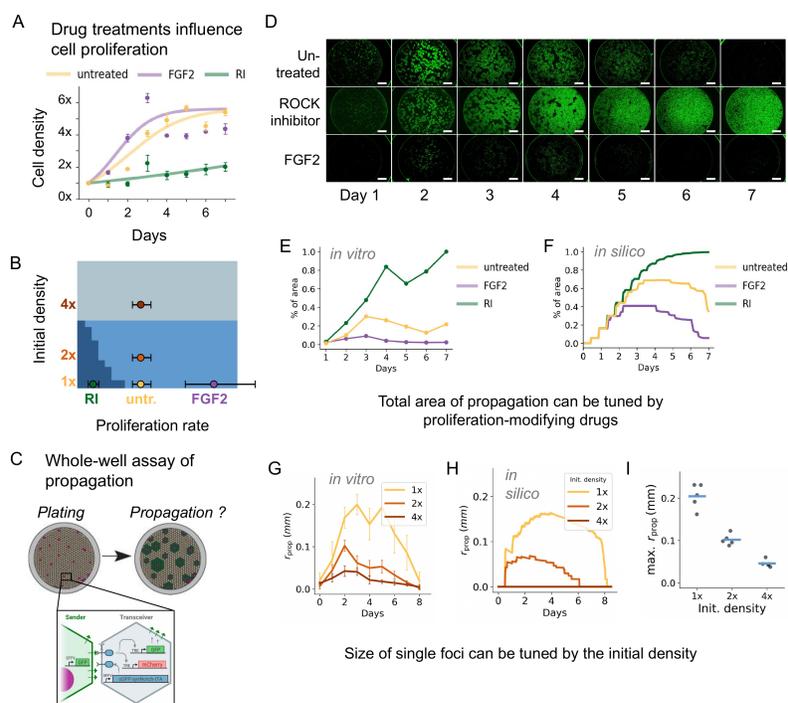


Figure 5: *In vitro* control of Transceiver activation area by manipulating either population growth rate or initial population density. (A) Time-series of cell density under growth-modulating drug conditions: Untreated (yellow), FGF2 (violet), or ROCK-inhibitor (RI; green). Cell count data (circles, mean \pm s.d.) were used to parameterize the logistic growth equation (solid lines). FGF2 induces faster population growth, while RI dramatically slows growth. "Untreated" sample reproduced from Fig. 3B. (B) Experimental perturbations alter the predicted behavior of transceivers. Phase diagram of transceiver propagation behavior as a function of proliferation rate (g) and initial cell density (ρ_0) (see Figure 4A). Circles indicate g (estimated mean and 90%CI) and ρ_0 for real *in vitro* perturbations. An increase in ρ_0 from 1x (yellow) should cause smaller foci (2x, orange) and eventually full attenuation of signaling (4x, red). Alternatively, slowing down growth should cause fast, uncontrolled propagation (low g , green) while speeding up growth should cause faster attenuation (high g , purple). (C) Schematic of the whole-well propagation assay. A co-culture of senders (purple) and transceivers (brown) is plated in a culture well at time $t = 0$. Each sender acts as a propagation focus (inset diagram), and the amount of ligand produced in the well (green) over time is measured by fluorescence imaging. This assay allows quantification of propagation without isolation of single propagation foci. (D) Time-series micrographs of 1:100 sender:transceiver co-cultures plated at 1x density (1250 cells/mm²) under various drug treatments. Senders and activated transceivers produce GFP (green). Compared to untreated, RI-treated wells propagate strongly throughout the time-course, while FGF2-treated wells propagate much weaker (days 1-4) and shut off earlier (day 5-6 instead of 7). Scale bar 1 mm. Note the well border produces a circular artifact. (E) *In vitro* and (F) simulated results for the assay in (D), shown as the percentage of the well covered by GFP fluorescence over time. (G) *In vitro* and (H) simulated propagation radii (r_{prop}) of single foci over time for different initial densities. Error bars in (G) indicate mean \pm s.d. (I) Peak propagation distance *in vitro* for different initial densities ($n = 5$, blue indicates mean).

drug treatment (colored circles, mean \pm s.d.; see Figure S9 for 2x and 4x initial densities). Treatment with 250 ng/mL FGF2 (purple) accelerated proliferation by a factor of 1.49 to 1.09 days⁻¹ (90% CI: 0.847 to 1.45 days⁻¹) compared to the untreated condition (yellow), while 50 μ M RI (green) slowed proliferation by a factor of 0.19 to 0.141 days⁻¹ (90% CI: 0.111 to 0.169 days⁻¹). The resulting best-fit growth curves are shown as solid curves in Figure 5A. These measured proliferation rates are shown in Figure 5B (circles, mean \pm 90% CI) superimposed on the phase diagram of behavior predicted by the computational model. In this model, RI treatment (green) should push the system into the dark blue region of "unlimited" propagation that does not shut down by day 8, while FGF2 treatment (purple) should remain in the light blue region of "self-limited" propagation (respectively, left and right in the X-direction in the phase diagram). Of note, neither drug was found to modulate synNotch activity *per se* (see Figure 1D and Figure S8G). See Figure S9 for full results of MLE.

To test these predictions, we devised a "whole well" assay of signal propagation. Figure 5C shows a schematic of this assay in which senders (purple) and transceivers (brown) are initially plated in a 1:100 ratio (top left diagram). Each sender (inset diagram, bottom) locally triggers a wave of GFP signaling (green), and in permissive conditions, these foci expand and fuse over time to occupy a significant percentage of the surface area of the culture well (top right diagram). This assay allows us to quantify transceiver propagation in cases where isolation of individual propagation foci is challenging, such as when propagation is so efficient that adjacent foci in the well rapidly fuse.

Using this assay, we tracked GFP fluorescence in whole culture wells over a seven-day time-course under different drug treatment conditions. Figure 5D shows that in the absence of drug treatment, transceivers initially produce GFP ligand and subsequently shut off expression by day 6-7 (first row, "untreated"), as in the single-foci experiments. In the presence of RI, however, GFP ligand propagation proceeds virtually indefinitely and fills the entire culture well (second row, "ROCK inhibitor"). Conversely, FGF2 treatment causes a more subtle activation of GFP ligand, followed rapidly by GFP depletion (third row, "FGF2"). The percentage of the well covered by GFP fluorescence is shown in Figure 5E (untreated: yellow, RI: green, FGF2: purple). We populated the morphospace also with the other combinations of proliferation rate/initial density combinations, the results follow similar trends and are shown in Fig. S10C,D.

Figure 5F shows computational simulations of these scenarios. For each experimental condition, a 150×150 hexagonal lattice was randomly seeded with senders and transceivers in a 1:100 ratio ($n = 10$ replicates) and numerically integrated over seven days using the population growth parameters fitted via MLE and a shorter cell-cell contact distance (see Methods). In the absence of treatment (yellow), a majority of the lattice quickly activates, followed by a gradual decline. In the presence of RI (green), propagation instead persists throughout the time-course, eventually saturating the lattice. In contrast, FGF2 treatment (purple) speeds up the progression from propagation to attenuation, causing lower levels of activation. Thus, our model corroborates the experimental evidence and demonstrates that the effects of RI and FGF2 treatment on cell proliferation (inhibition and activation, respectively) are sufficient to explain the starkly divergent outcomes of signaling activation between these conditions. We also note that the experimental propagation area was found to be smaller than predicted by modeling (contrast the peak activation in the Untreated and FGF2 conditions in Figure 5E and Figure 5F), even with a shorter contact distance. Such a deviation could be a result of the lower signal-to-noise ratio when imaging at low magnification, where low levels of fluorescence are more difficult to discern and isolate from background. Additionally, drug treatments appear to induce changes in signaling behavior within the first two days of growth, earlier than predicted by modeling (contrast the separation between curves in Figure 5E and Figure 5F). In addition to altering population growth, is therefore these drugs may affect other factors, such as the production rate of ligand and/or the process of fusion between adjacent signaling foci.

Alternatively, the Y axis of the phase diagram can be traversed by changing the initial cell density. Single-foci propagation assays were performed as described above, *in vitro* ($n = 5$) and *in silico*, starting from initial densities of 1x, 2x, or 4x confluent cell density. Figures 5G-I show that the behavior of the propagation follows the predictions of the model. In Figure 5G (*in vitro*; mean \pm s.d.) and Figure 5H (*in silico*), the propagation disc is smaller and deactivates faster at an initial density of 2x. At an initial density of 4x, there is no activation *in silico* and very little activation above baseline *in vitro*. Thus, as initial density increases, the critical attenuation density is reached more quickly and there is less time for signal to spread. Due to the persistence of mCherry after transceiver deactivation (see Figure 3A), an important property of transceiver patterning is the maximum size of the activation disc before deactivation. As shown in Figure 5I, the initial density strongly determines the maximum area of the propagation disc ($n = 5$ replicates in

black; blue lines indicate means). These perturbation experiments support the idea that population density dynamics are predictive of transceiver patterning. They also demonstrate the feasibility of treating two parameters of growth — initial cell density and the intrinsic rate of cell proliferation — as control parameters of patterning.

Tissue-scale cell density gradients generate spatial signaling activation gradients and kinematic waves

In order to define distinct regions of differentiation, embryonic tissues regulate the spatial distribution of chemical morphogens (Turing, 1952; Wolpert, 1969; Ashe and Briscoe, 2006). Recently, chemical gradients have been engineered to direct spatial differentiation using exogenous morphogens (Stapornwongkul et al., 2020; Toda, McKeithan, et al., 2020). Here, we instead set out to engineer spatial information in a non-genetic fashion in the form of cell density gradients, which can be decoded by transceivers into distinct spatial domains.

We hypothesized that a spatial gradient of cell density can elicit a gradient of GFP ligand. In Figure 6A, a phase diagram at an early time-point (left, computed as above at $t = 2.7$ days) shows that the initial cell density primarily determines whether a region of tissue is signaling or not signaling. Thus, we reason that a culture well inoculated with senders and transceivers in a gradient of cell density (top right) should become activated in areas where density is within the optimal range ($\rho_c^{\text{low}} < \rho < \rho_c^{\text{high}}$) and stay inactive in areas where density is above the optimal range ($\rho > \rho_c^{\text{high}}$). Figure 6A, bottom right, shows a computational example where an initial density gradient causes a gradient in the predicted amount of GFP ($[\text{GFP}]_{\text{ss}}$, shown in green; see Methods for details of calculation).

To test this prediction, we established an *in vitro* system where a 1:100 mixture of senders and transceivers was seeded in a tissue culture well. Based on the number of cells seeded, the density was 2x confluence (2500 cells/mm²) on average, but the initial seeding was biased towards one end of the well, forming a gradient of cell density. We then imaged the constitutive blue reporter (tagBFP) and the GFP ligand as a proxy of cell density and signaling activation, respectively. As shown in Figure 6B, at 2.7 days of culture the GFP output of the transceivers is patterned along the well in a way that recapitulate the cell density gradient. The yellow boxes indicate the region used to quantify the BFP and GFP gradients shown in 6C. Thus, spatial patterns of gene expression can be established with this circuit given the cell-density dependency, via establishing cell density gradients.

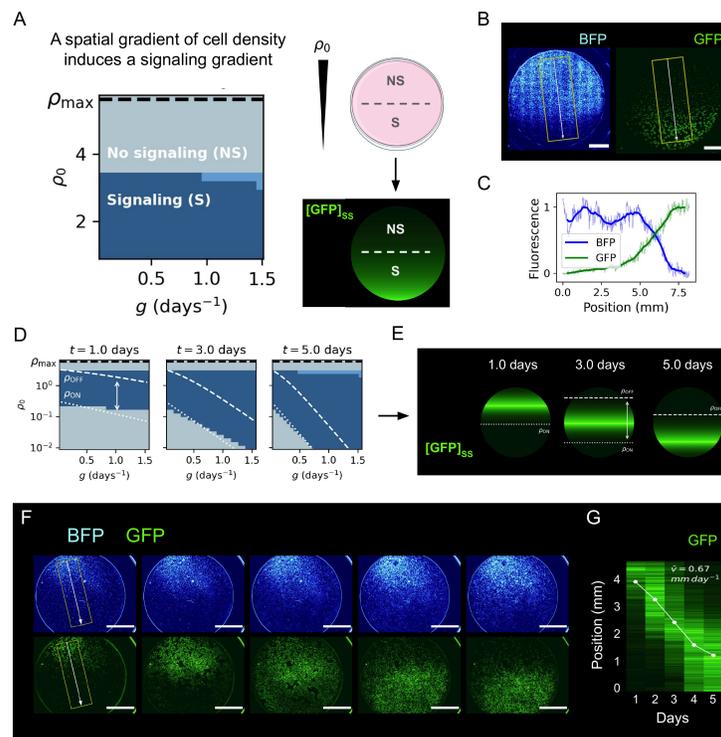


Figure 6: Spatial gradients of cell density produce long-range activation gradients and kinematic waves. (A) Modeling suggests that at early time-points, transceiver activation is determined primarily by the initial density ρ_0 (left, $t = 2.7$ days). Spatial variation in signaling, therefore, could be achieved by biasing the initial seeding of the cell culture well to create a spatial gradient of density (top right). If the the top half of the well is too dense for signaling while the bottom half is optimal, distinct regions of GFP signaling/no signaling should develop (bottom right). Green represents $[GFP]_{SS}$. (B-C) *In vitro* density and signaling gradients. (B) Stitched epifluorescence micrographs of a culture well inoculated with a sender:transceiver co-culture (1:100 ratio) in a spatial gradient of initial density (average of 2x) and imaged at 64 hrs (2.7 days) of culture. Scale bar 2 mm. TagBFP is constitutively expressed by Transceivers and acts as a cell density readout. Boxed region quantified in (C). (C) Fluorescence profiles showing the anti-correlated BFP and GFP gradients. (D-E) *In silico* modeling predicts long-range kinematic waves over time. (D) Phase diagrams at three time-points with logarithmic y-axes. The density range between ρ_c^{low} and ρ_c^{high} (white dotted and dashed lines) is optimal for signaling. (E) Given an initial density gradient, different areas of the well could enter/exit the optimal range at different times, creating a virtual (kinematic) wave of activation. Green is $[GFP]_{SS}$. (F-G) A synthetic kinematic wave generated by a density gradient *in vitro*. (F) Epifluorescence micrographs of a culture well seeded with 1:100 senders:transceivers in a spatial gradient of density (average of 1x) and imaged daily. From days 1-4, a wave spreads downwards as different regions become activated. Scale bar 2 mm. The boxed region is quantified in (G). See Supplemental Movies 9-12 for time-lapse movie of this dataset, and other examples of similar results. (G) Spatial profile of GFP fluorescence in (F) over time. White dots show the mean wavefront position, which has a mean velocity of 0.67 mm/day.

As with uniform cell density, we reason that gradients of cell density can produce rich signaling behavior over time due to the dynamics of cell population growth. In particular, because signaling activation occurs in our system within an optimal range of density (see 1E), we hypothesized that at lower densities, a spatial gradient of cell density could cause a virtual wave of activation across the well. Due to population growth, a region of the culture well that begins the time-course too sparse for activation ($\rho < \rho_c^{\text{low}}$) over time will enter the optimal range ($\rho_c^{\text{low}} < \rho < \rho_c^{\text{high}}$) and eventually exit this range ($\rho > \rho_c^{\text{high}}$). In Figure 6D phase diagrams at three consecutive time-points of simulation are annotated with the critical densities ρ_c^{low} and ρ_c^{high} , plotted respectively as white dotted and dashed lines. Over time, regions of the well with different initial densities (shown on a logarithmic Y-axis) should enter the optimal range between the two lines at different times. As shown using an example gradient in Figure 6E, this staggered activation creates the appearance of a large wave spreading through the well. However, this wave is in fact only virtual, or "kinematic," meaning each region is turning on and off independently based on its local cell density.

Our results in Figures 6F and 6G demonstrate a kinematic wave generated by an *in vitro* density gradient of sender and transceiver cells (1:100 ratio). The upper region of the well in Figure 6F starts at an efficient density for signaling but eventually becomes too dense. In contrast, the bottom region of the well is initially too sparse for signaling but enters the optimal range of density after 3-4 days in culture. Although these regions are many millimeters apart (scale bar 2mm), their entry and exit into the optimal range are staggered, creating the appearance of a wave moving downwards (see also Supplemental Movies 9-12 for time-lapse movie of this dataset, and other examples of similar results). The data in the yellow box were used to calculate the mean wavefront position as shown in Figure 6G, and the mean wave velocity was found to be 0.67 mm/day. This speed is 5.1 times faster than the speed of direct cell-to-cell propagation measured at 1x confluent density (see Figure 2H) and roughly twice the speed of directed fibroblast motility (Cornwell and Pins, 2010), making signal transduction and bulk transport unlikely causes of this phenomenon. Collectively, these results show that we can create spatial and spatio-temporal gradients of signal activation by generating gradients of cell density and exploiting the dynamics of population growth.

Discussion

The astonishing diversity of tissue patterning and morphology in our own bodies, compared to the relative uniformity of the starting material (fertilized oocyte), underscore a challenge of how the genetic circuits for patterning achieve precision in an ever-changing mechanical environment. In particular, patterning of cell growth (proliferation) generates complex inhomogeneities in cell density over space and time in all growing tissues. This observation provokes the fundamental question of whether and how mechano-chemical circuits expand achievable patterning behaviors, compared to chemical only circuits, to achieve some of the organization that we see in mammalian multicellular tissue shapes and patterns.

Here, we discovered a new effect of cell density on synNotch, which was previously unknown (Figure 1), and use it to build and study a signal-propagation genetic patterning circuit that can be influenced via cell density. The coupling between density and signal propagation means, in practice, that the physical state of a growing multicellular structure switches the synNotch circuit between propagating and non-propagating phases. This simple circuit displays many of the hallmarks of natural patterning systems that operate during development in mechanically non-uniform contexts. We observed behaviors such as self-limiting propagation (Figures 2 and 3), controlled propagation area and speed (Figures 4 and 5), which we could control by modulating growth of cells harboring the same identical simple genetic circuit. Even spatial phenomena like gradients and propagating waves could be obtained by generating spatially non-homogeneous cell densities, all with this only one simple genetic circuit (Figure 6). Similar control principles (signaling controlled by density) may have played a role in the evolution of natural patterning systems, allowing the control of minimal genetic patterning circuits via non-genetic mechanisms to generate increased variety of tissue patterning. And this type of control could be at play in present-day natural systems as well.

A remarkable feature of this circuit and its density-dependency is its close agreement with predictions from a simple computational systems model. Despite a lack of precise quantitative parameter values for many molecular interactions, the qualitative behaviors possible with this density-dependent multicellular circuit can be enumerated and explained from simple properties of the components and their interactions. More precise measurements could help to explain, eliminate, or exploit subtle differences between *in silico* and *in vitro* results. Improved model accuracy, together with finer control over cell proliferation and spatial distribution *in vitro*, would in

turn enable rapid prototyping of complex spatiotemporal patterning circuits.

