Multicellular Synthetic Biology in Mammalian Systems

Thesis by
Yitong Ma

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended March 8th, 2023

Yitong Ma
ORCID: 0000-0003-4446-7326

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Michael Elowitz for all the mentorship and support over the almost eight years as a graduate student, and even before that, as my SURF (summer internship) mentor when I was just a junior undergraduate. It was that wonderful time I spent in this lab as a undergraduate summer intern, that inspired the undergrad me to enter this field eventually. Michael's great vision, intuition, and pure enthusiasm in science has inspired me countless times during the years, and most of the ideas presented in this thesis can be traced back to some afternoon chat with him in his office or next to our lab's coffee machine. Michael was not only enormously supportive of my research, but also very patient when I face difficulties, or even wandered off a bit (like spending time with 3D printing and making a software). I am immensely grateful to have Michael as my PhD mentor.

I am also very grateful for my mentors when I started the lab, Mark Budde (who was also my SURF mentor) and Lacra Bintu. I have learned a great amount from them, including both the basics in the lab, and the knowledge that I needed for going through the PhD. They have always been very patient and gave me advice, even after they left the lab. Mark also trained me to use the lab espresso machine, which has been a great help a lot of times.

I also want to thank Kaiwen Luo, who entered the lab only about 1.5 years ago as graduate student. He has worked with me extensively on the "seeker-killer" and spearheaded a lot of the work. He also made the poster for my defense. I also want to thank Junqin Zhu, who SURFed twice in our lab, and are heavily involved in two of the projects I will present in this thesis. Those works would be impossible without her. I also want to thank Jo Leonardo, Leah Satant, and James Linton, our lab administrator and managers. They made our lab a wonderful and hospitable place for study, work, and research. I cannot imagine how our lab will run without their extraordinary efforts. I also want to thank Lucy Chong, Sheng Wang, Yihan Lin, Yaron, Anbeti, John Yong, Michael Flynn, Evan Mun, Felix Horns, Jan Gregrowicz, Martin Tran for all the disscussions about both science and life inside the labs, as well as everyone in the Elowitz lab that made the lab a great place for my PhD years.

I also want to thank Richard Murray, as my collaborator, leader of our joint "Bio-Control" DARPR project, as well as a member of my thesis committee. His management and scientific inputs have guided us throughout our joint project. Also lots

of thanks to Michaëlle Mayalu of the Murray lab. She contributed greatly to our synthetic population control project. I also learned about matlab and LaTeX from her. I also received lots of help from the discussions with Cindy Ren and Leopold Green from the Murray lab.

I am also very grateful to have worked with Dr. Mikhail Shapiro, and Abdullah Farooq, Mohamad Abedi, and Margaret Swift from his lab. They helped immensely with moving the auxin system into immune cells, and kindly taught us to work with mice, and in their animal facility for in vivo experiments. Besides those help on experiments, scientific discussions and writing with Mikhail and Michael together has always been very inspiring, and I got precious experience of looking at the problem from different angles.

Besides Michael and Richard, I also want to thank Dr. Mitch Guttman and Dr. Matt Thomson as my committee member. Their kind and generous suggestions have always been helpful for driving the projects forward.

I also received a lot of help from Jeff Park of the Matt Thomson lab with various sequencing experiment; Igor Antoshechkin of the Caltech Sequencing core with RNA sequencing; Jamie Tijerina and Shelly Diamond of the Caltech flowcytometry core with FACS sorting and advice on flowcytometry; Daniel Wagenaar of prototyping tools used in the researches. The projects presented here would have been impossible without their kind help.

During my almost eight years at Caltech as both a SURF and as a graduate student, I have also received countless help from lots of other people, including our BBE administrator Liz Ayala, SURF administrator Carol Casey, librarian Natalie Gilmore, and Paula Gaetos. Also as a foreign student, I want to express my special thanks to the ISP staff, including Laura Flower Kim and Daniel Yoder. Without them it would have been almost impossible to go through the daunting paper work. I also want to send my special thanks to Maia Jaser White, our director of Caltech chamber music. I am very grateful for her instructions regarding my violin skill, and luckily enough for me, I played alongside her on one of the Caltech recitals. Eight years is a long time, and it is impossible that I listed all the help I received from the Caltech community here. I just want to express my sincerest gratitude to the community here.

Lastly, I also want to express my thanks to my family, including both my parents Jun Ma and Guiping Wu in China and my wife Sophie Song now in San Francisco. My

parents have been so supportive of me pursuing the scientific path throughout my life. I have not been able to visit them since the COVID pandemic, but their support for me, even remotely, has helped me overcome so many difficulties. Sophie, who I married last year, has been a constant support since we met during college. I feel ultimately fortunate to have had her support with me through all the years of my PhD, and it would have been be impossible for me to reach this point here without her.