Future work can explore the impact of physical variables on other synthetic signaling pathways and circuit, and also produce multicellular designs that explicitly exploit physical-chemical coupling to control synthetic morphogenesis using synthetic gene circuit dynamics that can sense and respond to changes in the physical state of a multicellular structure potentially responding to features including shape, size, and mechanical stress in addition to density. Finally, closing the loop with the output of patterning circuits being a change in mechanical features of the multicellular structure would enable closed-loop control of patterning and morphogenesis. Mechano-chemical coupling phenomena provide a route towards constructing synthetic circuits that can modulate progression through morphogenesis in a stepwise fashion, for example, executing new gene expression programs sequentially following the completion of a morphogenetic program, towards the construction and control of circuits of complexity similar to the ones observed *in vivo*.

Methods

Constructs

Constructs Design

The pHR_SFFV_LaG17_synNotch_TetRVP64 (#79128). The pHR_SFFV_GFPligand (#79129) were provided by AddGene, whereas the pHR_TRE3G_mCherry_PGK_BFP was obtained as described in the original syn-Notch paper (Morsut et al., 2016). The rest of the constructs were cloned via In-Fusion cloning (Clontech #ST0345). Specifically, the plasmids used for engineering fibroblasts were cloned in the pHR plasmid for lentivirus production.

Lentivirus Production

Lentivirus was produced by co-transfecting the transfer plasmids (pHR) and vectors encoding packaging proteins (pMD2.G and pCMV-dR8.91). Plasmids were transfected by lipofectamine LTX transfection reagent (ThermoFisher Scientific) in HEK293-T cells plated the day before in 6-well plates at approximately 70% confluence (800000 cells/well). Supernatant containing viral particles was collected 2 days after transfection and filtered to eliminate death cells and cellular debris (cut-off 0.45 μm).

Cell Lines

L929 mouse fibroblast cells (ATCC# CCL-1) and HEK293 cells were cultured in DMEM (Invitrogen) containing 10% fetal bovine serum (Laguna Scientific) and tetracycline (100 ng/ml) when requested. We generated engineered cell lines that we call "Senders" in the manuscript. The L929 Senders were obtained by transduced to stably express surface GFP (GFP fused to the PDGFR transmembrane domain, Addgene construct #79129) and a nuclear infrared fluorescent marker (H2B-miRFP703, Addgene plasmid #112853). We generated engineered cell lines that we call "Transceivers" in the manuscript. L929 transceivers were obtained via transduction with three different virions. The first virion constitutively expresses an anti-GFP antibody (Lag17) fused to a syn-Notch receptor with the transcription factor tetracycline Trans-Activator (tTA) as its intracellular domain. This receptor harbors a myc-tag on its extracellular domain that can be visualized by immunostaining (Addgene #79128). A second virion expresses both the mCherry reporter and BFP, respectively under control of the Tetracycline responsive element (TRE-3G) promoter (cloned with the Infusion kit from the Addgene plasmid #133805) and the

constitutive PGK promoter (cloned with Infusion from Addgene plasmid #79120). The third virion expresses surface GFP ligand under transcriptional control of the TRE-3G promoter. "Receiver" cells were obtained from the Lim Lab, and were produced as described in Morsut et al., Cell 2016. Briefly, they contain the LaG17 synNotch TetRVP64 (#79128) anti-GFP synNotch receptor, and a TRE→mCherry reporter for synNotch induction visualization.

All cell cultures were maintained in an incubator at 37% humidity with 5% CO₂. For viral transduction, cells were plated in 6-well dishes to achieve approximately 10% confluence at the time of infection. For lentiviral transduction, 10–100 ml of each virus supernatant was added directly to cells, with 1 μ l of polybrene (Millipore Sigma) also added to increase infection efficiency. Viral media was replaced with normal growth media 48 hours post-infection. Cells were sorted for co-expression of each component of the pathways via fluorescence-activated cell sorting (FACS) on a FACS Aria2 (Beckton-Dickinson) and by staining for the appropriate myc-tag with fluorescence-tagged antibody where needed. A bulk-sorted population consisting of fluorescence-positive cells was established for "sender" cells. For single-cell clonal population establishment of transceivers, single cells were sorted by FACS into 96-well plates starting from populations of cells infected with lentiviral particles for the relevant expression constructs. After sorting, monoclonal population were expanded and screened for the activation of the GFP ligand after stimulation with anti-myc antibodies (Cell Signaling Technology) bound to an A/G plate (Thermo Scientific).

Experiments

Signaling modulation assay

Senders and receivers L929 cells were co-cultured in a 1:1 ratio in DMEM + 10% for 24h. Then, cells were detached and analyzed by FACS Aria2 (Beckton-Dickinson). We modulated one parameter only in each experiment. The experiments where ECM composition is modulated (fibronectin, matrigel, gelatin), the procedure used Fibronectin (Corning via VWR, VWR catalog #47743-728) was used without dilution and incubated for 1 hour at 37°C; matrigel (Corning via VWR, VWR catalog #47743-720) 10 μ l of Matrigel were diluted in 1ml of DMEM in ice and then incubated at 37°C for 45 minutes for the coating; gelatin (Sigma-Aldrich, catalog #G1890) at is provided ready to use and incubated for 1h at room temperature to prepare the plate before cell seeding at 1X density. For the stiffness modulation,

were performed at 1x density on a commercial plate with FN coating on the silicone bottom (CytoSoft 6-well Plates, Advanced Biomatrix). The set of experiment done in the presence of cytoskeletal modulators, the drugs Y-27632 to inhibit Rock (Stem Cell Technologies, #72304), used at 100 μ M; latrunculin (Sigma-Aldrich, Catalog #L5163) at 200 μ M; blebbistatin (Sigma-Aldrich Catalog #B0560) at 25 μ g/ml. All the drug treatments were done on tissue treated plastic dishes at 1X plating density. The set of experiments with modulated cell density at plating were done on standard tissue culture plastic, at the indicated densities, where 1X=121250 cells/cm², 2X=242500 c/cm², 4X= 485000 c/cm².

Mechanism experiments

For indications of volume (S3B), we measure the parameter "FSC" given as output from the FACS machine (BD), which is proportional to the volume of the cells. To obtain that value per cells, sender + receiver cells were cultured for 24h at the indicated densities on tissue culture plastic, detached by trypsinization, and run as single cells in a FACS machine FACSAria2 (Beckton-Dickinson). Similarly, to measure GFP ligand via FACS (S3D), we cultivated Sender + receiver cells for 24h at the indicated densities, detached the cells via trypsinization, and recorded the GFP channel output per cell from a FACS machine FACSAria2 (Beckton-Dickinson). To measure the GFP aggregates via microscopy (S3E and F), we did as follow: we plated sender + transceiver cells at 1:100 density at the indicated densities of 1X, 2X, or 4X. We then fixed the cells with 4% PFA for 10 minutes, and performed immunostaining for anti-GFP synNotch receptor with an anti-myc antibody Alexa 647-conjugated (Cell Signaling Technology, Catalog #2233S). Fluorescent microscopy images were captured with a Keyence digital microscope machine. To measure the aggregates surface area, the green fluorescent channel was isolated and analyzed by itself; a mask was obtained and the surface individual GFP aggregates was calculated in ImageJ with the function "Analyze Particles". For analysis of YAP localization (S3J), L929 parental cells were plated at the indicated densities of 1X or 4X for 24h, fixed in 4%PFA and stained with an anti-YAP/TAZ mouse antibody (Santa Cruz Biotechnology, Catalog #sc-376830 AF647), and imaged after DAPI counterstaining with a Keyence digital microscope. For the conditioned media experiment (S3K), sender and transceiver cells were plated at 1X or 4X density for one day, after 24 hours the media was harvested; in another plate, sender + transceiver cells at 1:1 ratio were plated at either 1X or 4X densities; plating media was conditioned media from either 1X or 4X densities. For the nuclear marker fluorescent protein accumulation,

sender + transceiver cells were plated at different confluences and far red channel was measured via FACS after trypsinization.

Propagation experiments

The L929 Sender and Transceiver cells, previously cultured in presence of tetracycline (100 ng/ml), were co-plated at ratio of 1:1000 in absence of tetracycline. Cells were imaged by automated inverted epifluorescence microscope (Keyence BZ-X710) at a magnification of 2X for the image acquisition of the whole well and at 20X for the image acquisition of the single spots. Gradients were generated by non-uniformly plating the cell suspension on one side of the well.

Growth curves

At day 0 cells were plated in 96 well plate at different densities, with and without drug treatments. Multiple replicates were plated to have one well for each day, until day 7. After imaging that was done to evaluate the extent of the propagation, cells were detached and resuspended in 100 μ l of PBS and counted three times by automatic cell counter (Countess II FL – life technologies).

Image analysis

Processing of fluorescence images

To quantify the profiles of fluorescence intensity in microscope images, analysis was performed by Fiji-ImageJ. Background was subtracted and binary masks for fluorescent signals were generated to automatically segment propagation spots and quantify the area of fluorescence.

Cell shape analysis

We sought to quantify the change in projected cell shape. In the absence of a membrane marker, zoomed-in fields of view were selected at random from bright field images taken before FACS analysis (1:1 Sender:Receiver ratio, imaged at 24h of co-culture; see Figure S1C). Using Python, outlines of a subset of cells in the field of view were drawn manually and their area and perimeter were calculated. Both Senders and Receivers were counted and not distinguished. The circularity index (Cerchiari et al., 2015) was calculated as $c = 4\pi \frac{\text{Area}}{\text{Perimeter}^2}$. The index ranges from 0 to 1, where $c = 1$ is a perfect circle.

Inference of cell motility

Cell motility speed in Figure S3I was inferred by particle image velocimetry (PIV) applied to brightfield images. In each density condition, bright field images of three regions of interest (ROIs) were taken hourly for 160 hours using an automated inverted epifluorescence microscope (Keyence BZ-X710) at 20X magnification. Every ten frames, the flow field was measured by comparing the locations of "particles" (cell fragments) in adjacent frames (i.e. 0hr to 1hr, 10hr to 11hr, etc.). Particle velocities were then grouped within a 64 x 64 pixel window to generate a vector field of velocities. Outlier velocities were replaced with the local mean in a 3 x 3 window kernel, and the mean velocity in the ROI was measured.

Statistical analysis

Comparison of fluorescence measurements by FACS

Due to the hundreds to thousands of measurements per sample generated by flow cytometry, significance testing between sample distributions can be highly sensitive to minor technical variations. For this reason, we instead calculate the statistical likelihood that a sample originates from one of two signaling states, represented by a positive control in the presence of sender cells and a negative control in the absence of senders, both plated at 1x density. These control samples were binned and their histograms treated as reference distributions $P(x | \text{"ON"})$ and $P(x | \text{"OFF"})$ that estimate the true distributions in the "signaling ON" and "signaling OFF" states, respectively. Each non-control distribution \mathbf{x} was first binned into a histogram and regularized (add 1 to each bin to avoid division by zero). Then each mCherry fluorescence value $x_i \in \mathbf{x}$ was compared to the reference distributions to calculate the likelihood that x_i originated from either the ON distribution ($L_{\text{ON}}(x_i) = P(x_i | \text{"ON"})$) or the OFF distribution ($L_{\text{OFF}}(x_i) = P(x_i | \text{"OFF"})$). These were used to compute a likelihood ratio LR

$$\text{LR}(\mathbf{x}) = \prod_{x_i \in \mathbf{x}} \frac{P(x_i | \text{"ON"})}{P(x_i | \text{"OFF"})} \quad (3.1)$$

$$= \prod_{x_i \in \mathbf{x}} \frac{L_{\text{ON}}(x_i)}{L_{\text{OFF}}(x_i)} \quad (3.2)$$

that is a ratio of the odds that the sample originates from the ON condition rather than the OFF condition (i.e. $\text{LR}(\mathbf{x}) = 5$ indicates the sample \mathbf{x} is 5 times more likely to be ON than OFF). Because these values can be quite large for deep samples such as those in flow cytometry, the log-likelihood ratio $\text{LLR}(\mathbf{x}) = \log_{10} \text{LR}$ was

calculated. A perturbed sample k was considered inhibitory (labeled ** in Figure 1) if $\text{LR}(\mathbf{x}_k) < 1$ ($\text{LLR}(\mathbf{x}_k) < 0$) - in other words, if it was more likely OFF than ON. This procedure was found to be robust to the number of bins used for the histograms (Figure S2A).

Estimation of population growth parameters

To model the dynamics of cell density in culture, we use the logistic equation, a differential equation that describes how population density $\rho(t)$ evolves over time.

$$\dot{\rho} = g \left(1 - \frac{\rho}{\rho_{\max}} \right); \rho(t = 0) = \rho_0 \quad (3.3)$$

At low densities ($0 < \rho \ll \rho_{\max}$), there is exponential growth at the intrinsic growth rate g . As density increases further, however, population density saturates as it asymptotically approaches the carrying capacity ρ_{\max} . We then applied maximum likelihood estimation (MLE) to infer parameters of the logistic equation from the cell density measurements from growth curve experiments (see Methods). All density counts were assumed to be normally distributed about the curve with constant standard deviation σ (homoscedastic). First, the parameters g , ρ_{\max} , and σ were estimated using data from untreated co-cultures starting from three different starting densities. We then studied the effects of pharmacologic treatments by estimating g and σ in the presence of ROCK-i and FGF2. The carrying capacity ρ_{\max} was held constant in order to alleviate degeneracies during fitting. Specifically, the logistic equation becomes underdetermined when the data has near-constant population dynamics. Due to measurement noise and low time-resolution, such data can correspond equally well to either very slow growth or very low carrying capacity, which obscure the true values of ρ_{\max} and g , respectively. To resolve this degeneracy, carrying capacity (ρ_{\max}) was assumed to be constant under pharmacologic perturbation. Although these two possibilities suggest different regulatory influences on cell behavior, our model of signaling only takes into account the density itself and is agnostic to any underlying processes regulating it. Thus, either case will produce the same predicted signaling behavior.

Mathematical modeling

Dynamical equations for Receiver and Transceiver signaling

Following a procedure of dynamical modeling and dimensional analysis (see Supplementary Text), we model the dynamical response of a Transceiver cell using a

system of delay differential equations.

$$\dot{s}_i = \alpha f(t - \tau) - s_i \quad (3.4)$$

$$\dot{r}_i = \alpha f(t - \tau) - \gamma r_i \quad (3.5)$$

Here, for a given cell i , s_i and r_i denote the amounts of ligand ("signal") and reporter, respectively. The maximum protein production rate α is equal for both species after non-dimensionalization. Protein degradation time-scales for ligand and reporter are γ_s and γ_r , and the time-scale of reporter kinetics relative to ligand kinetics is represented by ρ , and a time delay τ represents the time for protein manufacture and trafficking. Activation is taken to be a non-linear Hill-like function

$$f = \frac{(\beta I_i)^p}{k^p + (\beta I_i)^p + (\varepsilon s_i)^p} \quad (3.6)$$

Here, k and p are the threshold and Hill coefficient (ultrasensitivity) of activation. A Hill-like activation function was chosen because this form can be derived from the application of mass-action kinetics to a cooperative transcription factor assembly, such as the dimerization of tTA-VP16. Given a cell i and its neighbors $j \in \mathcal{N}(i)$, the total amount of input ligand is $I_i = \sum_j w_{ij} s_j$, where $W = (w_{ij})$ is the cell-cell contact matrix described below. The parameter ε is the strength of *cis*-inhibitory interactions by which the ligand inhibits synNotch activation in the same cell, indirectly inhibiting its own production. The effective amount of transcription factor depends on the cell density, with the most efficient activation occurring at 100% confluence (1250 cells/mm², or $\rho = 1$) and less efficient activation at greater or lesser densities. We model this effect with a density sensitivity factor

$$\beta = \begin{cases} e^{-m(\rho-1)}, & \text{if } \rho \geq 1 \\ e^{-m(\frac{1}{\rho}-1)}, & \text{if } 0 < \rho < 1. \end{cases} \quad (3.7)$$

When $\rho = 1$, $\beta = 1$, and at lower or higher densities, β decays to zero in a log-symmetric fashion (i.e. doubling the density has the same effect as halving it). Table 3.1 shows the parameter values used for all simulations.

The cell-cell contact matrix

The matrix $W = (w_{ij})$ represents the extent of cell-cell contact between the signal-sending cell j and signal-receiving cell i . This matrix is used to map the amount of signal presented to each cell as a function of the amount expressed by each other

Parameter	Meaning	Value used
α	Ligand production rate when induced	3.0
k	Inducible promoter threshold	0.02
p	Cooperativity (ultrasensitivity) of promoter	2.0
ε	Strength of ligand-receptor inhibition in <i>cis</i>	1.0
τ	Time delay for ligand production	0.30 [†]
γ	Decay rate of reporter relative to ligand	0.1
m	Sensitivity of synNotch signaling to density	1.0
r_{int}	Maximum distance for cell-cell contacts	3 cell diameters*
v_{thresh}	Ligand production rate threshold for phase calculation	0.5

Table 3.1: **Modeling parameters.**

[†]Expressed in dimensionless time-units; corresponds to 0.43 times the generation time of the cell line.

* $r_{\text{int}} = 1$ was used for whole-well simulations.

cell. Accordingly, W encodes the signal-sending "flux" between each cell and thus conserves mass. Because of this property, W is a Markov matrix ($\sum_j w_{ij} = \mathbf{1}$), and thus the signaling input to a cell $I_i = \sum_j w_{ij}s_j$ is a weighted mean of the ligand expressed by its neighbors. Weights were calculated by evaluating a Gaussian kernel centered on cell i at each neighbor j , truncating the kernel at a maximum contact radius r_{int} , and normalizing to satisfy the Markov constraint. Figure S5A shows an example of these calculated weights.

Integration of delay differential equations

Equations 3.4 were integrated over time using a forward-Euler integration scheme with fixed time-step of $dt = 0.002$ (≈ 4.7 min). Time delays were represented as an integer number of time-steps and resolved using the method of steps for delay differential equations, with all chemical species set to zero expression at time $t < 0$. Starting at time $t = 0$, sender cell expression of signaling ligand is set to $S_{\text{sender}(s)}(t \geq 0) = 1.0$. Simulations, as well as data analysis and plotting, were performed using a custom Python library (see Code Availability).

Cell density and length-scale in the model

In our model, we define the area of each hexagonal cell to be equal to the average area at the given density. For instance, at a density of 1250 cells/mm² the cell area

is $1250^{-1} \text{ mm}^2 = 800 \mu\text{m}^2$. The side length of a hexagon is then calculated from the area and used to infer linear distances on the plane of the lattice (propagation distances, scale bars, etc.). To circumvent computational complexity associated with re-meshing after each cell division, density changes are modeled by re-scaling cell sizes using the area relationship above, preserving the structure of the lattice. This choice may under-estimate propagation velocity due to the contribution of cell division to signal diffusion.

Alignment of *in vitro* and *in silico* time-scales

In our chemical reaction system, the time-scale of reaction kinetics is set by the degradation rate constant of the ligand species γ_s . In practice, this rate is difficult to measure directly. Instead, we relate this time-scale to the observed *in vitro* system by assuming that ligand degradation and dilution occurs at a constant rate across all conditions and that this rate is comparable to the intrinsic growth rate of the wild-type cell line ($g \approx \gamma_s$). For example, we assume that during the exponential phase of growth ligand decay is primarily due to dilution by growth and division rather than by active proteasomal degradation. Thus, we were able to relate the time units of the computational and *in vitro* systems.

Simulation of GFP at steady-state

Steady-state concentrations of GFP ($[\text{GFP}]_{\text{SS}}$) in transceivers were measured by simulation of senders and transceivers randomly seeded on a 40 x 40 hexagonal lattice in a 1:100 ratio ($n = 5$ replicates). Cell lattices were simulated for eight days at 1000 different densities ranging from $\rho = 10^{-3}$ to $\rho = 10$, with no population growth (constant density). At the final time-point, the mean concentration of GFP in transceiver cells was used as an estimate for the steady-state. $[\text{GFP}]_{\text{SS}}$ was calculated as a nearest neighbor interpolation of these data.

Identification of critical densities

The critical densities ρ_c^{low} and ρ_c^{high} were calculated from the steady-state GFP concentration in transceivers simulated at constant densities (see above section "Simulation of GFP at steady-state"). The critical densities were taken to be the densities at which $[\text{GFP}]_{\text{SS}}$ crosses the promoter threshold k . That is, $[\text{GFP}]_{\text{SS}}(\rho_{\text{crit}}) = k$ was calculated by linear interpolation of the simulated data.

Data Availability

Supplementary data are available on the CaltechDATA repository at <https://doi.org/10.22002/q8n10-tsk03> (datasets) and <https://doi.org/10.22002/3vkth-rwr73> (videos).

Code Availability

Code used to perform data analysis, mathematical simulation, and plotting is available at <https://doi.org/10.5281/zenodo.11285554>.

Acknowledgements

We thank M. Johnson for her critical review of this manuscript and valuable feedback; J. Courte and A. Subramanian for critical scientific discussions; J. Bois and M. Elowitz for sharing teaching materials for mathematical modeling; S.O. for coding advice; and the Morsut and Thomson labs for their scientific input and support. This work was partially funded by the Human Frontier Science Program (HFSP) Organisation, funding M.S. with Long Term Fellowship (LT000469/2019-L); the National Institute of General Medicine of the NIH award number R35 GM138256; the National Science Foundation award number CBET-2034495 RECODE; Heritage Medical Research Institute; and The David and Lucile Packard Foundation. Figures 1A, 1B, 2A, 2B, and 2E created with BioRender.com.

References

- Alfonzo-Méndez, Marco A. et al. (Feb. 16, 2022). “Dual clathrin and integrin signaling systems regulate growth factor receptor activation”. In: *Nature Communications* 13.1. Number: 1 Publisher: Nature Publishing Group, p. 905. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28373-x. URL: <http://www.nature.com/articles/s41467-022-28373-x> (visited on 07/05/2022).
- Amano, Mutsuki, Masanori Nakayama, and Kozo Kaibuchi (Sept. 2010). “Rho-Kinase/ROCK: A Key Regulator of the Cytoskeleton and Cell Polarity”. In: *Cytoskeleton (Hoboken, N.j.)* 67.9, pp. 545–554. ISSN: 1949-3584. DOI: 10.1002/cm.20472. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3038199/> (visited on 09/29/2021).
- Ashe, Hilary L. and James Briscoe (Feb. 1, 2006). “The interpretation of morphogen gradients”. In: *Development* 133.3, pp. 385–394. ISSN: 0950-1991. DOI: 10.1242/dev.02238. URL: <https://doi.org/10.1242/dev.02238> (visited on 09/27/2022).
- Binshtok, Udi and David Sprinzak (Jan. 1, 2018). “Modeling the Notch response”. In: *Advances in experimental medicine and biology* 1066, pp. 79–98. ISSN: 0065-

2598. DOI: 10.1007/978-3-319-89512-3_5. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6879322/> (visited on 07/18/2020).
- Cai, Xiaomin, Kuei-Chun Wang, and Zhipeng Meng (2021). “Mechanoregulation of YAP and TAZ in Cellular Homeostasis and Disease Progression”. In: *Frontiers in Cell and Developmental Biology* 9. ISSN: 2296-634X. URL: <https://www.frontiersin.org/article/10.3389/fcell.2021.673599> (visited on 05/31/2022).
- Cerchiari, Alec E. et al. (Feb. 17, 2015). “A strategy for tissue self-organization that is robust to cellular heterogeneity and plasticity”. In: *Proceedings of the National Academy of Sciences* 112.7. Publisher: Proceedings of the National Academy of Sciences, pp. 2287–2292. DOI: 10.1073/pnas.1410776112. URL: <https://www.pnas.org/doi/10.1073/pnas.1410776112> (visited on 03/16/2022).
- Chan, Chii J., Carl-Philipp Heisenberg, and Takashi Hiragi (Sept. 25, 2017). “Co-ordination of Morphogenesis and Cell-Fate Specification in Development”. In: *Current Biology* 27.18, R1024–R1035. ISSN: 0960-9822. DOI: 10.1016/j.cub.2017.07.010. URL: <https://www.sciencedirect.com/science/article/pii/S0960982217308680> (visited on 05/31/2022).
- Chanet, Soline and Adam C. Martin (Jan. 1, 2014). “Chapter Thirteen - Mechanical Force Sensing in Tissues”. In: *Progress in Molecular Biology and Translational Science*. Ed. by Adam J. Engler and Sanjay Kumar. Vol. 126. Mechanotransduction. Academic Press, pp. 317–352. DOI: 10.1016/B978-0-12-394624-9.00013-0. URL: <https://www.sciencedirect.com/science/article/pii/B9780123946249000130> (visited on 05/31/2022).
- Chiasson-MacKenzie, Christine et al. (Sept. 1, 2018). “Merlin/ERM proteins regulate growth factor-induced macropinocytosis and receptor recycling by organizing the plasma membrane:cytoskeleton interface”. In: *Genes & Development* 32.17. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, pp. 1201–1214. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.317354.118. URL: <http://genesdev.cshlp.org/content/32/17-18/1201> (visited on 07/05/2022).
- Cohen, Roie and David Sprinzak (Oct. 2021). “Mechanical forces shaping the development of the inner ear”. In: *Biophysical Journal* 120.19, pp. 4142–4148. ISSN: 00063495. DOI: 10.1016/j.bpj.2021.06.036. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0006349521005464> (visited on 01/07/2022).
- Collier, Joanne R. et al. (Dec. 21, 1996). “Pattern Formation by Lateral Inhibition with Feedback: a Mathematical Model of Delta-Notch Intercellular Signalling”. In: *Journal of Theoretical Biology* 183.4, pp. 429–446. ISSN: 0022-5193. DOI:

- 10.1006/jtbi.1996.0233. URL: <https://www.sciencedirect.com/science/article/pii/S0022519396902337> (visited on 06/23/2022).
- Cornwell, Kevin G. and George D. Pins (Dec. 2010). “Enhanced Proliferation and Migration of Fibroblasts on the Surface of Fibroblast Growth Factor-2-Loaded Fibrin Microthreads”. In: *Tissue Engineering Part A* 16.12. Publisher: Mary Ann Liebert, Inc., publishers, pp. 3669–3677. ISSN: 1937-3341. DOI: 10.1089/ten.tea.2009.0600. URL: <http://www.liebertpub.com/doi/full/10.1089/ten.tea.2009.0600> (visited on 09/27/2022).
- Davies, Jamie (Apr. 1, 2017). “Using synthetic biology to explore principles of development”. In: *Development* 144.7, pp. 1146–1158. ISSN: 0950-1991. DOI: 10.1242/dev.144196. URL: <https://doi.org/10.1242/dev.144196> (visited on 05/31/2022).
- De Belly, Henry et al. (Feb. 4, 2021). “Membrane Tension Gates ERK-Mediated Regulation of Pluripotent Cell Fate”. In: *Cell Stem Cell* 28.2, 273–284.e6. ISSN: 1934-5909. DOI: 10.1016/j.stem.2020.10.018. URL: <https://www.sciencedirect.com/science/article/pii/S1934590920305348> (visited on 07/05/2022).
- Detillieux, Karen A. et al. (Jan. 1, 2003). “Biological activities of fibroblast growth factor-2 in the adult myocardium”. In: *Cardiovascular Research* 57.1, pp. 8–19. ISSN: 0008-6363. DOI: 10.1016/S0008-6363(02)00708-3. URL: [https://doi.org/10.1016/S0008-6363\(02\)00708-3](https://doi.org/10.1016/S0008-6363(02)00708-3) (visited on 02/07/2022).
- Dupree, Melissa A. et al. (Oct. 15, 2006). “Fibroblast Growth Factor 2 Induced Proliferation in Osteoblasts and Bone Marrow Stromal Cells: A Whole Cell Model”. In: *Biophysical Journal* 91.8, pp. 3097–3112. ISSN: 0006-3495. DOI: 10.1529/biophysj.106.087098. URL: <https://www.sciencedirect.com/science/article/pii/S0006349506720239> (visited on 02/07/2022).
- Ebrahimkhani, Mo R and Miki Ebisuya (Oct. 1, 2019). “Synthetic developmental biology: build and control multicellular systems”. In: *Current Opinion in Chemical Biology*. Synthetic Biology • Synthetic Biomolecules 52, pp. 9–15. ISSN: 1367-5931. DOI: 10.1016/j.cbpa.2019.04.006. URL: <https://www.sciencedirect.com/science/article/pii/S1367593119300468> (visited on 05/31/2022).
- Ebrahimkhani, Mo R. and Michael Levin (May 21, 2021). “Synthetic living machines: A new window on life”. In: *iScience* 24.5, p. 102505. ISSN: 2589-0042. DOI: 10.1016/j.isci.2021.102505. URL: <https://www.sciencedirect.com/science/article/pii/S2589004221004739> (visited on 05/31/2022).
- Feroze, Rafey et al. (Feb. 1, 2015). “Mechanics of blastopore closure during amphibian gastrulation”. In: *Developmental Biology* 398.1, pp. 57–67. ISSN: 0012-1606. DOI: 10.1016/j.ydbio.2014.11.011. URL: <https://www.sciencedirect.com/science/article/pii/S0012160614005880> (visited on 05/31/2022).

- Formosa-Jordan, Pau and Marta Ibañes (Apr. 29, 2014). “Competition in Notch Signaling with Cis Enriches Cell Fate Decisions”. In: *PLOS ONE* 9.4. Publisher: Public Library of Science, e95744. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0095744. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0095744> (visited on 06/22/2022).
- Formosa-Jordan, Pau, Marta Ibañes, et al. (July 1, 2012). “Regulation of neuronal differentiation at the neurogenic wavefront”. In: *Development* 139.13, pp. 2321–2329. ISSN: 0950-1991. DOI: 10.1242/dev.076406. URL: <https://doi.org/10.1242/dev.076406> (visited on 06/22/2022).
- Fukaya, Takashi (May 24, 2021). “Dynamic regulation of anterior-posterior patterning genes in living *Drosophila* embryos”. In: *Current Biology* 31.10, 2227–2236.e6. ISSN: 0960-9822. DOI: 10.1016/j.cub.2021.02.050. URL: <https://www.sciencedirect.com/science/article/pii/S0960982221002943> (visited on 05/31/2022).
- Gierer, A. and H. Meinhardt (Dec. 1972). “A theory of biological pattern formation”. In: *Kybernetik* 12.1, pp. 30–39. ISSN: 0023-5946. DOI: 10.1007/BF00289234.
- Goehring, Nathan W. and Stephan W. Grill (Feb. 1, 2013). “Cell polarity: mechanochemical patterning”. In: *Trends in Cell Biology* 23.2, pp. 72–80. ISSN: 0962-8924. DOI: 10.1016/j.tcb.2012.10.009. URL: <https://www.sciencedirect.com/science/article/pii/S0962892412002012> (visited on 05/31/2022).
- Gordon, Wendy R., Kelly L. Arnett, and Stephen C. Blacklow (Oct. 1, 2008). “The molecular logic of Notch signaling – a structural and biochemical perspective”. In: *Journal of Cell Science* 121.19, pp. 3109–3119. ISSN: 0021-9533. DOI: 10.1242/jcs.035683. URL: <https://doi.org/10.1242/jcs.035683> (visited on 09/12/2022).
- Hadjivasilou, Zena, Ginger L. Hunter, and Buzz Baum (Nov. 30, 2016). “A new mechanism for spatial pattern formation via lateral and protrusion-mediated lateral signalling”. In: *Journal of The Royal Society Interface* 13.124. Publisher: Royal Society, p. 20160484. DOI: 10.1098/rsif.2016.0484. URL: <https://royalsocietypublishing.org/doi/10.1098/rsif.2016.0484> (visited on 06/21/2022).
- Hannezo, Edouard and Carl-Philipp Heisenberg (June 27, 2019). “Mechanochemical Feedback Loops in Development and Disease”. In: *Cell* 178.1, pp. 12–25. ISSN: 0092-8674. DOI: 10.1016/j.cell.2019.05.052. URL: <https://www.sciencedirect.com/science/article/pii/S0092867419306233> (visited on 04/12/2021).
- Heisenberg, Carl-Philipp and Yohanns Bellaïche (May 23, 2013). “Forces in Tissue Morphogenesis and Patterning”. In: *Cell* 153.5, pp. 948–962. ISSN: 0092-8674. DOI: 10.1016/j.cell.2013.05.008. URL: <https://www.sciencedirect.com/science/article/pii/S0092867413005734> (visited on 01/31/2022).

- Ho, Christine and Leonardo Morsut (May 11, 2021). “Novel synthetic biology approaches for developmental systems”. In: *Stem Cell Reports* 16.5, pp. 1051–1064. ISSN: 2213-6711. DOI: 10.1016/j.stemcr.2021.04.007. URL: <https://www.sciencedirect.com/science/article/pii/S2213671121002022> (visited on 05/31/2022).
- Howard, Jonathon, Stephan W. Grill, and Justin S. Bois (June 2011). “Turing’s next steps: the mechanochemical basis of morphogenesis”. In: *Nature Reviews Molecular Cell Biology* 12.6. Number: 6 Publisher: Nature Publishing Group, pp. 392–398. ISSN: 1471-0080. DOI: 10.1038/nrm3120. URL: <https://www.nature.com/articles/nrm3120> (visited on 01/04/2021).
- Hufnagel, Lars et al. (Mar. 6, 2007). “On the mechanism of wing size determination in fly development”. In: *Proceedings of the National Academy of Sciences* 104.10. Publisher: Proceedings of the National Academy of Sciences, pp. 3835–3840. DOI: 10.1073/pnas.0607134104. URL: <https://www.pnas.org/doi/10.1073/pnas.0607134104> (visited on 09/13/2022).
- Hunter, Ginger L. et al. (Feb. 11, 2019). “A role for actomyosin contractility in Notch signaling”. In: *BMC Biology* 17.1, p. 12. ISSN: 1741-7007. DOI: 10.1186/s12915-019-0625-9. URL: <https://doi.org/10.1186/s12915-019-0625-9> (visited on 05/31/2022).
- Khamaisi, Bassma et al. (Oct. 21, 2022). “Functional Comparison between Endogenous and Synthetic Notch Systems”. In: *ACS Synthetic Biology* 11.10. Publisher: American Chemical Society, pp. 3343–3353. DOI: 10.1021/acssynbio.2c00247. URL: <https://doi.org/10.1021/acssynbio.2c00247> (visited on 02/07/2023).
- Ko, Clint S. and Adam C. Martin (Jan. 1, 2020). “Chapter Five - The cellular and molecular mechanisms that establish the mechanics of Drosophila gastrulation”. In: *Current Topics in Developmental Biology*. Ed. by Lilianna Solnica-Krezel. Vol. 136. Gastrulation: From Embryonic Pattern to Form. Academic Press, pp. 141–165. DOI: 10.1016/bs.ctdb.2019.08.003. URL: <https://www.sciencedirect.com/science/article/pii/S0070215319300729> (visited on 05/31/2022).
- Kopan, Raphael and Ma. Xenia G. Ilagan (Apr. 17, 2009). “The Canonical Notch Signaling Pathway: Unfolding the Activation Mechanism”. In: *Cell* 137.2, pp. 216–233. ISSN: 0092-8674. DOI: 10.1016/j.cell.2009.03.045. URL: <https://www.sciencedirect.com/science/article/pii/S0092867409003821> (visited on 09/12/2022).
- Kosako, Hidetaka et al. (Dec. 2000). “Rho-kinase/ROCK is involved in cytokinesis through the phosphorylation of myosin light chain and not ezrin/radixin/moesin proteins at the cleavage furrow”. In: *Oncogene* 19.52. Number: 52 Publisher: Nature Publishing Group, pp. 6059–6064. ISSN: 1476-5594. DOI: 10.1038/sj.onc.1203987. URL: <https://www.nature.com/articles/1203987> (visited on 02/07/2022).

- Lam, Calvin et al. (Apr. 15, 2022). “Parameterized Computational Framework for the Description and Design of Genetic Circuits of Morphogenesis Based on Contact-Dependent Signaling and Changes in Cell–Cell Adhesion”. In: *ACS Synthetic Biology* 11.4. Publisher: American Chemical Society, pp. 1417–1439. DOI: 10.1021/acssynbio.0c00369. URL: <https://doi.org/10.1021/acssynbio.0c00369> (visited on 09/12/2022).
- Le Roux, Anabel-Lise et al. (Aug. 19, 2019). “The plasma membrane as a mechanochemical transducer”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1779. Publisher: Royal Society, p. 20180221. DOI: 10.1098/rstb.2018.0221. URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.2018.0221> (visited on 05/31/2022).
- Lovendahl, Klaus N., Stephen C. Blacklow, and Wendy R. Gordon (2018). “The Molecular Mechanism of Notch Activation”. In: *Molecular Mechanisms of Notch Signaling*. Ed. by Tilman Borggrefe and Benedetto Daniele Giaimo. Advances in Experimental Medicine and Biology. Cham: Springer International Publishing, pp. 47–58. ISBN: 978-3-319-89512-3. DOI: 10.1007/978-3-319-89512-3_3. URL: https://doi.org/10.1007/978-3-319-89512-3_3 (visited on 09/12/2022).
- Mack, Julia J. et al. (Nov. 20, 2017). “NOTCH1 is a mechanosensor in adult arteries”. In: *Nature Communications* 8.1. Number: 1 Publisher: Nature Publishing Group, p. 1620. ISSN: 2041-1723. DOI: 10.1038/s41467-017-01741-8. URL: <http://www.nature.com/articles/s41467-017-01741-8> (visited on 05/31/2022).
- Mao, Qiyang and Thomas Lecuit (Jan. 1, 2016). “Chapter Thirty-Seven - Mechanochemical Interplay Drives Polarization in Cellular and Developmental Systems”. In: *Current Topics in Developmental Biology*. Ed. by Paul M. Wassarman. Vol. 116. Essays on Developmental Biology, Part A. Academic Press, pp. 633–657. DOI: 10.1016/bs.ctdb.2015.11.039. URL: <https://www.sciencedirect.com/science/article/pii/S0070215315002057> (visited on 05/31/2022).
- Matsuda, Mitsuhiro et al. (Apr. 17, 2012). “Synthetic Signal Propagation Through Direct Cell-Cell Interaction”. In: *Science Signaling* 5.220. Publisher: American Association for the Advancement of Science Section: Research Article, ra31–ra31. ISSN: 1945-0877, 1937-9145. DOI: 10.1126/scisignal.2002764. URL: <http://stke.sciencemag.org/content/5/220/ra31> (visited on 07/14/2021).
- Matsuo, Eri et al. (Nov. 1, 2021). “Substrate stiffness modulates endothelial cell function via the YAP-Dll4-Notch1 pathway”. In: *Experimental Cell Research* 408.1, p. 112835. ISSN: 0014-4827. DOI: 10.1016/j.yexcr.2021.112835. URL: <https://www.sciencedirect.com/science/article/pii/S001448272100389X> (visited on 05/31/2022).
- Meloty-Kapella, Laurence et al. (June 12, 2012). “Notch Ligand Endocytosis Generates Mechanical Pulling Force Dependent on Dynamin, Epsins, and Actin”. In: *Developmental Cell* 22.6, pp. 1299–1312. ISSN: 1534-5807. DOI: 10.1016/j.