ABSTRACT

In multicellular organisms, different types of cells use intercellular signals to communicate and regulate population dynamics, and further coordinate complex behaviors. This presents a rarely tapped into potential for mammalian synthetic biology, which was largely restricted to engineering a single cell type in the past to mimic and use similar multicellular designs to achieve more functionalities. However, with current synthetic biology tools and designs, there are several major challenges to achieve a multicellular circuit. Challenges include precise and tunable control over cell type switching, having an orthogonal cell-cell communication signal, and robust control of cell populations.

To address these challenges, this thesis presents a system for tunable regulating of gene expression with DNA methylation, an auxin-based module for mammalian cell-cell communication, and a robust circuit for population control in mammalian cells. I further applied these work to engineering immune cells to show the potential of multicellular circuits in immunotherapies. Together, these works demonstrated the possibility of constructing multicellular circuits in mammalian systems, and that multicellular circuit can further extend the scope of synthetic biology to achieve more complex functions.

PUBLISHED CONTENT AND CONTRIBUTIONS

Ma, Yitong, Mark W. Budde, Junqin Zhu, et al. (2023). "Promoter CpG content regulates DNMT-dependent silencing dynamics." In: *ACS Synthetic Bio*. Undergoing review.

Y.M. participated in the conception of the project, conducted the experiments, prepared data, and wrote the manuscript.

Ma, Yitong, Mark W. Budde, Michaëlle N. Mayalu, et al. (2022). "Synthetic mammalian signaling circuits for robust cell population control." In: *Cell* 185.6, pp. 967–979. DOI: https://doi.org/10.1016/j.cell.2022.01.026.

Y.M. participated in the conception of the project, conducted the experiments, prepared data, and wrote the manuscript.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AAV. Adeno-associated virus.

AID. Auxin inducible degron.

CAR. Chimeric antigen receptor.

CHO cells. Chinese hamster ovary cells, see CCL-61 at ATCC.

DNMT. DNA methyl-transferase, responsible for the transfer of a methyl group from methyl donor, to the 5-position of cytosine residues in DNA (Jin and Robertson, 2013).

FACS. Fluorescence assisted cell sorting.

GPCR. G-protein coupled receptor.

IAA. Indole-3-acetic acid, CAS87-51-4.

IAM. indole-3-acetamide, CAS879-37-8.

IL-2. Interleukin 2.

IRES. Internal ribosome entry site.

NAA. 1-napthalenatic acid, CAS86-87-3.

NAM. 1-naphthaleneacetamide, CAS86-86-2.

NGS. Next-generation sequencing, in this thesis we use the ”sequencing by synthesis platform” by Illumina Inc..

PBMC. Peripheral blood mononuclear cells.

SCF complex. Skp, Cullin, F-box containing complex.

scFv. Single-chain variable fragments. An artificial constructs composed of the immunoglobulin heavy and light chain variable regions connected by a peptide linker.

TF. Transcription factor.

TSS. Transcription start site.

Chapter 1

INTRODUCTION: SYNTHETIC BIOLOGY FROM SINGLE CELLULAR TO MULTICELLULAR

Synthetic biology originally aims to understand fundamental biological design principles, such as gene expression dynamics (Elowitz, Levine, et al., 2002; Cai, Friedman, and Xie, 2006), oscillatory behaviors (Elowitz and Leibler, 2000; Potvin-Trottier et al., 2016), and ligand-receptor interactions (Sprinzak et al., 2010; Antebi et al., 2017), by building circuits from well-characterized components in a well-controlled context (often called the bottom-up approach). More recently, various studies started to apply synthetic biology to fields like cancer biology (reviewed by Ruder, Lu, and Collins, 2011), immunotherapies (Choe et al., 2021; J. H. Cho, Collins, and Wong, 2018; also reviewed by Noll and Fraietta, 2022) and tissue engineering (Guye et al., 2016; Skylar-Scott et al., 2022) for clinical potentials.

Though some of the examples above yield brilliant results with circuits that are simple but elegant, it is likely that future application will require synthetic circuits with more and more complexity. For example, a hypothetical engineered cell for immunotherapies will need multiple receptors for antigen or environment detection, logic circuit to process inputs from the receptors, ”payload” genes for the desired outputs, as well as gene editing for resistance to the hostile tumor micro-environment (Shih and Y. Y. Chen, 2022). However, adding more components, especially into mammalian cells, faces a massive challenge in terms of gene delivery. Currently most reliable delivery methods, mostly including viral delivery via adeno-associated virus (AAV) or lentivirus, have very limited capacity per construct (Tornabene et al., 2019; Counsell et al., 2017). This can be further complicated due to promoter interference (Chtarto et al., 2003; Curtin et al., 2008; Amendola et al., 2005), if different components need to be driven by different promoters for responses to separate inputs. In contrast, delivery methods that can accommodate larger ”cargo” like lipofection or electroporation, are less robust especially in harder-to-transfect on cell types (Dullaers et al., 2004; Cao et al., 2010). One potential circumvention is to delivery multiple virus into the same cell, but there has been evidence about high titer of AAV transduction lowering stem cell viability (P. Lin et al., 2012; Marasini et al., 2017). On the lentivirus side, due to its mutagenic potential of reverse transcription and integration, most clinically relevant protocols aim for integration copy number

as low as possible (L. Fernández et al., 2019; Ghassemi et al., 2022). This limitation provokes the question that, is there a more scalable approach for integrating more complex circuit into cells?

The more promising approach is to distribute different modules of the circuit into different cells (types, or strains), thus creating multicellular synthetic circuits. Much like cells in multicellular organism or bacterial communities, cells carry out simple tasks and coordinate actions through cell-cell communications for more sophisticated system-level functions (more in section 1.1). This approach potentially reduces the amount of required engineering per cell type, and scales the circuit up by adding in more cell types.

Additionally, multicellular circuits also provide the following benefits over traditional single cellular circuits:

1. **Reusability of components.** Dividing the circuit components into different cell strains enables the reuse of the same components, but in different cells.

   Traditionally in single cell synthetic biology, the up-scaling is constrained by the number of an ”orthogonal” components, that do not interfere with each other when put together in a single cell. For example, recent work from our lab demonstrated that one can construct all seven basic logic gates with three proteases with pairwise orthogonality (Gao et al., 2018). However, a straightforward scale up of this concept will require 3xN orthogonal proteases for a circuit with N logic gates.

   Furthermore, even theoretically orthogonal and scalable systems can have unexpected cross talk when implemented into the same cell. For example, the synthetic zinc-finger system that in theory should be able to generate hundreds of synthetic transcriptional factors (TFs) by combinatory rearrangement, faces challenges to create an orthogonal set of six TFs (Khalil et al., 2012). In another example, synthetic Notch (synNotch) and chimeric antigen receptor (CAR), two systems derived from entirely different origins (Notch and TCR, respectively), can have unexpected cross interactions when expressed on the same cell surface (I. Zhu et al., 2022).

   These problems can potentially be solved by multicellular synthetic circuits, as each cell strain provides a physically separated, thus orthogonal context for its sub-circuits.

2. **Cell type specialization.** Another unique advantage of multicellular synthetic circuits is the potential to harness the inherent specification of different cell types.

   For example, in cell-based therapies against solid tumors, the engineered cells need to infiltrate, recognize, amplify themselves, and kill cancer cells. The most common cell type to engineering for this purpose currently are T cells due to their exceptional ability of self amplification upon target engagement and killing capacity. However, they can have poor infiltration against solid tumors (Scarfò and Maus, 2020). On the other hand, macrophages are known to be good at infiltration (Klichinsky et al., 2020; Aalipour et al., 2019), but lack the capability of self-amplification and killing effectively. A multicellular circuit could program the infiltration and detection function into macrophages, and then relay the information to engineered T cells for killing, thus taking advantage of both cells' natural specialty.

1.1 Inspiration from natural multicellular circuits

Cells in bacterial communities and multicellular organisms naturally operate multicellular circuits to achieve complex functions. In this section, we will discuss several natural circuits and try to answer what are required components for these types of circuit.

In microbial communities, cells can decide when to disperse from their multicellular structure, biofilm, based on environmental cues produced or caused by other cells in the biofilm, like specific fatty-acid signals, or simply depletion of oxygen (Rumbaugh and Sauer, 2020). In the case of quorum sensing cells secrete diffusive signals to inform each other about the current population density (quorum) and coordinate cooperative behaviors (Waters and Bassler, 2005; Papenfort and Bassler, 2016).

A classic example in animal development is the "French flag" model, in which a diffusive molecule (morphogen) is produced by a group of cells and shapes a spatial gradient in the population. The gradient is further perceived by cells, and triggers different irreversible changes (differentiation) in the receiving cells (Bier and De Robertis, 2015). A more complex example is the somitogenesis in vertebrates, in which a gene oscillation on single cell level are fixed into a spatial repetitive pattern (somites) (Gibb, Maroto, and Dale, 2010). This process is achieved by a circuit consisting of a gene osccilator on single cell level, contact dependent Notch signal for

local synchronization, and several diffusive molecules to form opposing gradients for final determination of cell state.

Our immune system provides a series of examples of multicellular circuits, where the regulation (both via cell cycle regulation and differentiation) of population size of different types is key to the functions. For example, IL-2's secretion by immune cells (mostly CD4 T and dentritic cells) are important for primary expansion of both CD8 and CD4 T cells upon antigen recognition, and are further produced by activated T cells, turning into a positive feedback driving the T cells' clonal expansion. IL-2 also tip the balance between the differentiation of Th17 cell and survival of Treg cells, ultimately determining the extension or termination of inflammation (Boyman and Sprent, 2012).

Despite the complexity in the wiring of the natural circuits above, we can conclude and identify the essential components from these circuits, and develop synthetic counterparts for synthetic application.

1.2 Modules for ideal multicellular circuit

Based on observing above natural examples (section 1.1), we propose that the following components are keys for multicellular circuits (schematics in Figure 1.1).