- devcel.2012.04.005. URL: <https://www.sciencedirect.com/science/article/pii/S1534580712001438> (visited on 05/31/2022).
- Morsut, Leonardo et al. (Feb. 11, 2016). “Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors”. In: *Cell* 164.4. Publisher: Elsevier, pp. 780–791. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.01.012. URL: [http://www.cell.com/cell/abstract/S0092-8674\(16\)00052-0](http://www.cell.com/cell/abstract/S0092-8674(16)00052-0) (visited on 07/14/2021).
- Mulberry, Nicola and Leah Edelstein-Keshet (Nov. 19, 2020). “Self-organized multicellular structures from simple cell signaling: a computational model”. In: *Physical Biology* 17.6, p. 066003. ISSN: 1478-3975. DOI: 10.1088/1478-3975/abb2dc.
- Murray, J. D., ed. (2002). *Mathematical Biology*. Vol. 17. Interdisciplinary Applied Mathematics. New York, NY: Springer. ISBN: 978-0-387-95223-9.
- Pan, Yuanwang et al. (Nov. 8, 2016). “Differential growth triggers mechanical feedback that elevates Hippo signaling”. In: *Proceedings of the National Academy of Sciences* 113.45. Publisher: Proceedings of the National Academy of Sciences, E6974–E6983. DOI: 10.1073/pnas.1615012113. URL: <https://www.pnas.org/doi/10.1073/pnas.1615012113> (visited on 09/13/2022).
- Petrovic, Jelena et al. (June 1, 2014). “Ligand-dependent Notch signaling strength orchestrates lateral induction and lateral inhibition in the developing inner ear”. In: *Development* 141.11, p. 2313. DOI: 10.1242/dev.108100. URL: <http://dev.biologists.org/content/141/11/2313.abstract>.
- Pitha, Ian et al. (Nov. 14, 2018). “Rho-Kinase Inhibition Reduces Myofibroblast Differentiation and Proliferation of Scleral Fibroblasts Induced by Transforming Growth Factor β and Experimental Glaucoma”. In: *Translational Vision Science & Technology* 7.6, p. 6. ISSN: 2164-2591. DOI: 10.1167/tvst.7.6.6. URL: <https://doi.org/10.1167/tvst.7.6.6> (visited on 02/07/2022).
- Raballo, Rossana et al. (July 1, 2000). “Basic Fibroblast Growth Factor (Fgf2) Is Necessary for Cell Proliferation and Neurogenesis in the Developing Cerebral Cortex”. In: *Journal of Neuroscience* 20.13. Publisher: Society for Neuroscience Section: ARTICLE, pp. 5012–5023. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.20-13-05012.2000. URL: <https://www.jneurosci.org/content/20/13/5012> (visited on 02/07/2022).
- Santorelli, Marco, Calvin Lam, and Leonardo Morsut (Oct. 1, 2019). “Synthetic development: building mammalian multicellular structures with artificial genetic programs”. In: *Current Opinion in Biotechnology*. Tissue, Cell and Pathway Engineering 59, pp. 130–140. ISSN: 0958-1669. DOI: 10.1016/j.copbio.2019.03.016. URL: <https://www.sciencedirect.com/science/article/pii/S0958166918301617> (visited on 05/31/2022).

- Santos, Gabriela Leão et al. (Sept. 1, 2019). “Inhibition of Rho-associated kinases suppresses cardiac myofibroblast function in engineered connective and heart muscle tissues”. In: *Journal of Molecular and Cellular Cardiology* 134, pp. 13–28. ISSN: 0022-2828. DOI: 10.1016/j.yjmcc.2019.06.015. URL: <https://www.sciencedirect.com/science/article/pii/S0022282818312598> (visited on 02/07/2022).
- Schiffhauer, Eric S. and Douglas N. Robinson (Jan. 24, 2017). “Mechanochemical Signaling Directs Cell-Shape Change”. In: *Biophysical Journal* 112.2, pp. 207–214. ISSN: 0006-3495. DOI: 10.1016/j.bpj.2016.12.015. URL: <https://www.sciencedirect.com/science/article/pii/S0006349516343181> (visited on 05/31/2022).
- Schlissel, Gavin and Pulin Li (Oct. 6, 2020). “Synthetic Developmental Biology: Understanding Through Reconstitution”. In: *Annual Review of Cell and Developmental Biology* 36.1. Publisher: Annual Reviews, pp. 339–357. ISSN: 1081-0706. DOI: 10.1146/annurev-cellbio-020620-090650. URL: <https://doi.org/10.1146/annurev-cellbio-020620-090650> (visited on 05/31/2022).
- Scott, Lewis E., Seth H. Weinberg, and Christopher A. Lemmon (2019). “Mechanochemical Signaling of the Extracellular Matrix in Epithelial-Mesenchymal Transition”. In: *Frontiers in Cell and Developmental Biology* 7. ISSN: 2296-634X. URL: <https://www.frontiersin.org/article/10.3389/fcell.2019.00135> (visited on 05/31/2022).
- Shaya, Oren et al. (Mar. 2017). “Cell-Cell Contact Area Affects Notch Signaling and Notch-Dependent Patterning”. In: *Developmental Cell* 40.5, 505–511.e6. ISSN: 15345807. DOI: 10.1016/j.devcel.2017.02.009. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1534580717300783> (visited on 10/05/2020).
- Sjöqvist, Marika and Emma R. Andersson (Mar. 1, 2019). “Do as I say, Not(ch) as I do: Lateral control of cell fate”. In: *Developmental Biology*. Signaling pathways in development 447.1, pp. 58–70. ISSN: 0012-1606. DOI: 10.1016/j.ydbio.2017.09.032. URL: <https://www.sciencedirect.com/science/article/pii/S0012160617304967> (visited on 06/21/2022).
- Sloas, D. Christopher and John T. Ngo (May 1, 2022). *Synthetic Mechanotransduction Using Engineered SynNotch Receptors*. Section: New Results Type: article. bioRxiv, p. 2022.05.01.490205. DOI: 10.1101/2022.05.01.490205. URL: <https://www.biorxiv.org/content/10.1101/2022.05.01.490205v1> (visited on 05/31/2022).
- Spector, Ilan et al. (Feb. 4, 1983). “Latrunculins: Novel Marine Toxins That Disrupt Microfilament Organization in Cultured Cells”. In: *Science* 219.4584. Publisher: American Association for the Advancement of Science, pp. 493–495. DOI: 10.1126/science.6681676. URL: <https://www.science.org/lookup/doi/10.1126/science.6681676> (visited on 09/29/2021).

- Sprinzak, David and Stephen C. Blacklow (2021). “Biophysics of Notch Signaling”. In: *Annual Review of Biophysics* 50.1. _eprint: <https://doi.org/10.1146/annurev-biophys-101920-082204>, pp. 157–189. DOI: 10.1146/annurev-biophys-101920-082204. URL: <https://doi.org/10.1146/annurev-biophys-101920-082204> (visited on 09/12/2022).
- Sprinzak, David, Amit Lakhanpal, Lauren LeBon, et al. (June 9, 2011). “Mutual Inactivation of Notch Receptors and Ligands Facilitates Developmental Patterning”. In: *PLOS Computational Biology* 7.6. Publisher: Public Library of Science, e1002069. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002069. URL: <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002069> (visited on 06/23/2022).
- Sprinzak, David, Amit Lakhanpal, Lauren Lebon, et al. (May 6, 2010). “Cis-interactions between Notch and Delta generate mutually exclusive signalling states”. In: *Nature* 465.7294, pp. 86–90. ISSN: 1476-4687. DOI: 10.1038/nature08959.
- Stapornwongkul, Kristina S. et al. (Oct. 16, 2020). “Patterning and growth control in vivo by an engineered GFP gradient”. In: *Science* 370.6514. Publisher: American Association for the Advancement of Science Section: Research Article, pp. 321–327. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abb8205. URL: <http://science.sciencemag.org/content/370/6514/321> (visited on 07/15/2021).
- Stassen, Oscar M. J. A., Tommaso Ristori, and Cecilia M. Sahlgren (Dec. 21, 2020). “Notch in mechanotransduction – from molecular mechanosensitivity to tissue mechanostasis”. In: *Journal of Cell Science* 133.24, jcs250738. ISSN: 0021-9533. DOI: 10.1242/jcs.250738. URL: <https://doi.org/10.1242/jcs.250738> (visited on 09/12/2022).
- Straight, Aaron F. et al. (Mar. 14, 2003). “Dissecting Temporal and Spatial Control of Cytokinesis with a Myosin II Inhibitor”. In: *Science* 299.5613. Publisher: American Association for the Advancement of Science, pp. 1743–1747. DOI: 10.1126/science.1081412. URL: <https://www.science.org/lookup/doi/10.1126/science.1081412> (visited on 09/29/2021).
- Teague, Brian P., Patrick Guye, and Ron Weiss (Sept. 1, 2016). “Synthetic Morphogenesis”. In: *Cold Spring Harbor Perspectives in Biology* 8.9. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, a023929. DOI: 10.1101/cshperspect.a023929. URL: <http://cshperspectives.cshlp.org/content/8/9/a023929> (visited on 04/07/2022).
- Theodoris, Christina V. et al. (Mar. 12, 2015). “Human Disease Modeling Reveals Integrated Transcriptional and Epigenetic Mechanisms of NOTCH1 Haploinsufficiency”. In: *Cell* 160.6, pp. 1072–1086. ISSN: 0092-8674. DOI: 10.1016/j.

- cell.2015.02.035. URL: <https://www.sciencedirect.com/science/article/pii/S0092867415002019> (visited on 05/31/2022).
- Thompson, D'Arcy Wentworth (1917). *On growth and form*.
- Toda, Satoshi, Lucas R. Blauch, et al. (July 13, 2018). "Programming self-organizing multicellular structures with synthetic cell-cell signaling". In: *Science (New York, N.Y.)* 361.6398, pp. 156–162. ISSN: 1095-9203. DOI: 10.1126/science.aat0271.
- Toda, Satoshi, Nicholas W Frankel, and Wendell A Lim (Oct. 1, 2019). "Engineering cell–cell communication networks: programming multicellular behaviors". In: *Current Opinion in Chemical Biology. Synthetic Biology • Synthetic Biomolecules* 52, pp. 31–38. ISSN: 1367-5931. DOI: 10.1016/j.cbpa.2019.04.020. URL: <https://www.sciencedirect.com/science/article/pii/S1367593118302047> (visited on 05/31/2022).
- Toda, Satoshi, Wesley L. McKeithan, et al. (Oct. 16, 2020). "Engineering synthetic morphogen systems that can program multicellular patterning". In: *Science* 370.6514. Publisher: American Association for the Advancement of Science Section: Report, pp. 327–331. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abc0033. URL: <http://science.sciencemag.org/content/370/6514/327> (visited on 07/14/2021).
- Tschumperlin, Daniel J. (2011). "Mechanotransduction". In: *Comprehensive Physiology*. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cphy.c100016>. John Wiley & Sons, Ltd, pp. 1057–1073. ISBN: 978-0-470-65071-4. DOI: 10.1002/cphy.c100016. URL: <http://onlinelibrary.wiley.com/doi/abs/10.1002/cphy.c100016> (visited on 05/31/2022).
- Turing, Alan Mathison (Aug. 14, 1952). "The chemical basis of morphogenesis". In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237.641. Publisher: Royal Society, pp. 37–72. DOI: 10.1098/rstb.1952.0012. URL: <https://royalsocietypublishing.org/doi/10.1098/rstb.1952.0012> (visited on 02/18/2022).
- Uehata, Masayoshi et al. (Oct. 1997). "Calcium sensitization of smooth muscle mediated by a Rho-associated protein kinase in hypertension". In: *Nature* 389.6654. Number: 6654 Publisher: Nature Publishing Group, pp. 990–994. ISSN: 1476-4687. DOI: 10.1038/40187. URL: <https://www.nature.com/articles/40187> (visited on 02/07/2022).
- Urdy, Séverine (2012). "On the evolution of morphogenetic models: mechanochemical interactions and an integrated view of cell differentiation, growth, pattern formation and morphogenesis". In: *Biological Reviews* 87.4. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-185X.2012.00221.x>, pp. 786–803. ISSN: 1469-185X. DOI: 10.1111/j.1469-185X.2012.00221.x. URL: <http://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-185X.2012.00221.x> (visited on 05/31/2022).

- Vasilopoulos, Georgios and Kevin J. Painter (Mar. 1, 2016). “Pattern formation in discrete cell tissues under long range filopodia-based direct cell to cell contact”. In: *Mathematical Biosciences* 273, pp. 1–15. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2015.12.008. URL: <https://www.sciencedirect.com/science/article/pii/S0025556415002709> (visited on 06/22/2022).
- Veerman, Frits, Moritz Mercker, and Anna Marciniak-Czochra (Dec. 27, 2021). “Beyond Turing: far-from-equilibrium patterns and mechano-chemical feedback”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2213. Publisher: Royal Society, p. 20200278. DOI: 10.1098/rsta.2020.0278. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0278> (visited on 05/31/2022).
- Velazquez, Jeremy J. et al. (Apr. 1, 2018). “Programming Morphogenesis through Systems and Synthetic Biology”. In: *Trends in Biotechnology* 36.4, pp. 415–429. ISSN: 0167-7799. DOI: 10.1016/j.tibtech.2017.11.003. URL: <https://www.sciencedirect.com/science/article/pii/S0167779917303013> (visited on 05/31/2022).
- Weijts, Bart et al. (Dec. 14, 2018). “Blood flow-induced Notch activation and endothelial migration enable vascular remodeling in zebrafish embryos”. In: *Nature Communications* 9.1. Number: 1 Publisher: Nature Publishing Group, p. 5314. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07732-7. URL: <http://www.nature.com/articles/s41467-018-07732-7> (visited on 05/31/2022).
- Wolpert, L. (Oct. 1, 1969). “Positional information and the spatial pattern of cellular differentiation”. In: *Journal of Theoretical Biology* 25.1, pp. 1–47. ISSN: 0022-5193. DOI: 10.1016/S0022-5193(69)80016-0. URL: <https://www.sciencedirect.com/science/article/pii/S0022519369800160> (visited on 02/18/2022).
- Yarmola, Elena G. et al. (Sept. 8, 2000). “Actin-Latrunculin A Structure and Function: DIFFERENTIAL MODULATION OF ACTIN-BINDING PROTEIN FUNCTION BY LATRUNCULIN A *”. In: *Journal of Biological Chemistry* 275.36. Publisher: Elsevier, pp. 28120–28127. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M004253200. URL: [https://www.jbc.org/article/S0021-9258\(19\)65093-3/abstract](https://www.jbc.org/article/S0021-9258(19)65093-3/abstract) (visited on 09/29/2021).
- Zallen, Jennifer A. and Bob Goldstein (July 1, 2017). “Cellular mechanisms of morphogenesis”. In: *Seminars in Cell & Developmental Biology*. Extracellular Vesicles 67, pp. 101–102. ISSN: 1084-9521. DOI: 10.1016/j.semcdb.2017.06.008. URL: <https://www.sciencedirect.com/science/article/pii/S1084952117303452> (visited on 05/31/2022).
- Zarkesh, Ibrahim et al. (Apr. 12, 2022). “Synthetic developmental biology: Engineering approaches to guide multicellular organization”. In: *Stem Cell Reports* 17.4, pp. 715–733. ISSN: 2213-6711. DOI: 10.1016/j.stemcr.2022.02.004. URL: <https://www.sciencedirect.com/science/article/pii/S2213671122000935> (visited on 05/31/2022).

Supplementary Videos

Supplementary Video 1. Propagation of GFP signal by transceiver cells around a single sender cell (purple) plated at 1X confluency (1250 cells/mm²). Shown are microscope images of overlaid far red channel (purple), bright field (gray), and green channel (green) of a focus of propagation around a single sender cell through the transceiver cells (gray). Images taken every 1 hour, over a total of 1-72h. Scale bar 50um.

Supplementary Video 2. Propagation of GFP signal by transceiver cells around a single sender cell (purple) plated at 2X confluency (2500 cells/mm²). Shown are microscope images of overlaid far red channel (purple), bright field (gray), and green channel (green) of a focus of propagation around a single sender cell through the transceiver cells (gray). Images taken every 1 hour, over a total of 1-72h. Scale bar 50um.

Supplementary Video 3. Propagation of GFP signal by transceiver cells around a single sender cell (purple) plated at 4X confluency (5000 cells/mm²). Shown are microscope images of overlaid far red channel (purple), bright field (gray) and green channel (green) of a focus of propagation around a single sender cell through the transceiver cells (gray). Images taken every 1 hour, over a total of 1-72h. Scale bar 50um.

Supplementary Video 4. Renderings of model-simulated GFP expression in an 80×80 lattice of transceivers and a single sender cell (purple) at constant densities of $\rho = 1, 2,$ and 4 (1250, 2500, and 5000 cells/mm²). For details of modeling, simulation, and calculation of length scale, see Methods: Mathematical modeling.

Supplementary Video 5. Attenuation of propagation of GFP signal by transceiver cells around sender cells (purple) plated at the beginning at 1X confluency (1250 cells/mm²); shown here is a timelapse video from pictures taken every 30min, starting from 3 days after beginning of culture, for a total of 5 days. Shown are microscope images of overlaid far red channel (purple), bright field (gray), and green channel (green) of a focus of propagation around a single sender cell through the transceiver cells (gray). Scale bar 100um.

Supplementary Video 6. Attenuation of propagation of GFP signal by transceiver cells around sender cells (purple) plated at the beginning at 2X confluency (2500 cells/mm²); shown here are timelapse with pictures taken every 30min, starting from 3 days after beginning of culture, for a total of 5 days. Shown are microscope images of overlaid

far red channel (purple), bright field (gray), and green channel (green) of a focus of propagation around a single sender cell through the transceiver cells (gray). Scale bar 100um.

Supplementary Video 7. Renderings of model simulations with population growth. Increasing cell density over time was modeled by re-scaling cell size (see Methods: Mathematical modeling). Population growth dynamics were modeled by fitting a logistic growth equation to experimental cell count data (see Methods: Experiments and Methods: Statistical analysis).