1. **Intracellular logic.** These circuits, as the traditional focus of synthetic biology, are spatially separated in different cell types, and are responsible for processing the inputs and producing outputs on single cell level. These circuits, as demonstrated previously, can be transcription based (R. Zhu et al., 2022; Khalil et al., 2012), post-translation-based (Gao et al., 2018), RNA logic (Matsuura et al., 2018; Wroblewska et al., 2015), splicing-based (North et al., 2022), or a combination of all of the above.

2. **Contact dependent signal.** Contact dependent synthetic signal programs the cells to communicate with their immediate neighbors.

   Systems with this function have been developed over the years by re-engineering natural receptors like Notch (to synNotch and SNIPR) (Morsut et al., 2016; I. Zhu et al., 2022), TCR (to CAR) (Noll and Fraietta, 2022; Y. Liu et al., 2021) or cytokine receptors (Kojima, Scheller, and Fussenegger, 2018). Though notably CAR mentioned above do not provide fully synthetic

outputs, the natural output (in this case T-cell cytotoxicity) can be convenient in a lot of applications. [1]

3. **Diffusive signal.** Diffusive signal in a multicellular circuit can either provide global/long-distance communications or spatial information if produced by a group of spatially defined cells (similar to "morphogens" in development biology).

   Previous examples of this class include repurposed various natural systems like GPCR (Barnea et al., 2008), Sonic Hedgehog (P. Li and Elowitz, 2019), cytokine-receptor (Sockolosky et al., 2018), or fully modular designs (MESA, Daringer et al., 2014). Additionally, synthetic systems produce and sense metabolites from either plant (M.-T. Chen and Weiss, 2005) or bacteria (Bacchus et al., 2012) has been developed for cell-cell communication in mammslian system. Though most systems above are not designed to send and receive orthogonal signals (with the exception of the IP-CRE1 system from plant, (M.-T. Chen and Weiss, 2005)), some of them (e.g. synthetic GPCR-based systems and MESAs) can potentially perceive synthetic secreted proteins and peptides.

4. **Control of cell proliferation and differentiation.** Another major aspect of multicellular circuit is the ability to control population dynamics (both absolute size and relative ratio of populations), either via driving differentiation or regulate proliferation based on input. Control of proliferation can be achieved via regulation of proliferation (e.g. controlled cytokine release for immune cells (Allen et al., 2022), or expression of cell cycle arrestors (Satyanarayana and Kaldis, 2009)), or cell death (e.g. controlled apoptosis (Straathof et al., 2005; Gao et al., 2018)). On the other hand, drive of differentiation has been achieved by either rewiring of the endogenous differentiation circuit (Kitada et al., 2018), or creating fully synthetic bi-stable gene switch to lock cells into a certain desired state (cell-type) (R. Zhu et al., 2022).

1.3 Content of this thesis

In this thesis, I will present an auxin-based diffusive signaling module for mammalian cell-cell communication and a robust circuit for population control in mam-

[1] This and the next point about receptors for surface bound and diffusive signals are well reviewed by Manhas et al., 2022 from a receiver angle.

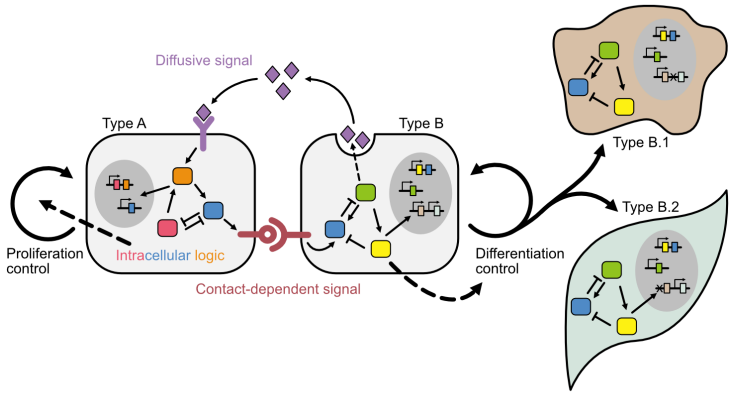


Figure 1.1: Components for the ideal multicellular synthetic circuit.
Different cell types use their intracellular logic to process inputs from, as well as send output signals to other cells, regulating cell proliferation and differentiation on a multicellular level.

malian cells. I will also present a DNA methylation-based, tunable gene switch, that could potentially drive stable cell-type differentiation.

Finally, I will present our ongoing advances on constructing multicellular circuit in immune cells, demonstrating the potential of multicellular circuits in clinically relevant applications.

Together, these works demonstrated the possibility of constructing multicellular circuits in mammalian systems, and that multicellular circuit can further extend the scope synthetic biology to achieve more complex functions.

Chapter 2

SYNTHETIC AUXIN-BASED MAMMALIAN SIGNALING CIRCUITS FOR ROBUST CELL POPULATION CONTROL

2.1 Introduction

Cells use intercellular communication systems to sense and control their own population density. In microbial communities, cells exchange diffusive signals to coordinate cooperative behaviors and form biofilms through the process of quorum sensing (Waters and Bassler, 2005; Papenfort and Bassler, 2016) or manipulation of distribution of the redox small molecules (Dietrich et al., 2013). In multicellular organisms, intercellular communication is essential to enable precise developmental patterning (Gibb, Maroto, and Dale, 2010; Bier and De Robertis, 2015), control immunological responses (Boyman and Sprent, 2012; C.-C. Chen et al., 2015; Muldoon et al., 2020), and coordinate organism-level physiology (Keener and Sneyd, 2009, Chapter 16).

Synthetic intercellular communication systems could allow the engineering of inherently multicellular behaviors not possible with cell-autonomous circuits (Toda, Blauch, et al., 2018). For example, in bacteria, foundational synthetic biology studies showed how coupling quorum sensing systems to cell death could enable bacteria to limit their own population size (You et al., 2004; Scott and Hasty, 2016), or drive synchronized oscillations of drug release for therapeutic applications (Din et al., 2016). In mammalian cells, an orthogonal, or "private" communication channel that allows specific communication between engineered cells could enable engineering of analogous circuits (Figure 2.1).

The ideal private communication system for mammalian population control would use a diffusible signal, avoid undesired interactions with non-engineered cells, permit external control over the strength of signaling, and operate in a broad variety of cell types. It should also allow direct and rapid control of diverse target protein activities to allow flexible interfacing within cells, and exhibit minimal immunogenicity to facilitate potential biomedical applications.

Synthetic intercellular communication channel based on auxin

In the past, mammalian cells have been engineered to produce, sense, and process signals from natural ligands by rewiring signaling pathways such as Nodal-Lefty, Sonic Hedgehog, and Notch (Matsuda et al., 2012; P. Li and Elowitz, 2019; Sekine, Shibata, and Ebisuya, 2018), and by repurposing the amino acid tryptophan as a signaling molecule (Bacchus et al., 2012). However, these approaches are not orthogonal to endogenous systems. On the other hand, synNotch receptors allow multiple orthogonal communication channels, but depend on cell contact interactions to mediate diffusible signaling (Toda, McKeithan, et al., 2020).

Auxins, a class of plant-specific hormones that coordinate growth and behavior including root initiation, embryogenesis, and tropism (Tanaka et al., 2006), represent an excellent candidate for this role. Molecularly, auxin induces protein-protein interactions between the F-box transport inhibitor response 1 (TIR1) protein and its target proteins. This leads to the assembly of a Skp, Cullin, F-box containing (SCF) complex, which in turn recruits E2 ubiquitin ligases that target specific proteins for degradation (Reitsma et al., 2017). Because TIR1 and its targets are absent in mammals, auxin does not regulate endogenous mammalian proteins. However, ectopic expression of TIR1 from rice (osTIR1) is sufficient to confer auxin-dependent degradation of proteins engineered to contain a minimal auxin inducible degron (mAID, or AID for simplicity in this paper) (Nishimura et al., 2009; Natsume et al., 2016). Thus, auxin is orthogonal to endogenous mammalian pathways, but can enable direct control of cellular activities through engineered protein targets. Additionally, in yeast, ectopic expression of bacterial indole-3-acetic acid hydrolase was shown to catalyze auxin production from an inactive precursor indole-3-acetamide (IAM), allowing control over auxin production (Khakhar et al., 2016). Nevertheless, a full auxin sending and receiving signaling system, which is necessary for population control, has not been established in mammalian cells.

In this thesis, we engineer the auxin pathway to act as a private mammalian communication channel. Combining auxin-synthesizing enzymes and auxin transporters, and employing alternative auxin precursors, we show that the auxin pathway can be used for synthetic quorum sensing in mammalian cells.

Robust mammalian population control circuit

With a established system capable of sending and receiving a diffusible signal simultaneously, we can construct the simplest population control circuit by coupling

a quorum-sensing circuit to cell death, similar to what has been achieved in bacteria (You et al., 2004; Scott and Hasty, 2016). This circuit configuration essentially created a negative feedback loop on between population size and cell net growth (Figure 2.1, right panel). However, a critical challenge for any population control circuit, including the negative feedback circuit mentioned above, is evolutionary robustness. By limiting growth, a population control circuit inherently selects for "cheater" mutations that escape regulation.

Previous works has established several type of circuit to counter this "cheater" problem. In bacteria, toxin-antitoxin systems and periodic strain replacement can prevent cheater escape (Stirling et al., 2018; Liao et al., 2019). However, these systems use components that do not function in mammalian cells or are not cell-autonomous. In the mammalian context, a seminal analysis of natural cell population size control systems in cytokine and glucose sensing circuits revealed a paradoxical architecture, in which a single signal stimulates both proliferation and death of the same target cell population, to actively select against cheaters (Hart et al., 2014). In this paradoxical design, mutations that diminish signal sensing lead to cell death and are eliminated (Karin and Alon, 2017). Despite its power and elegance, the paradoxical architecture has not, to our knowledge, been demonstrated synthetically in living cells.