Supplementary Video 8. Renderings of model simulations demonstrating three distinct modes, or phases, of transceiver signaling that emerge in a time-span of $t = 8$ days depending on growth parameters of initial density (ρ_0) and intrinsic growth rate (g); see Figure 4A. In the attenuated phase, initial density is greater than the critical density ρ_c^{high} and little or no propagation occurs. In the self-limited phase, density begins below ρ_c^{high} but crosses it during the time-course, leading to collapse of the signaling disc. In the unlimited phase, initial density is below ρ_c^{high} and the growth rate is too slow to reach ρ_c^{high} by the end of the time-course. Attenuated: $\rho_0 = 4.25$, $g = 1.2$. Self-limited: $\rho_0 = 1.0$, $g = 1.2$. Unlimited: $\rho_0 = 1.0$, $g = 0.4$.

Supplementary Video 9. Time course stitching of images taken every 24h for 5 days of the dataset shown in Main Fig. 6F. Shown is a field of view of the entire culture well; blue is a constitutive marker that tracks cell density; green is the signaling ligand produced upon activation. On the bottom, line profile graphs taken in the center of the well parallel to the direction of green signal propagation and cell density gradient of: left, blue signal; right, green signal.

Supplementary Video 10. Another example of traveling front wave due to a gradient in cell density. Shown is only the green channel in the entire well for 4 days with images taken every 24h.

Supplementary Video 11. Another example of traveling front wave due to a gradient in cell density. Shown is only the green channel in the entire well for 4 days with images taken every 24h.

Supplementary Video 12. Another example of traveling front wave due to a gradient in cell density. Shown is only the green channel in the entire well for 4 days with images taken every 24h.

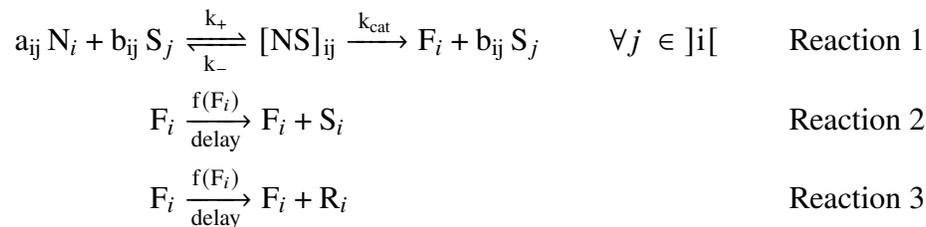
Mathematical modeling of multicellular contact-dependent signaling with synNotch

A dynamical model of synNotch-based transceiver signaling

To analyze the dynamics of synNotch signal transduction within a multicellular tissue, we seek a minimal mathematical description of the phenomenology of this system. Notch-ligand signaling has been extensively studied and modeled (Collier et al., 1996; Murray, 2002; Sprinzak et al., 2010; Formosa-Jordan et al., 2012; Shaya et al., 2017; Binshtok and Sprinzak, 2018). Building on these models, we begin with a description of the system in terms of chemical rate kinetics and a time delay for protein production. We then simplify the resulting dynamical equations under a set of stated assumptions and apply a dimensional analysis procedure to arrive at a minimal description of the system. Finally, the model is augmented with additional regulatory terms for density and regulation in *cis*.

We first consider the elementary chemical reactions involved when a cell j expressing a cell-surface Notch ligand comes in contact with another cell i expressing a (syn)Notch receptor. Subsequent to ligand-receptor binding, the receptor is endocytosed and cleaved by ADAM and γ -secretase proteases, freeing the Notch intracellular domain (NICD) then into the cytosol, where it translocates to the nucleus and acts as a transcription factor.

For our transceiver cells, which contain synNotch, the signaling ligand is GFP and the NICD contains the dimerizing transcription factor VP16-tTA, which stimulates the production of cell-surface GFP ligand and the cytosolic fluorescent reporter mCherry. This reaction system is summarized:



Here, the notation N_i represents the concentration of chemical N in cell i . Reaction 1 represents activation of Notch in cell i in *trans* by ligands on all the other cells j that are adjacent to i (denoted $]i[$). A fraction of receptors and ligands a_{ij} and b_{ij} participate in the reaction. Upon ligand-receptor binding with rate k_+ , complexes of $[NS]_{ij}$ form and can dissociate with rate k_- or trigger cleavage of N_i , producing F_i at a rate k_{cat} . We assume S_j is not degraded during *trans*-activation, although relaxing this assumption has not been shown to lead to significant differences in

Notch-ligand signaling models to date (Sprinzak et al., 2010; Khait et al., 2016). In reactions 2 and 3 the NICD F_i activates production of S_i and R_i . We model promoter activation using a Hill function $f(F_i) = F_i^p / (\kappa^p + F_i^p)$, where κ is the threshold concentration for activation and p is the Hill coefficient describing the cooperativity (ultrasensitivity) of the response, or the sharpness of the transition between on and off states.

Protein production requires many intermediate reactions such as transcription, translation, and post-translational processing that may delay the effect of transcriptional regulation on protein concentrations. We represent the delay for protein production as a single delay time T' for the signaling ligand S and reporter R . (Note that S 's dynamics do not depend on R , so the two can be given separate delays without much consequence). Using mass action kinetics, we express these reactions as a system of differential equations with time delays (DDEs)¹ in each cell i .

With the introduction of time delay terms, the time evolution equations for the elementary reactions are

$$\dot{N}_i = \beta_N + \sum_{j \in |i|} (k_- [NS]_{ij} - k_+ w_{ij} N_i S_j) - \gamma_N N_i \quad (3.8)$$

$$[N\dot{S}]_{ij} = k_+ w_{ij} N_i S_j - (k_- + k_{\text{cat}}) [NS]_{ij} \quad \forall j \in |i| \quad (3.9)$$

$$\dot{F}_i = k_{\text{cat}} \sum_{j \in |i|} [NS]_{ij} - \gamma_F F_i \quad (3.10)$$

$$\dot{S}_i = \alpha_S f(F_i(T - T')) - \gamma_S S_i \quad (3.11)$$

$$\dot{R}_i = \alpha_R f(F_i(T - T')) - \gamma_R R_i. \quad (3.12)$$

Here, T is time, and $F_i(T - T')$ indicates that the effects of gene induction are subject to a time delay T' . β_N is the basal receptor production rate, α_S and α_R are the inducible production rates of synNotch-induced ligand and reporter, and γ_F , γ_S , and γ_R are protein decay rates due to degradation and dilution. Note that we take the basal production rates of S and R (without induction) to be negligible based on the clones we selected, but basal production can easily be added to the model.

¹DDEs have been used similarly to study cellular processes with multi-step regulation including cell cycle regulation (Busenberg and Tang, 1994; Srividhya and Gopinathan, 2006), circadian oscillations (Smolen, Baxter, and Byrne, 2002), and p53 regulatory dynamics (Batchelor et al., 2008) and are well-suited for modeling nonlinear phenomena in a succinct framework (Glass, Jin, and Riedel-Kruse, 2021).

We have also defined a cell-cell contact weight $w_{ij} = a_{ij}b_{ij}$ that summarizes the amount of contact at the ij junction. The matrix $W = (w_{ij}) = W^T$ determines the signaling "flux" between cells in the lattice. In the tissue interior, conservation of mass constraints require that the total amount of ligand received in signaling should equal the total ligand presented ($\sum_i W_{ij}s_j = \sum_j s_j$), and thus W satisfies the Markov property $\sum_j w_{ij} = \mathbf{1}$. Intuitively, then, $\sum_{j \in \text{neighbors } i} w_{ij}[NS]_{ij}$ is a weighted sum over all neighbors j .² Weights were assigned based on cell-cell distance such that closer cells are engaged in stronger cell-cell contacts (see Figure S5A and refer to Methods: Mathematical Modeling).

We then assume that Notch-ligand complex association and dissociation occur on a much shorter time-scale than protein cleavage, production, or degradation and thus apply a steady-state approximation $[NS]_{ij}(T) \approx [NS]_{ij,ss}$. Solving for the steady-state, $[NS]_{ij,ss} = \frac{k_+}{k_- + k_{\text{cat}}} w_{ij} N_i S_j$. Next we apply an assumption that the cleavage turnover rate is slower than dissociation ($k_{\text{cat}} \ll k_-$) such that $[NS]_{ij,ss} \approx K_D^{-1} w_{ij} N_i S_j$, where $K_D = \frac{k_-}{k_+}$ is the dissociation constant for Reaction 1. To simplify notation, we also define the total amount of input ligand $I_i \equiv \sum_{j \in \text{neighbors } i} w_{ij} S_j$.

With these substitutions:

$$\dot{N}_i = \beta_N - \frac{k_{\text{cat}}}{K_D} N_i I_i - \gamma_N N_i \quad (3.13)$$

$$\dot{F}_i = \frac{k_{\text{cat}}}{K_D} N_i I_i - \gamma_F F_i \quad (3.14)$$

$$\dot{S}_i = \alpha_S f(F_i(T - T')) - \gamma_S S_i \quad (3.15)$$

$$\dot{R}_i = \alpha_R f(F_i(T - T')) - \gamma_R R_i \quad (3.16)$$

Next, we assume that ligand-receptor binding has a minimal effect on receptor levels relative to turnover due to degradation/dilution ($\frac{k_{\text{cat}}}{K_D} I_i \ll \gamma_N$)³. This results in the steady-state approximation $N_i(T) \approx N_{i,ss} = \frac{\beta_N}{\gamma_N}$. Furthermore, we assume that NICD is degraded quickly in the cytosol⁴ and thus we apply a separation of time scales approximation $F_i(T) \approx \frac{k_{\text{cat}} \beta_N}{K_D \gamma_N \gamma_F} I_i$. In the regime consistent with these assumptions,

²At lattice boundaries, some ligand is presented to the environment. However, we can recover the Markov property of W by including the environment as a component of the system with $s_{\text{env}} = 0$. Note that there is no material flux across cell boundaries, so no ligand is lost across this free boundary.

³Fluorescent staining of synNotch receptor during transceiver signaling suggests that receptor levels are not heavily depleted (see Figure S3E). While it is not yet clear whether synNotch levels are quite this stable during induction, we nevertheless apply this approximation for its powerful simplifying effect on the model.

⁴Degradation of synthetic derivatives of Notch has not yet been studied. However in natural

we arrive at the following modeling equations.

$$\dot{S}_i = \alpha_S f(F_i, T - T') - \gamma_S S_i \quad (3.17)$$

$$\dot{R}_i = \alpha_R f(F_i, T - T') - \gamma_R R_i \quad (3.18)$$

Promoter activation by a transcription factor (TF) is commonly modeled as a non-linear response that is weak under a certain threshold concentration of TF and strong above it, up to a saturating maximum rate.

We then apply a standard dimensional analysis procedure to represent the system in terms of simplified dimensionless quantities. Time is made dimensionless by substituting $t \equiv \gamma_S T$ and $\tau \equiv \gamma_S T'$. Similarly, concentrations are substituted as $s \equiv \frac{\gamma_S}{\beta_N} S$ and $r \equiv \frac{\alpha_S \gamma_S}{\alpha_R \beta_N} R$. We define $\alpha \equiv \frac{\alpha_S}{\beta_N}$ as the production rate of the inducible promoter and $k \equiv \frac{K_D \gamma_N \gamma_F \gamma_S}{k_{\text{cat}} \beta_N^2} \kappa$. We also define $\gamma \equiv \frac{\gamma_R}{\gamma_S}$ as the relative time-scale of reporter protein degradation kinetics.

$$\dot{s}_i = \alpha f(t - \tau) - s_i \quad (3.19)$$

$$\dot{r}_i = \alpha f(t - \tau) - \gamma r_i \quad (3.20)$$

$$f(t) = \frac{I_i^p}{k^p + I_i^p} \quad (3.21)$$

All parameter values used in simulation are supplied in Table 1 in the main text.

Density-dependent signal transduction

We model the observed dependence of synNotch signaling on cell density by considering that the input signal I is less effective at densities above or below 100% confluence (1250 cells/mm², or $\rho = 1$). We multiply I by a density sensitivity factor β that depends on density as follows.

$$\beta = \begin{cases} e^{-m(\rho-1)}, & \text{if } \rho \geq 1 \\ e^{-m(\frac{1}{\rho}-1)}, & \text{if } 0 < \rho < 1 \end{cases} \quad (3.22)$$

Therefore, the input signal is most effective ($\beta = 1$) when $\rho = 1$. As density increases above this value, β decays exponentially to zero at a rate m as $\rho \rightarrow \infty$, and

contexts, the NICD has been shown to be actively degraded by endosomal machinery (Carrieri and Dale, 2017), and transcription factors found in the cytosol generally require accessory molecules to protect them from rapid degradation (Cartwright and Helin*, 2000; Oeckinghaus and Ghosh, 2009).

similarly for decreasing density, $\lim_{\rho \rightarrow 0} \beta = 0$. Note that this function is symmetric about $\rho = 1$ on a logarithmic scale; for example, doubling the density has the same effect as halving it. Thus, the sensitivity to density in both sub- and super-confluent regimes is tuned by the same parameter m . With this density sensitivity factor, our system can be described as follows.

$$\dot{s}_i = \alpha f(t - \tau) - s_i \quad (3.23)$$

$$\dot{r}_i = \alpha f(t - \tau) - \gamma r_i \quad (3.24)$$

$$f(t) = \frac{(\beta I_i)^P}{k^P + (\beta I_i)^P} \quad (3.25)$$

Cis-inhibitory regulation

Classically, Notch receptors that are activated by a ligand in *trans* have been shown to be nevertheless inhibited by the same ligand in *cis* (Celis and Bray, 1997; Micchelli, Rulifson, and Blair, 1997; Klein, Brennan, and Arias, 1997; Jacobsen et al., 1998). It is thought that this occurs when Notch and its ligand(s) on the same cell bind and mutually inactivate each other (Sakamoto et al., 2002; Cordle et al., 2008; Sprinzak et al., 2010). We model this process by adding an inhibitory term to the denominator of the regulatory function f , tuned by a free parameter ε representing the strength of this effect. In the presence of signal ligand in *cis*, further ligand production is thus dampened. This coarse-grained approach is intended to generalize over Notch dynamics and minimize model complexity while retaining the salient features of *cis*-inhibition.

With this addition, our system is modeled as follows.

$$\dot{s}_i = \alpha f(t - \tau) - s_i \quad (3.26)$$

$$\dot{r}_i = \alpha f(t - \tau) - \gamma r_i \quad (3.27)$$

$$f(t) = \frac{(\beta I_i)^P}{k^P + (\beta I_i)^P + (\varepsilon s_i)^P} \quad (3.28)$$

In effect, *cis*-inhibition acts as a negative-feedback on signal production. Because there is a time delay τ for protein production, feedback is delayed, resulting in oscillatory behavior. In our regime, we observe dampened oscillations towards steady-state (Figure S5B).

Supplementary Figures

References

- Batchelor, Eric et al. (May 9, 2008). “Recurrent Initiation: A Mechanism for Triggering p53 Pulses in Response to DNA Damage”. In: *Molecular Cell* 30.3. Publisher: Elsevier, pp. 277–289. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2008.03.016. URL: [https://www.cell.com/molecular-cell/abstract/S1097-2765\(08\)00240-2](https://www.cell.com/molecular-cell/abstract/S1097-2765(08)00240-2) (visited on 08/27/2021).
- Binshtok, Udi and David Sprinzak (Jan. 1, 2018). “Modeling the Notch response”. In: *Advances in experimental medicine and biology* 1066, pp. 79–98. ISSN: 0065-2598. DOI: 10.1007/978-3-319-89512-3_5. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6879322/> (visited on 07/18/2020).
- Busenberg, S. and B. Tang (1994). “Mathematical models of the early embryonic cell cycle: the role of MPF activation and cyclin degradation”. In: *Journal of Mathematical Biology* 32.6, pp. 573–596. ISSN: 0303-6812. DOI: 10.1007/BF00573462.
- Carrieri, Francesca A. and Jacqueline Kim Dale (2017). “Turn It Down a Notch”. In: *Frontiers in Cell and Developmental Biology* 4. ISSN: 2296-634X. URL: <https://www.frontiersin.org/articles/10.3389/fcell.2016.00151> (visited on 09/30/2022).
- Cartwright, P. and K. Helin* (Aug. 1, 2000). “Nucleocytoplasmic shuttling of transcription factors”. In: *Cellular and Molecular Life Sciences CMLS* 57.8, pp. 1193–1206. ISSN: 1420-9071. DOI: 10.1007/PL00000759. URL: <https://doi.org/10.1007/PL00000759> (visited on 09/30/2022).
- Celis, J.F. de and S. Bray (Sept. 1, 1997). “Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the *Drosophila* wing”. In: *Development* 124.17, pp. 3241–3251. ISSN: 0950-1991. DOI: 10.1242/dev.124.17.3241. URL: <https://doi.org/10.1242/dev.124.17.3241> (visited on 09/02/2021).
- Collier, Joanne R. et al. (Dec. 21, 1996). “Pattern Formation by Lateral Inhibition with Feedback: a Mathematical Model of Delta-Notch Intercellular Signalling”. In: *Journal of Theoretical Biology* 183.4, pp. 429–446. ISSN: 0022-5193. DOI: 10.1006/jtbi.1996.0233. URL: <https://www.sciencedirect.com/science/article/pii/S0022519396902337> (visited on 06/23/2022).
- Cordle, Jemima et al. (Aug. 2008). “A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition”. In: *Nature Structural & Molecular Biology* 15.8. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 8 Primary_atype: Research Publisher: Nature Publishing Group, pp. 849–857. ISSN: 1545-9985. DOI: 10.1038/nsmb.1457. URL: <https://www.nature.com/articles/nsmb.1457> (visited on 09/02/2021).
- Formosa-Jordan, Pau et al. (July 1, 2012). “Regulation of neuronal differentiation at the neurogenic wavefront”. In: *Development* 139.13, pp. 2321–2329. ISSN: 0950-

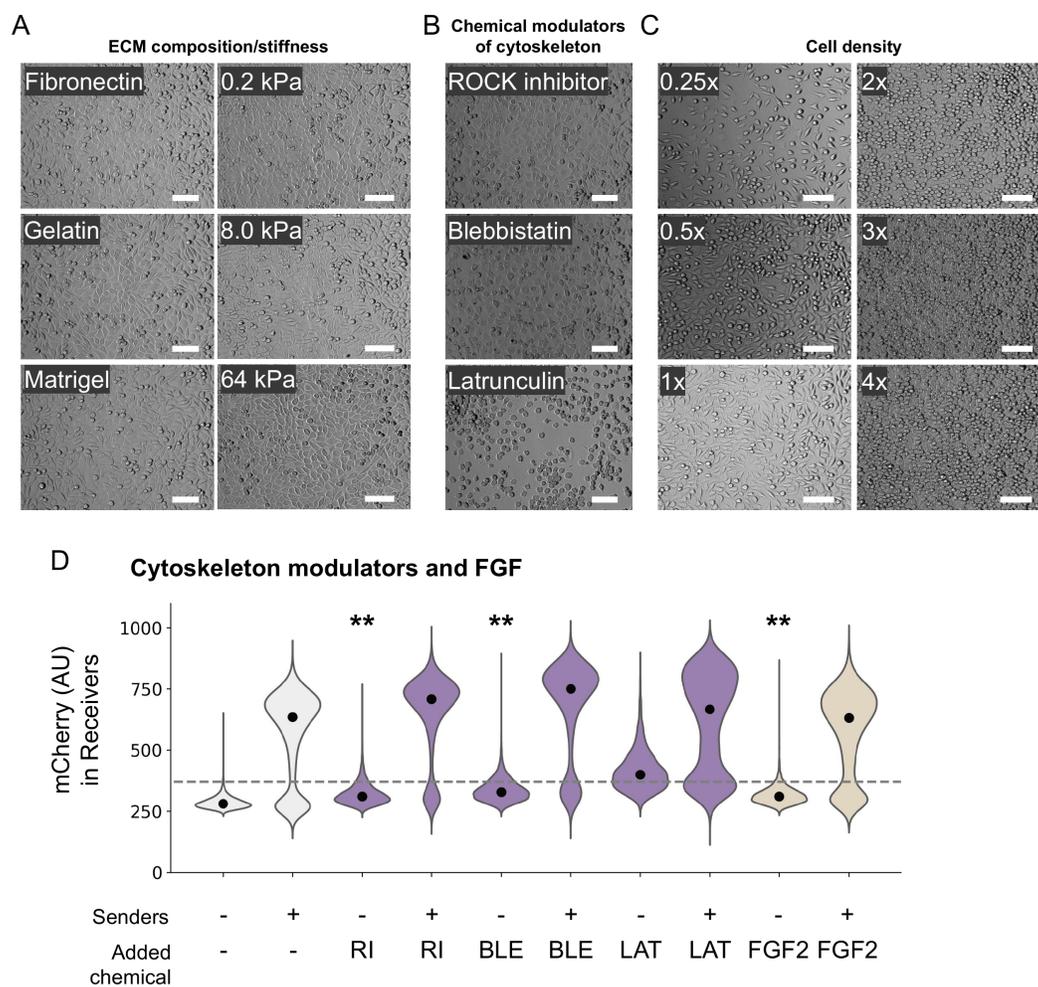


Figure S1: **Bright field imaging and additional controls for synNotch signaling assay** (A-C) Bright field images of L929 cells captured before the FACS assays reported in Figure 1. At this time point, cells were grown for 24h (A) on substrates with different compositions and stiffnesses, (B) in the presence of chemical modulators of cytoskeletal tension, and (C) starting from different cell densities at the time of plating. 1x denotes 1250 cells/cm². Scale bar 100 μ m. (D) Violin plots showing distributions of mCherry fluorescence intensity in Receivers, measured by FACS at 24h after co-culture with sender cells. Dotted line is the threshold separating high/low mCherry fluorescence. Black dot indicates median. ** indicates the sample is more likely OFF than ON, as determined by the log-likelihood ratio (LLR) statistical test. RI: ROCK-inhibitor, BLE: Blebbistatin, LAT: Latrunculin-A. FGF2: Fibroblast growth factor 2.

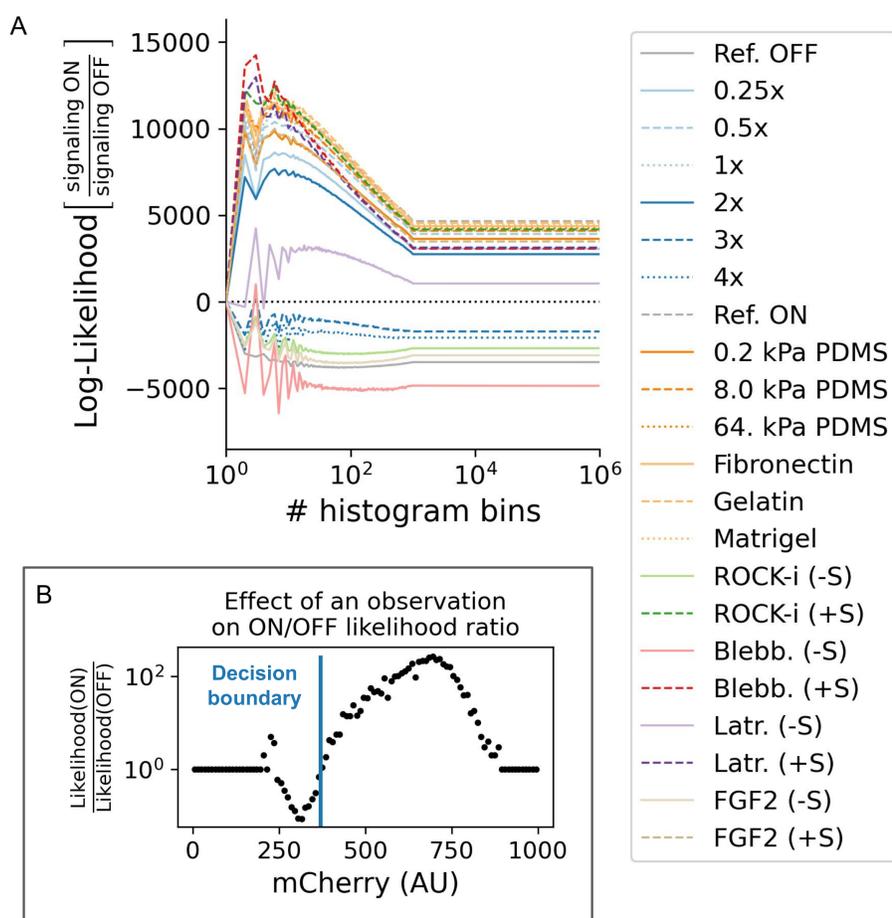


Figure S2: **Robustness of LLR statistical comparison.** After assaying the perturbed samples using the FACS-based assay for synNotch signaling, a likelihood ratio was computed to categorize the activation state of each sample as either ON or OFF. These states were exemplified by reference samples (OFF: Senders absent; ON: Senders present; both: 1x density, no ECM or additional chemicals). Reference samples were binned and regularized (1 observation added to each bin) to yield continuous empirical probability distributions corresponding to each state. Then, for each perturbed sample, the log-likelihood ratio (LLR) assigns each sample to its closest reference distribution. (A) LLR is robust to the binning procedure used for reference distributions. Curves show how the LLR of each FACS sample varies as the number of bins increases from 1 to 1,000. A point above the x -axis (horizontal dotted line) is more likely to be ON, while a point below is more likely to be OFF. With more than 4 bins, the number of bins does not affect the ON/OFF categorization by LLR. 1000 bins were used for all other figures. (B) To determine the ON/OFF threshold ("decision boundary") for a single cell, the LR was calculated for a hypothetical cell observed in each bin, and the threshold (370) was found where this curve crosses from OFF to ON.

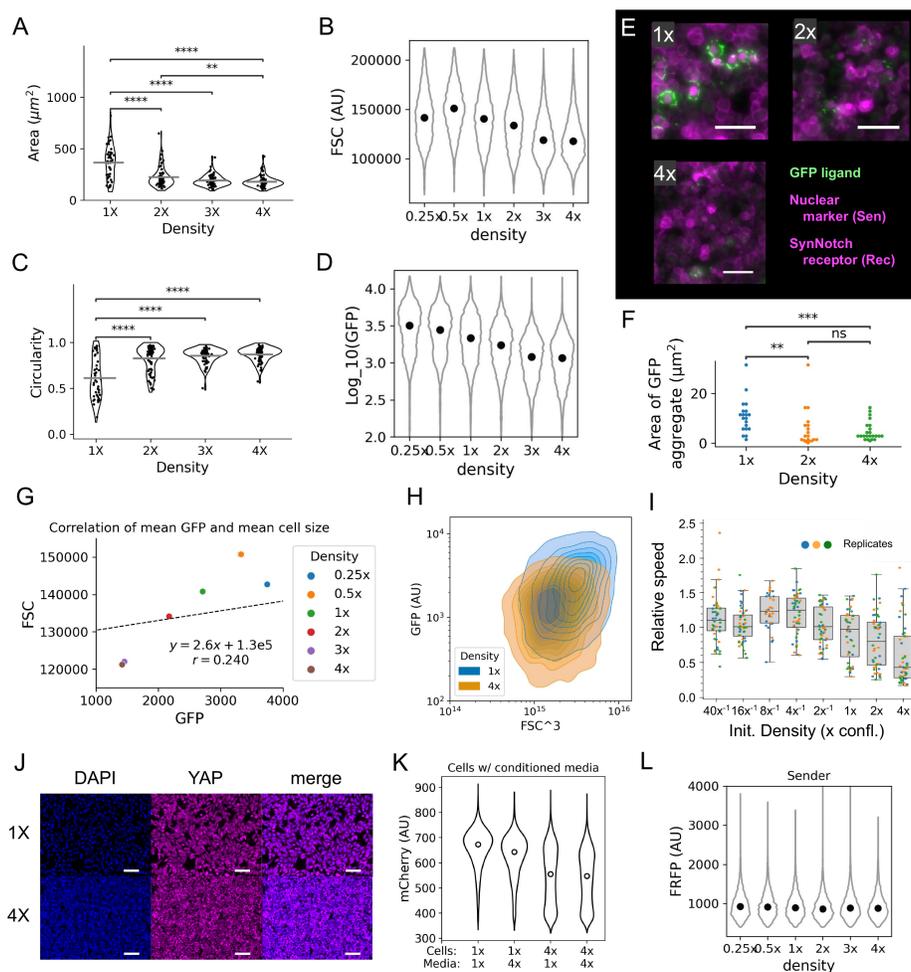


Figure S3: Mechanism of density-dependency synNotch signaling. (A) Violin plots showing the distributions of the 2D projected cell area at different cell densities. Sender:receiver co-cultures were imaged 24h after plating (images shown in Figure SS1C), and the 2D projected area of individual cells was measured from bright field images (black circles are cells; gray line indicates mean) as described in Methods: Image analysis. As density increases above 1x (1250 cells/mm²) the surface area occupied per cell decreases. **** $p < 10^{-4}$ by Mann-Whitney U test. (B) FACS plots rendered as violin plots of Forward Scatter (FSC, proxy for cell volume) of sender cells at 24h coming from co-cultures of sender/receivers at the indicated densities. Black dot indicates median. This shows that cell volume decreases at higher plating densities. (C) Violin plots showing distributions of circularity of single cells, using the same image data in (A) (black circles are cells; gray line indicates mean). Circularity scales from 0 (irregularly shaped) to 1 (perfectly circular), as described in Methods: Image analysis. This shows that cells from higher confluency cultures are more circular than the ones coming from lower densities. **** $p < 10^{-4}$, ** $p < 10^{-2}$ by Mann-Whitney U test. (Continued on next page.)

Figure S3: **(Continued from previous page)** (D) FACS fluorescence distributions of green channel intensity in GFP-ligand producing sender cells at 24h coming from co-cultures of sender/receivers at the indicated densities. Black dot indicates median. This shows that cells from higher confluency cultures have less GFP intensity. (E) Sample micrograph pictures of sender:receivers 1:50 co-cultures at the indicated densities at 24h; magenta marks both nuclei in sender cells, and synNotch receptor in receiver cells. Overlaid green channel marks GFP ligand. Scale bar 50 μm . (F) Swarm plot of the size of GFP aggregates on sender membranes in sender-transceiver co-cultures plated at the indicated densities and imaged at 24hr. Ligand aggregates are significantly smaller at 2x and 4x density than at 1x. Each dot is an aggregate. 62 aggregates were counted from 5 sender cells in each condition. *** $p < 10^{-3}$, ** $p < 10^{-2}$, and ns $p > 0.1$ by Mann-Whitney U test. (G) contour plot of FACS data depicting green channel intensity and FSC (size) of cell distributions of sender cells from sender/receiver co-cultures at 24h after seeding at the different densities indicated via the color (blue 1x, orange 4x); this shows a trend whereby both cell size and GFP intensity decreases at higher confluence. (H) Scatter plot of sender cell FSC (cell size) and GFP (green channel) fluorescence, color-coded by density condition. Mean \pm standard error of the mean. Linear correlation analysis shows FSC is positively correlated with GFP (Pearson's $r = 0.24$). (I) Box-and-whisker plots of average cell motility speed (units are arbitrary) at different plating densities of L929 sender/receiver co-cultures over 160 hours of co-culture, capturing 1 frame/hour. Every 10 frames, the velocity field in a high-magnification field of view was estimated using particle image velocimetry (PIV, described in Methods: Image analysis). Each dot is the average velocity of the field, and time-lapses for three fields of view in the same culture well were used (shown in different colors). Central line is median. Box encapsulates 1st quartiles. Whiskers denote minimum and maximum range after removal of outliers by the IQR method. This shows that cell motility trends towards decreasing at higher cell densities. (J) Sample micrograph pictures of parental L929 cells at the indicated densities at 24h stained with blue nuclear marker (DAPI) and anti-YAP antibody (magenta), shown alongside a merge of the 2 channels. Scale Bar is 50 μm . This shows that YAP localization does not dramatically changes in these cells at 1X confluency vs 4X confluency. (K) FACS plots rendered as violin plots of red channel intensity in receiver cells at 24h coming from co-cultures of sender/receivers at the indicated densities, and treated with the indicated conditioned media. The conditioned media was obtained from cultures grown at the indicated densities for 24h, as detailed in the methods. Black dots are the median of the distribution. (L) FACS plots rendered as violin plots of nuclear marker intensity in sender cells (FRFP, far red channel) at 24h coming from co-cultures of sender/receivers at the indicated densities. Black dots indicate medians.

1991. DOI: 10.1242/dev.076406. URL: <https://doi.org/10.1242/dev.076406> (visited on 06/22/2022).
- Glass, David S., Xiaofan Jin, and Ingmar H. Riedel-Kruse (Mar. 19, 2021). “Non-linear delay differential equations and their application to modeling biological network motifs”. In: *Nature Communications* 12.1, p. 1788. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21700-8. URL: <https://www.nature.com/articles/s41467-021-21700-8> (visited on 08/27/2021).
- Jacobsen, T.L. et al. (Nov. 15, 1998). “Cis-interactions between Delta and Notch modulate neurogenic signalling in *Drosophila*”. In: *Development* 125.22, pp. 4531–4540. ISSN: 0950-1991. DOI: 10.1242/dev.125.22.4531. URL: <https://doi.org/10.1242/dev.125.22.4531> (visited on 09/02/2021).
- Khait, Itzhak et al. (Jan. 12, 2016). “Quantitative Analysis of Delta-like 1 Membrane Dynamics Elucidates the Role of Contact Geometry on Notch Signaling”. In: *Cell Reports* 14.2. Publisher: Elsevier, pp. 225–233. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2015.12.040. URL: [http://www.cell.com/cell-reports/abstract/S2211-1247\(15\)01464-3](http://www.cell.com/cell-reports/abstract/S2211-1247(15)01464-3) (visited on 01/06/2021).
- Klein, Thomas, Keith Brennan, and Alfonso Martinez Arias (Sept. 1, 1997). “An Intrinsic Dominant Negative Activity of Serrate That Is Modulated during Wing Development in *Drosophila*”. In: *Developmental Biology* 189.1, pp. 123–134. ISSN: 0012-1606. DOI: 10.1006/dbio.1997.8564. URL: <https://www.sciencedirect.com/science/article/pii/S001216069798564X> (visited on 09/02/2021).
- Micchelli, C.A., E.J. Rulifson, and S.S. Blair (Apr. 15, 1997). “The function and regulation of cut expression on the wing margin of *Drosophila*: Notch, Wingless and a dominant negative role for Delta and Serrate”. In: *Development* 124.8, pp. 1485–1495. ISSN: 0950-1991. DOI: 10.1242/dev.124.8.1485. URL: <https://doi.org/10.1242/dev.124.8.1485> (visited on 09/02/2021).
- Murray, J. D., ed. (2002). *Mathematical Biology*. Vol. 17. Interdisciplinary Applied Mathematics. New York, NY: Springer. ISBN: 978-0-387-95223-9.
- Oeckinghaus, Andrea and Sankar Ghosh (Oct. 2009). “The NF- κ B Family of Transcription Factors and Its Regulation”. In: *Cold Spring Harbor Perspectives in Biology* 1.4, a000034. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a000034. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2773619/> (visited on 09/30/2022).
- Sakamoto, Kei et al. (Jan. 15, 2002). “Intracellular Cell-Autonomous Association of Notch and Its Ligands: A Novel Mechanism of Notch Signal Modification”. In: *Developmental Biology* 241.2, pp. 313–326. ISSN: 0012-1606. DOI: 10.1006/dbio.2001.0517. URL: <https://www.sciencedirect.com/science/article/pii/S0012160601905172> (visited on 09/02/2021).

- Shaya, Oren et al. (Mar. 2017). “Cell-Cell Contact Area Affects Notch Signaling and Notch-Dependent Patterning”. In: *Developmental Cell* 40.5, 505–511.e6. ISSN: 15345807. DOI: 10.1016/j.devcel.2017.02.009. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1534580717300783> (visited on 10/05/2020).
- Smolen, Paul, Douglas A. Baxter, and John H. Byrne (Nov. 1, 2002). “A Reduced Model Clarifies the Role of Feedback Loops and Time Delays in the Drosophila Circadian Oscillator”. In: *Biophysical Journal* 83.5. Publisher: Elsevier, pp. 2349–2359. ISSN: 0006-3495. DOI: 10.1016/S0006-3495(02)75249-1. URL: [https://www.cell.com/biophysj/abstract/S0006-3495\(02\)75249-1](https://www.cell.com/biophysj/abstract/S0006-3495(02)75249-1) (visited on 08/27/2021).
- Sprinzak, David et al. (May 6, 2010). “Cis-interactions between Notch and Delta generate mutually exclusive signalling states”. In: *Nature* 465.7294, pp. 86–90. ISSN: 1476-4687. DOI: 10.1038/nature08959.
- Srividhya, J. and M. S. Gopinathan (Aug. 7, 2006). “A simple time delay model for eukaryotic cell cycle”. In: *Journal of Theoretical Biology* 241.3, pp. 617–627. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2005.12.020. URL: <https://www.sciencedirect.com/science/article/pii/S002251930500559X> (visited on 09/02/2021).