In this thesis, we used the auxin cell-cell communication module (see 2.1) to construct and analyze synthetic population control circuits with different architectures. Specifically, we constructed and compared negative feedback and paradoxical control systems that regulate their own population size through auxin quorum sensing. While both circuits limit population size at early times, the paradoxical system enhances evolutionary stability, as predicted theoretically, extending the duration of population control.

2.2 Engineered mammalian cells sense population with auxin-based communication channel

In this section, we engineered mammalian cells (CHO cells) to sense, respond to and produce auxin, by ectopic expression of components from plants and agrobacterium. We further engineered cells to sense and produce simultaneously, and achieved synthetic quorum sensing in mammalian system.

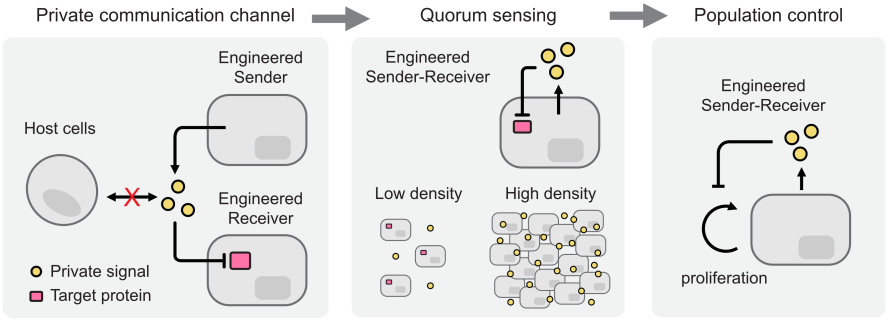


Figure 2.1: The plant hormone auxin allows private-channel mammalian cell-cell communication.
Left: an ideal mammalian private-channel communication system would allow engineered cells to send and respond to an orthogonal signal that does not interact with host cells. Middle: Engineered cells that can send and receive the signal simultaneously can respond to their own population size. Right: coupling the sending-receiving function with cell survival in a negative feedback loop enables population control.

Engineered mammalian cells sense auxin

To establish and characterize auxin regulation of mammalian cell growth, we coupled auxin sensing to drug resistance and fluorescence. Specifically, we fused blasticidin S deaminase (BlastR) (Kimura, Takatsuki, and I. Yamaguchi, 1994), whose protein product is necessary for survival in the presence of blasticidin, to AID and mCherry domains, allowing auxin-dependent degradation and fluorescent readout of protein concentration, respectively. We then stably integrated this chimeric gene, along with a constitutively co-expressed osTIR1, in CHO-K1 cells to create an auxin-sensitive "Receiver" cell line (Figure 2.2A).

To validate auxin regulation of mCherry-AID-BlastR, we cultured Receivers in media containing different concentrations of two auxin variants: either the major natural auxin, indole-3-acetic acid (IAA), or a synthetic auxin analog, 1-napthalenatic acid (NAA) (Figure 2.2B). Both auxins reduced mCherry fluorescence in a dose-dependent manner, with EC50 values of 0.11 μM and 0.76 μM, respectively. Addition of IAA to media containing blasticidin was sufficient to degrade BlastR and inhibit cell survival (Figure 2.2C). This effect was dose-dependent with both blasticidin and IAA. Comparing fluorescence of mCherry-AID-BlastR in Figure 2.2B with cell survival in 2.2C shows that a small amount of mCherry-AID-BlastR is sufficient to enable survival in blasticidin. These results confirmed that the AID

domain and osTIR1 together are sufficient to enable net growth regulation by auxin (IAA) in mammalian cells.

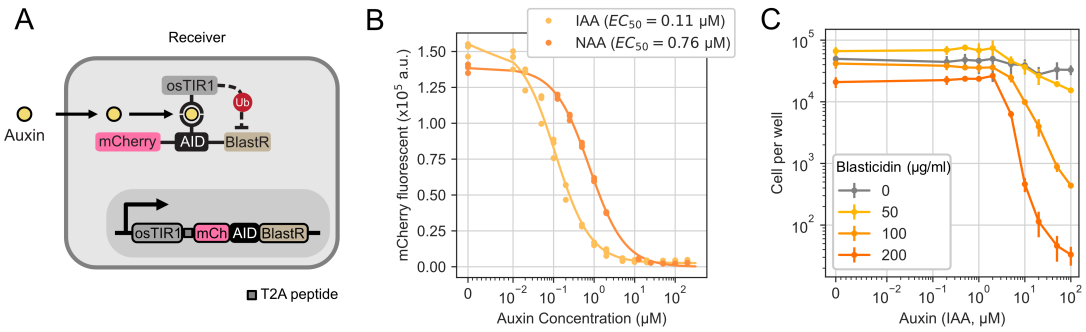


Figure 2.2: Engineered auxin receivers responds to auxin.
(A) Auxin Receiver cells constitutively express a fluorescent target fusion protein, mCherry-AID-BlastR, as well as the F-box protein osTIR1. In the presence of auxin (yellow circle), osTIR1 and the AID-tagged target protein assemble into an SCF complex, which allows ubiquitylation of the target protein, leading to target degradation. Both proteins are encoded on a single transcript, with an intervening T2A ribosomal skip sequence (grey square) to yield separate proteins (Szymczak et al., 2004).
(B) Auxin regulates intracellular protein levels. The response of mCherry fluorescence in Receiver cells (A) to two different species of auxin (IAA and NAA) was measured after two days of treatment (dots). The responses follow Michaelis-Menten kinetics (fitted lines, Supplementary text), with indicated EC50 values.
(C) Auxin regulates cell density. Cells were treated with a combination of IAA and blasticidin at different concentrations for four days and passaged once. Cells were counted by flow cytometry (n=3, error bar = standard deviation). In (B) and (C), the x axis uses a symmetric log (symlog) scale to include a value of 0.

Engineered mammalian cells produce auxin

In addition to sensing, a population control system requires that cells produce auxin at levels sufficient to trigger responses in receiving cells (Khakhar et al., 2016). Auxin can be synthesized in two enzymatic steps: (1) oxidation of L-tryptophan to indole-3-acetamide (IAM) and (2) hydrolysis of IAM to IAA (Figure 2.3A) (Sitbon et al., 1992). By itself, the second enzymatic step can produce either IAA or NAA from precursors IAM or 1-naphthaleneacetamide (NAM), respectively, enabling direct control of auxin production (Figure 2.4A) (Kawaguchi et al., 1991). To identify enzymes that efficiently catalyze this reaction, we compared thirteen indole-3-acetamide hydrolases from bacteria and plants (Figure 2.3B, left), transiently ex-

pressed them individually in Receivers, and measured their ability to downregulate AID-tagged mCherry fluorescence by flow cytometry.

Three enzymes reduced mCherry to levels comparable to that produced by addition of IAA itself (Figure 2.3B). Among these, we selected A. tumefaciens iaaH for further use. We stably integrated it in Receivers to create a Sender-Receiver cell line (Figure 2.4B). After 2 days of culturing Sender-Receivers in media containing the IAM precursor, the resulting conditioned media, diluted into an equal volume of fresh media for optimum cell growth, reduced mCherry-AID-BlastR to levels comparable to those generated by saturating concentrations of IAA (Figure 2.4C and 2.4D). The Sender-Receiver line was also able to produce the auxin NAA (which, as shown below, has some advantages compared to IAA) from its corresponding precursor NAM (Figure 2.4D), albeit with a diminished response compared to NAA, consistent with the higher EC50 of NAA compared to IAA. These results show that iaaH expression in mammalian cells can efficiently produce both auxins from corresponding precursors.

The ability to synthesize auxin in mammalian cells without exogenous precursors could facilitate in vivo applications of population control circuits. Bacterial auxin (IAA) synthesis pathways use a tryptophan 2-monooxygenase, iaaM (also known as aux1 or TMO), to synthesize IAM from L-tryptophan (Figure 2.3A). Sender-Receivers expressing iaaM (from P. savastanoi), cultured in media without precursors, produced auxin concentrations in conditioned media that were sufficient to degrade the auxin reporter in Receiver cells (Figure 2.3C). Furthermore, iaaM-expressing Receivers produced precursor in conditioned media, allowing auxin production by Sender-Receivers (Figure 2.3C). These results demonstrate the two step iaaM-iaaH auxin synthesis pathway can operate in mammalian cells without exogenous precursors. However, in the following experiments, we used iaaH with added precursors to allow external control of auxin synthesis.

To estimate the spatial range of auxin signaling, we seeded a field of Sender-Receivers adjacent to a larger region of Receivers, and applied media in a 1.5% agarose gel to prevent non-diffusive transport (Methods, section A.3). After 48 hours, mCherry fluorescence was reduced in Receivers proximal to the Sender-Receiver region, forming a long-range gradient. This effect occurred in the presence, but not the absence, of the IAM precursor, consistent with a dependence on IAA production (Figure 2.4E, bottom). Image analysis revealed auxin response in Receivers declining to half its maximum value at a length scale of 15.6±0.85

millimeters (Methods, section A.3), or approximately 750 cell diameters, from the source region, within 48 hours (Figure 2.4E, top), consistent with expectations for a small molecule the size of auxin diffusing in buffered solutions (Robinson, J. E. Anderson, and J. L. Lin, 1990) (Figure 2.3D and Supplementary text 2.5). These results show that auxin can provide information about global cell density over an extended region, and further demonstrate the possibility of using auxin as a synthetic gradient-forming morphogen for applications in synthetic developmental biology (Teague, Guye, and Weiss, 2016).

Engineered cells sense and respond to population density

We next asked whether Sender-Receiver cells could sense their own population density by producing auxin at a rate proportional to population size and sensing auxin concentration in the local environment (quorum sensing, Figure 2.1, middle panel).