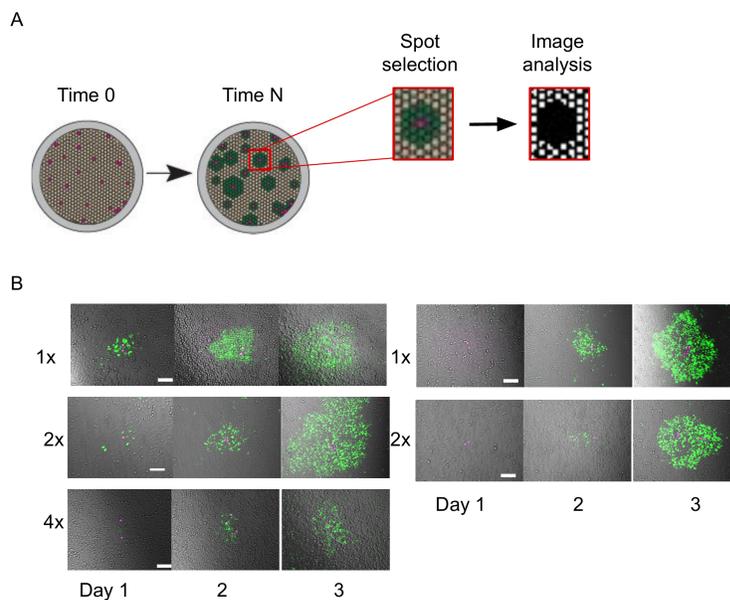


Figure S4: Examination of single propagation foci. (A) Schematic of experimental setup and image analysis pipeline for quantifying the size of individual propagation focus. In each culture well, Sender cells (purple) are sparsely seeded among Transceivers, which are initially inactive (brown) and become activated over time by signal propagation (green). Sufficiently isolated foci are identified, and these foci are imaged at multiple sequential time-points. The area of the propagation focus is obtained via thresholding of the green fluorescence channel (described in the Methods). (B) Micrograph pictures of single isolated foci of propagation at the indicated densities, captured at the indicated timepoints (day 1-3). These pictures represent replicates of Figure 2C. These and other similar propagation foci were used for calculation of propagation distance and velocity (Figures 2D and 2H). Grayscale: bright field. Green: GFP (signaling ligand and activation area). Purple: FRFP (nuclear marker in Senders). Scale bar 100 μm .

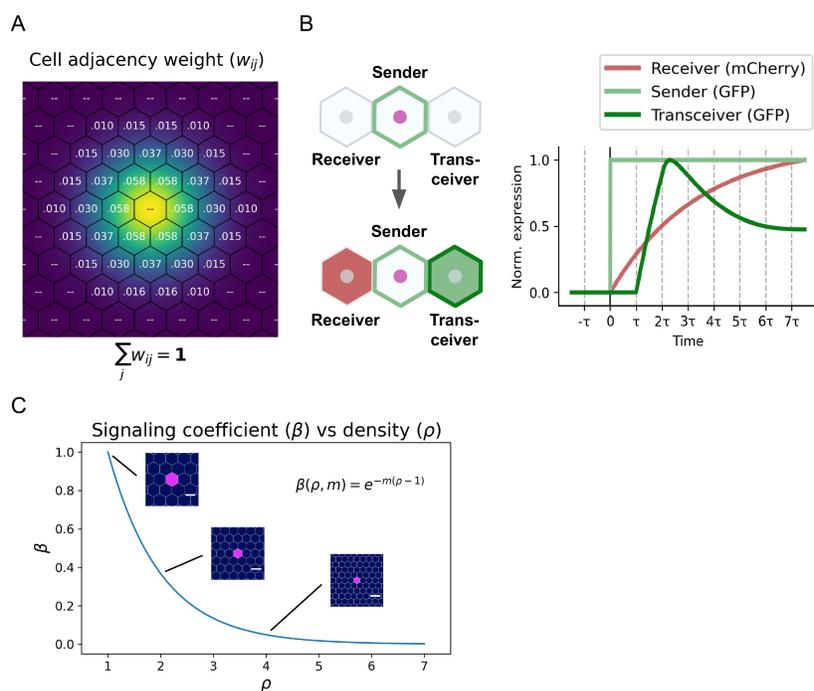


Figure S5: Mathematical modeling of cell-cell contact, signaling dynamics, and density-dependent signaling. (A) Weights of all cell-cell contacts between a given cell i (centered) and its neighbors. For each neighbor j , the weight w_{ij} (inset number) is multiplied by its ligand expression to yield the amount of ligand involved in signaling from j to i . For each cell i , its adjacency weights w_{ij} sum to 1. Weights were calculated from all pairwise cell-cell distances by applying a Gaussian kernel (depicted in color), enforcing a maximum distance, and normalizing (see Methods and Materials). For all simulations shown, the Gaussian kernel has covariance matrix $\Sigma = d_c \cdot \begin{pmatrix} 1.5 & \\ & 1.5 \end{pmatrix}$, and cell-cell adjacency was truncated above $3d_c$, where d_c is the cell-cell distance on the lattice. (B) A time-course illustrating gene expression dynamics between cell types in the model. Senders begin expressing a constant amount of ligand at time 0. While Receiver activation is monotonic, the Transceiver response is non-monotonic. Due to the time delay and negative feedback in their dynamical response, Transceivers produce an initial pulse of ligand that quickly overshoots the equilibrium concentration and then gradually approaches equilibrium. Simulation time is presented in terms of the time delay parameter τ . All quantities are normalized to highlight the difference in relative dynamics. (C) Effect of cell density on signaling efficiency in the computational model. To represent the phenomenology of impaired signal transduction at super-confluent cell density, we multiply the amount of ligand presented by the cell by a coefficient β (y-axis) that exponentially decays as a function of cell density ρ (x-axis). $\rho = 1$ corresponds to the density at confluence (1250 cells/mm^2). The parameter m tunes the sensitivity of signaling to density, and $m = 1.0$ was chosen for all shown simulations. The inset images show renderings of the cell lattice at the indicated density. Scale bar $25\mu\text{m}$. As density increases, cell area is reduced while shape and cell-cell adjacency are preserved.

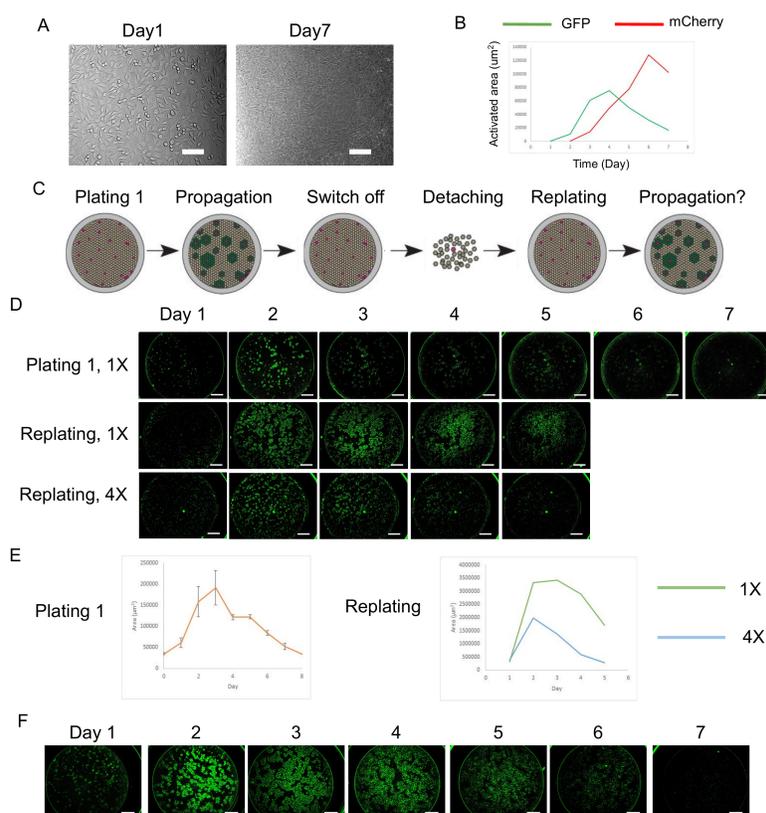


Figure S6: Density-based attenuation of signal propagation. (A) Bright field images of the culture shown in Figure 3A, at the indicated time points; shows differences in cell density at 1 day vs 7 days of culture. Scale bar 100 μm . (B) Quantification of GFP and mCherry area over time for the images in Figure 3A. GFP intensity peaks at day 4 and falls substantially by day 7. (C) Diagram description of the assay to investigate whether reducing the cell density reverses the attenuation of signaling. First, a 1:50 co-culture of Senders:Transceivers ("Plating 1") is grown at 1x density for 7 days; the propagation is allowed to occur ("Propagation" phase) and to turn off ("switch off" phase) to ensure both activation and attenuation. Then, cells are detached ("Detaching") and re-cultured at lower densities ("Replating") to assess whether they are still able to propagate the signal ("Propagation?" test). (D) First row: Daily time-course showing micrographs of green fluorescence for the Plating 1-propagation-switchOff phases. Second and third row: micrographs pictures from the Replating phase grown at 1X (second row) and at 4X (third row). After detachment, a replating at 1x retained competency for signal propagation, while a replating at 4x showed rapid attenuation after an initial activation. Scale bars 1 mm. (E) Quantifications of the area of GFP activation during Plating 1 (error bars indicate standard deviation, $n = 2$) and Replating at 1x and 4x densities. (F) Micrographs used for the quantifications of signal propagation shown in Figure 3C. Green epifluorescence was captured from a whole culture well, and propagation foci were selected and analyzed using the image analysis pipeline described in the Methods. Scale bars 1 mm.

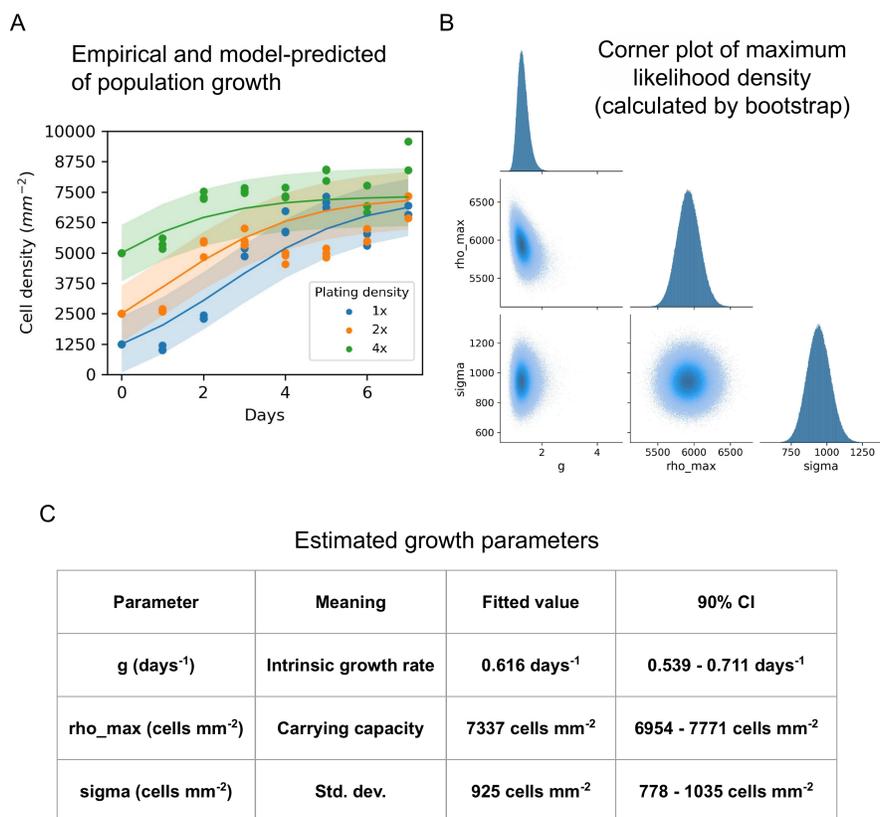


Figure S7: **Maximum likelihood estimation of population growth parameters.** A maximum-likelihood estimation (MLE) procedure was used to identify the intrinsic cell proliferation rate g , the population carrying capacity ρ_{\max} , and the standard deviation σ . (A) Sender-Transceiver co-culture density was measured daily (scattered points, $n=3$) for initial densities of 1x, 2x, and 4x. The solid lines are simulated growth curves using MLE parameters (shaded regions are 80% probability). (B) 1- and 2-D distributions of parameters ($n=10^6$ bootstrap replicates). Identified regions are well-bounded with little correlation, indicating successful fitting. Occasional high values of g (fast growth) suggest that higher time-resolution may be desired at early time-points. (C) Summary table of fitted parameters with confidence intervals from bootstrap.

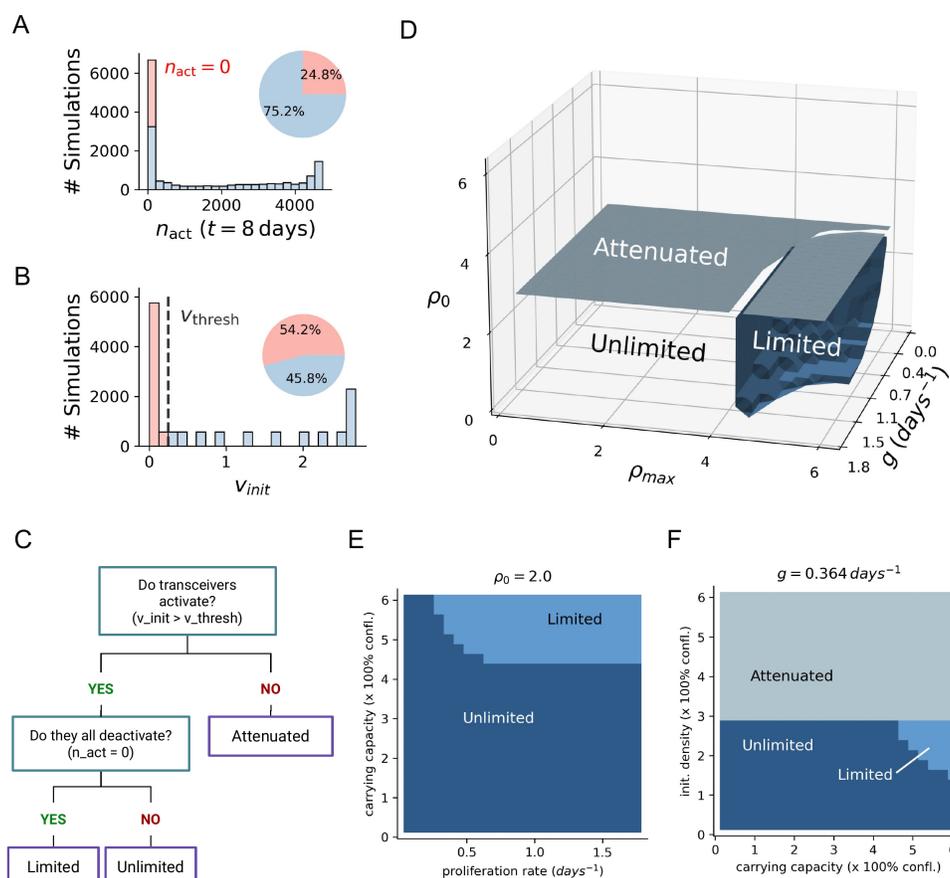
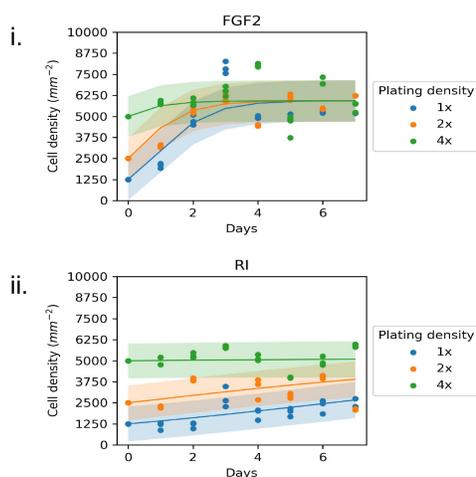


Figure S8: Categorization of transceiver responses into phases. (A-C) Simulated transceiver monolayers were categorized into three distinct phases based on the number of activated transceiver cells (n_{act}) at time $t = 8$ days and the initial rate of GFP production (v_{init}). See Figure 4 and Methods for simulation details and mathematical definitions. (A) Histogram of n_{act} at $t = 8$ days. The pie chart shows that at this time-point, 24.8% of simulated monolayers (red) have no activated transceivers ($n_{\text{act}} = 0$), and 75.2% (blue) have at least one. (B) Histogram of v_{init} shows a bimodal distribution, with the two modes separated by a chosen threshold value $v_{\text{thresh}} = 0.25$. The pie chart shows that 54.2% of simulations do not show activation ($v_{\text{init}} < v_{\text{thresh}}$, red), and 45.8% of simulations (blue) do. (C) Decision tree for assigning phase based on v_{init} and v_{thresh} . (D-F) Signaling phase also depends on the carrying capacity. Logistic growth of cell density depends on the intrinsic proliferation rate g , initial density ρ_0 , and carrying capacity ρ_{max} . Therefore, the signaling phase should depend on ρ_{max} in addition to g and ρ_0 . (D) Phase boundaries in 3D. For ρ_{max} less than a critical density, cell density does not increase enough to trigger signaling attenuation. Thus, the limited phase does not exist. (E-F) Phase diagrams generated by selecting slices of (D) at the indicated parameter values.

A Estimated growth parameters

	Estimated value (90% CI)		
	g days ⁻¹	rho_max cells mm ⁻²	sigma cells mm ⁻²
Untreated	0.616 (0.539 - 0.711)	7337 (6954 - 7771)	925 (778 - 1035)
FGF2	1.297 (1.032 - 1.681)	5918 (5672 - 6186)	965 (812 - 1080)
ROCK-i	0.173 (0.127 - 0.225)	5120 (4567 - 5856)	826 (694 - 920)

B Data and model predictions



C Maximum likelihood density corner plots (calc. by bootstrap)

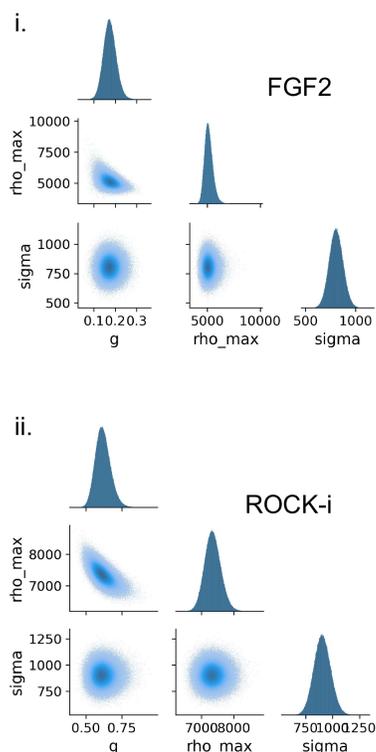


Figure S9: Estimation of growth parameters under drug-induced perturbations. The maximum-likelihood estimation (MLE) procedure employed in Figure SS7 was used to identify the intrinsic cell proliferation rate g and carrying capacity ρ_{\max} (parameters of logistic population growth) and the standard deviation σ . (A) Summary of results of the MLE procedure with 90% confidence intervals ($n=10^6$ bootstrap replicates). "Untreated" condition reproduced from Figure SS7. (B) In the indicated treatment conditions, sender-Transceiver co-culture density was measured daily for initial densities of 1x, 2x, and 4x (colored circles, $n=3$). Solid lines are simulated growth curves using MLE parameters (shaded regions are 80% probability). (C) 1- and 2-D distributions of parameters estimated by bootstrapping of MLE. Identified regions are well-bounded with little correlation, indicating successful fitting. Occasional high values of g (fast growth) suggest that higher time-resolution may be desired at early time-points. See Figure SS7 for complete results from the "Untreated" condition.

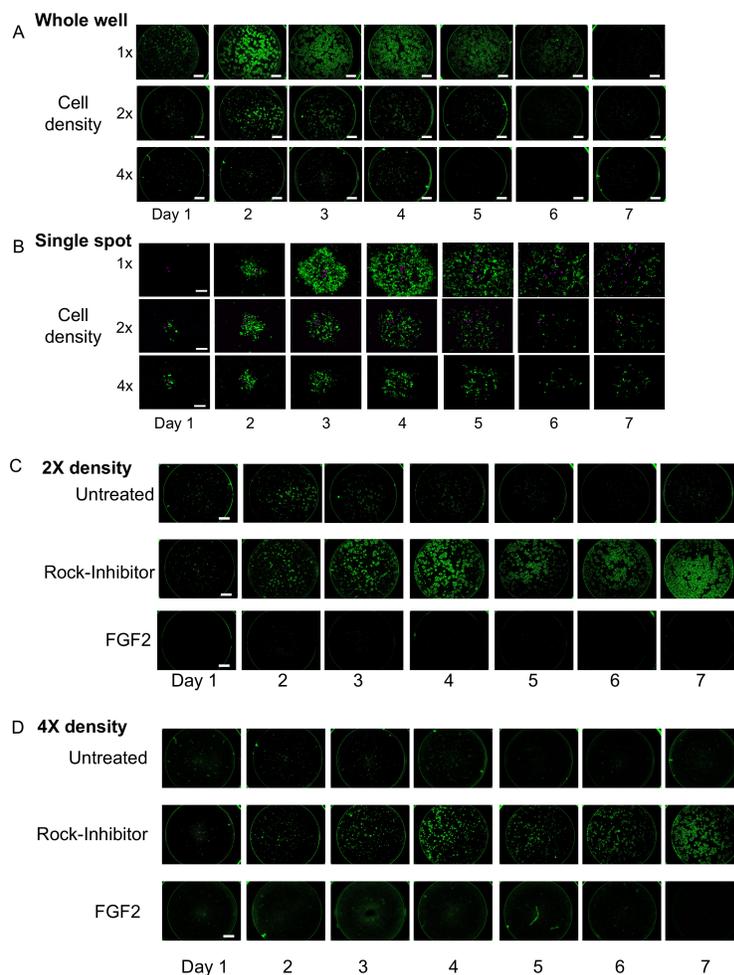


Figure S10: **Confluence and drugs** (A) Whole-well micrographs of 1:100 sender:transceiver cultures imaged every day for 7 days at the indicated initial cell densities (1X, 2X, 4X). First row (1x) reproduced from Figure 3. (B) Micrographs of single foci from sender-transceiver cultures at the indicated densities. Green is GFP, and purple is the nuclear marker present in sender cells. (C-D) Micrographs of green fluorescent channel for 1:100 sender-transceiver cultures over a 7-day time-course under the indicated drug treatments at initial densities of (C) 2x and (D) 4x confluence. Compared to 1X (shown in Figure 5D), the drug treatments show similar trends: FGF2 decreases signal propagation, while Rock-inhibitor increases it. In all conditions propagation is a little slower in the 4x condition than in 2x. All scale bars in the figure are 1 mm.

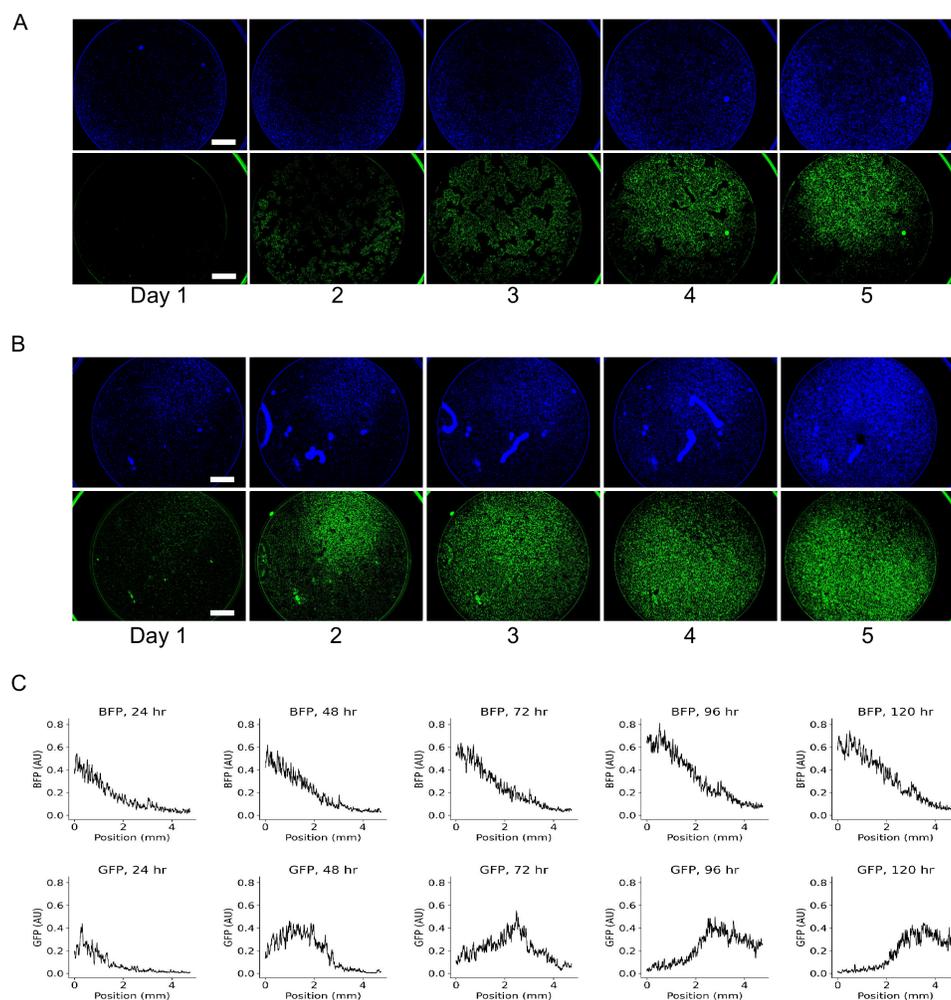


Figure S11: Kinematic wave experiment: replicates and quantification. (A-B) Micrographs of BFP and GFP fluorescence (top and bottom rows, respectively) from whole wells of sender-transceiver cultures plated at a mean initial density of 1x confluence. The heterogeneous distribution of cell density is demonstrated by BFP fluorescence (constitutively expressed in transceivers; blue channel), and the resulting kinematic wave of transceiver activation is shown by GFP fluorescence (green channel). Scale bars 1 mm. (C-D) Quantification of the fluorescence profile in the boxed area in Figure 6F over time. The BFP gradient (first row) is maintained throughout the time-course, while the peak in GFP fluorescence (second row) migrates over time.

CONCLUDING REMARKS

4.1 Overcoming barriers to engineering complex behavior with *in silico* circuit design and synthetic mechano-chemical coupling

Cellular behavior is determined by the dynamic activity of networks, or circuits, of biomolecules interacting with one another and the cellular microenvironment (Lim, Lee, and Tang, 2013). Many studies have used mathematical modeling to reveal the quantitative principles of how certain circuits generate specific phenotypes and thus have produced molecular blueprints for engineering diverse phenotypes including transcriptional oscillations (Elowitz and Leibler, 2000), single-cell polarization (Chau et al., 2012), and multicellular differentiation (Zhu et al., 2022). Despite the large variety of circuits in nature, most of these synthetic engineering efforts have focused on small signaling circuits (with 4 or fewer molecular components) that respond only to chemical stimuli such as the presence of ligands or small molecules (Manhas et al., 2022), due to two methodological barriers: the computational difficulty of studying larger circuits and the lack of molecular tools that respond to physical stimuli. This thesis addresses these barriers by developing a novel computational platform for designing large circuits and demonstrating a multicellular signaling paradigm that uses physical information, particularly cell density, to produce novel patterning phenotypes.

A crucial initial step in biological circuit design is to identify which circuit architectures, or topologies, will most reliably produce a given phenotype. When natural examples and human ingenuity are insufficient, circuit topologies can be discovered *de novo* using computational methods, the most common being exhaustive enumeration and simulation of all topologies in a given search space. As outlined in Chapter 1, for systems with $n > 4$ biomolecular components, this space of topologies becomes too large for enumeration. Chapter 2 of this thesis presents the software platform CircuiTree, which searches for circuit topologies with a given phenotype more efficiently using Monte Carlo tree search (MCTS). Commonly used for automated path planning and game-playing artificial intelligence (AI), MCTS efficiently traverses the search space by using the UCB selection policy from reinforcement learning (RL) to balance rapid discovery and exploitation of promising high-reward

areas with the exploration of potentially better alternatives. We show that CircuiTree designs 3-component stochastic transcriptional oscillators (a well-studied problem) approximately 100 times more efficiently than exhaustive enumeration by discovering and exploiting known network motifs. We use a parallelized version of CircuiTree to search for 5-component oscillators that oscillate even after the deletion of a random circuit component. After 5 million iterations searching a space of ~ 10 million topologies with 300 CPUs in just 2.5 days of real time, CircuiTree finds 1,368 candidate oscillators, the most fault-tolerant of which contain many interleaved oscillatory motifs, a design feature we call “motif multiplexing”. Multiple sub-oscillators are common among circadian clock circuits across divergent taxa, suggesting that this design feature may have evolved to make these large circuits robust to mutations during evolution (Wagner, 2005). Overall, this new open-source design platform facilitates the scalable study and design of larger biomolecular circuits.

Despite the development of a large variety of engineered proteins (and particularly synthetic receptors) in recent years, there is a shortage of synthetic tools that translate changes in the mechanical environment of the cell into a molecular signal such as a change in gene expression (so-called mechano-chemical coupling). Chapter 3 reports the discovery of cell density-controlled activation of a synthetic Notch (SynNotch) signaling system and demonstrates how mechano-chemical coupling can be used in tandem with mathematical modeling to program cell state over space and time. In this chapter, we describe a SynNotch signaling system that is suppressed at very high and very low densities of cell culture. When placed in a signal-propagation circuit, density-sensitive SynNotch can be used to control the spatial extent of signaling waves. Guided by a mathematical model of multicellular signaling, we show that signaling disc size can be controlled by manipulating parameters of cell proliferation such as growth rate and seeding density. Finally, we use spatial gradients of seeding density to achieve spatiotemporal patterning phenomena such as spatially distinct activation regions and kinematic waves, the latter of which have not been reported in any previous synthetic patterning circuit. Altogether, these results provide a novel modality for studying and engineering mechanico-chemical feedback and showcase its potential for expanding the repertoire of engineerable phenotypes to include spatiotemporal patterning.

4.2 Future perspectives

The mathematical methods developed in this thesis can be applied to design and study molecular circuits for a wide range of applications and, in general, to address many combinatorial problems in biology. The software package CircuiTree can be used to optimize circuit topologies for existing synthetic systems amenable to mathematical modeling, such as MultiFate (Zhu et al., 2022) and SynNotch (Morsut et al., 2016; Toda et al., 2020). Given the large set of synthetic molecular tools that have been developed in recent years, CircuiTree may be particularly suited to address the difficult problem of how to combine multiple tools synergistically. Notably, the greater efficiency of CircuiTree over traditional approaches opens the door to phenotypes that may have relatively expensive simulations, such as reaction-diffusion based patterning systems, “multicellular” circuits (having two or more distinct cell types), mechano-chemical feedback circuits, and large deterministic or stochastic systems. More generally, CircuiTree could be adapted to address other problems of combinatorial optimization in biology by casting them as decision-making problems. For instance, given a set of observations (say, single-cell sequencing measurements in multiple conditions), it could search a large space of possibilities for the network architectures that best explain the data by using a reward function that computes a maximum-likelihood estimation or Bayesian information criterion. It could also perhaps be applied to protein sequence design problems by exploring peptide sequence space using an RL framework.

As CircuiTree is the first application of RL to biological circuit design, there remain many fundamental questions and methodological improvements that could be addressed.

- What are good criteria for convergence and early stopping of the search?
- Anecdotally, search efficiency increases with the number and size of solution clusters or “motifs” in the search space. Can we quantitatively describe this clustering, and how specifically does search time scale with the degree of clustering?
- What is the best way to extract patterns (motifs) from the results of a tree search? Anecdotally, the random sampling method described in Chapter 2 can require a large number of samples if the search space is large and solutions are sparse.

- Can parameter sampling be done more efficiently than the random uniform sampling used in Chapter 2, perhaps by incorporating Bayesian sampling methods like approximate Bayesian computation (Woods et al., 2016)?
- In game-playing studies, the performance per iteration is known to decrease as the number of parallel threads increases (Segal, 2011). For circuit design, this degradation in performance may differ between decision trees. Can the number of parallel threads be adapted on-the-fly to minimize the overall expected sampling regret?
- Some design problems may require search spaces (referred to as Grammars in the codebase) that are not built-in to the package. These Grammars could be developed and added in an open-source manner.
- For simulations with long execution times (roughly a minute or longer), could heuristic methods be used to quickly approximate the reward/value function? In game-playing, deep learning has been used to speed up reward evaluations, with great success.
- In what cases would other RL approaches like policy-gradient and Q-learning methods be preferable to MCTS?

Future work with density-dependent SynNotch (ddsNotch) should further investigate the causal mechanism of density-dependence, which has remained elusive. Other future work could improve the resolution with which cells are patterned on the growth substrate, for instance by using lithography to deposit fibronectin with high spatial resolution. This could enable a high degree of control over spatiotemporal activation, as opposed to the “plate tilting” method described in Chapter 3. Additionally, the transcriptional output of ddsNotch could be engineered to regulate cell proliferation, thus forming a tunable feedback control loop to control proliferation in a manner complementary to the auxin-based control circuit devised by Ma and colleagues [[ref Ma et al]]. The transceiver circuit (with or without density-dependence) could also be adapted to study other spatiotemporal patterning phenomena such as reaction-diffusion patterning.

Natural tissues are composed of multiple cell types, each expressing distinct regulatory networks. Combining the methodologies from both projects in this thesis, CircuiTree could be used with multicellular modeling simulations to study the principles of how tissues self-organize and maintain their integrity.

- How do mature tissues resist mechanical disruptions, such as the metastatic cascade during carcinogenesis? Why are some individuals and some tissues more susceptible to metastatic invasion than others?
- What regulatory circuits do growing and regenerating tissues use to balance structural plasticity and integrity?
- How does the intricate microstructure of tissues such as the liver and kidney self-organize from cell-level signaling and differentiation circuits?

Finally, future projects should explore how the work described in this thesis can catalyze the development of synthetic circuits for therapeutic and biomedical applications. For instance, could CircuiTree be combined with mechanical cell models to design a cell therapy that structurally stabilizes a tissue suffering from carcinoma in situ, thus preventing imminent metastatic invasion? One contemporary barrier in cell therapy is the poor penetration of therapeutic chimeric antigen receptor T cells (CAR-T cells) into the so-called “cold” interior of some solid tumors. Can we improve the penetration of solid tumors by, for example, designing a second therapeutic cell type that remodels the extracellular matrix to slow growth and encourage penetration? In general, the author believes that multicellular therapies (that is, cell therapies consisting of multiple engineered cell types) are a promising avenue for future translational work in synthetic biology.

References

- Chau, Angela H. et al. (Oct. 12, 2012). “Designing Synthetic Regulatory Networks Capable of Self-Organizing Cell Polarization”. In: *Cell* 151.2, pp. 320–332. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.08.040. URL: <https://www.sciencedirect.com/science/article/pii/S009286741201121X> (visited on 01/31/2023).
- Elowitz, Michael B. and Stanislas Leibler (Jan. 2000). “A synthetic oscillatory network of transcriptional regulators”. In: *Nature* 403.6767. Number: 6767 Publisher: Nature Publishing Group, pp. 335–338. ISSN: 1476-4687. DOI: 10.1038/35002125. URL: <https://www.nature.com/articles/35002125> (visited on 10/17/2023).
- Lim, Wendell A., Connie M. Lee, and Chao Tang (Jan. 24, 2013). “Design Principles of Regulatory Networks: Searching for the Molecular Algorithms of the Cell”. In: *Molecular Cell* 49.2. Publisher: Elsevier, pp. 202–212. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2012.12.020. URL: [https://www.cell.com/molecular-cell/abstract/S1097-2765\(13\)00004-X](https://www.cell.com/molecular-cell/abstract/S1097-2765(13)00004-X) (visited on 10/17/2023).

- Manhas, Janvie et al. (Mar. 2022). “The evolution of synthetic receptor systems”. In: *Nature Chemical Biology* 18.3. Number: 3 Publisher: Nature Publishing Group, pp. 244–255. ISSN: 1552-4469. DOI: 10.1038/s41589-021-00926-z. URL: <https://www.nature.com/articles/s41589-021-00926-z> (visited on 02/08/2023).
- Morsut, Leonardo et al. (Feb. 11, 2016). “Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors”. In: *Cell* 164.4. Publisher: Elsevier, pp. 780–791. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2016.01.012. URL: [http://www.cell.com/cell/abstract/S0092-8674\(16\)00052-0](http://www.cell.com/cell/abstract/S0092-8674(16)00052-0) (visited on 07/14/2021).
- Segal, Richard B. (2011). “On the Scalability of Parallel UCT”. In: *Computers and Games*. Ed. by H. Jaap van den Herik, Hiroyuki Iida, and Aske Plaat. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp. 36–47. ISBN: 978-3-642-17928-0. DOI: 10.1007/978-3-642-17928-0_4.
- Toda, Satoshi et al. (Oct. 16, 2020). “Engineering synthetic morphogen systems that can program multicellular patterning”. In: *Science* 370.6514. Publisher: American Association for the Advancement of Science Section: Report, pp. 327–331. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abc0033. URL: <http://science.sciencemag.org/content/370/6514/327> (visited on 07/14/2021).
- Wagner, Andreas (Aug. 16, 2005). “Circuit topology and the evolution of robustness in two-gene circadian oscillators”. In: *Proceedings of the National Academy of Sciences* 102.33. Publisher: Proceedings of the National Academy of Sciences, pp. 11775–11780. DOI: 10.1073/pnas.0501094102. URL: <https://www.pnas.org/doi/full/10.1073/pnas.0501094102> (visited on 10/19/2023).
- Woods, Mae L. et al. (June 17, 2016). “A Statistical Approach Reveals Designs for the Most Robust Stochastic Gene Oscillators”. In: *ACS Synthetic Biology* 5.6. Publisher: American Chemical Society, pp. 459–470. DOI: 10.1021/acssynbio.5b00179. URL: <https://doi.org/10.1021/acssynbio.5b00179> (visited on 10/13/2023).
- Zhu, Ronghui et al. (Jan. 21, 2022). “Synthetic multistability in mammalian cells”. In: *Science* 375.6578. Publisher: American Association for the Advancement of Science, eabg9765. DOI: 10.1126/science.abg9765. URL: <https://www.science.org/doi/10.1126/science.abg9765> (visited on 04/14/2023).