We analyzed the dependence of reporter expression on cell population density in the presence of concentrations of each precursor (NAM and IAM) sufficient to generate saturating levels of auxin. With the NAM precursor, reporter fluorescence decreased in response to increasing population density (Figure 2.5A, darker blue line). By contrast, addition of IAM generated a strong, but density-independent, decrease in fluorescence (Figure 2.5A, light blue line). We reasoned that this density-independence could result from IAA's limited membrane permeability at neutral pH (Raven, 1975), potentially causing newly synthesized IAA to accumulate intracellularly. Though the rate of exchange is sufficient to fully induce receiver cells (on the timescale of days, Figure 2.4D and 2.4E), it could be insufficient to prevent intracellular accumulation of IAA due to the faster rate of auxin production by iaaH (typically several μM per minute) in the same cell (Mishra et al., 2016). To overcome this issue, we stably expressed the auxin exporter PIN2 from Arabidopsis thaliana (Petrásek et al., 2006) in Sender-Receiver cells (Sender-Receiver-PIN2 cells) (Figure 2.5B, right). PIN2 expression produced a modest decrease in auxin sensing, suggesting the transporter was functional (Figure 2.6A), but allowed quorum sensing across most of the full dynamic range of auxin concentration sensing (Figure 2.5B and 2.6B). Cells responded similarly to cell density across different culture media volumes, indicating quorum sensing responded to cell density rather than absolute cell number (Figure 2.5C). Together, these results establish that Sender-Receiver cells can sense their own population density in two ways: using NAM without PIN2, or using either precursor with PIN2.

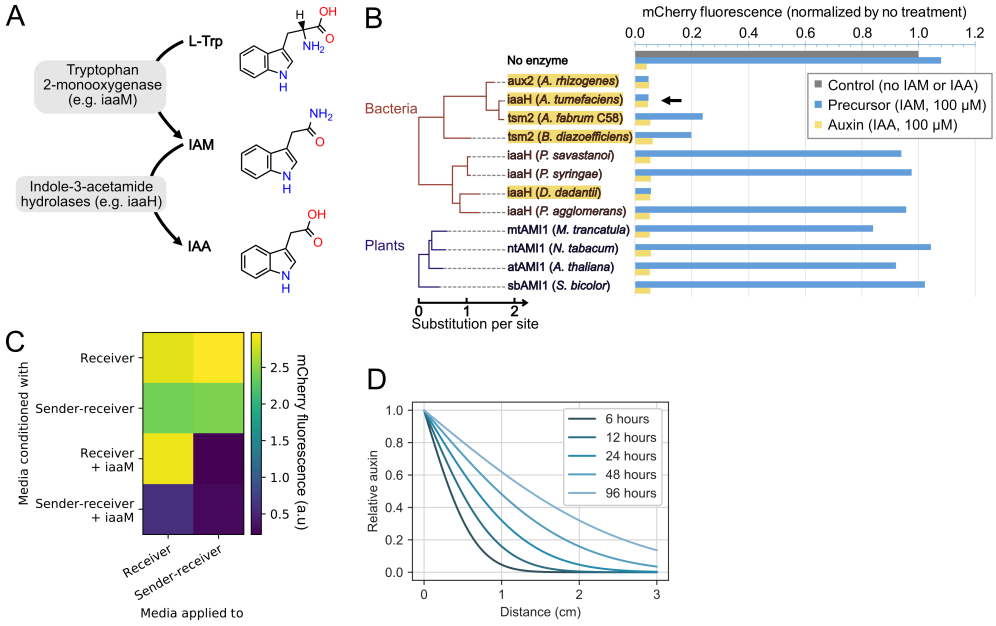


Figure 2.3: Expression of enzymes from agrobacterium enables cells to produce auxin.
(A) IAA can be synthesized from L-tryptophan (L-Trp) in a two-step pathway. (B) Twelve candidate indole-3-acetamide hydrolases were screened for the ability to convert IAM to IAA. The phylogenetic tree was created with CLUSTAL-OMEGA (Sievers et al., 2011) and phyML3.0 (Guindon et al., 2010) using default settings. Candidate enzymes were transfected into Receiver cells, and auxin production was assessed by mCherry-AID degradation (Figure 2.4A) after 48 hours. The fluorescent results are normalized to the control without auxin. Candidates that could down-regulate the mCherry by over 50% are highlighted in yellow. iaaH from A. tumefaciens (black arrow) was selected for subsequent experiments.
(C) Media conditioned by iaaM (from P. savastanoi) expressing cells contains substrate for auxin production by iaaH. Cells transfected with iaaM were used to condition the media for 48 hours, as shown in Figure 2.4C. Conditioned media was combined with an equal amount of fresh media and then applied to reporter cells for 48 hours before flow cytometry.
(D) Auxin diffusion observed in Figure 2.4E matches theoretical predictions (see also Supplementary text 2.5).

Linking quorum sensing to cell survival could enable cell population control (Figure 3.1, right). To test this possibility, we seeded cells at different densities in media containing blasticidin to make cell growth dependent on BlastR levels, as well as the precursor IAM. After 4 days, cells seeded at high, but not low, densities exhibited reduced cell numbers compared to cells plated at the same density without IAM

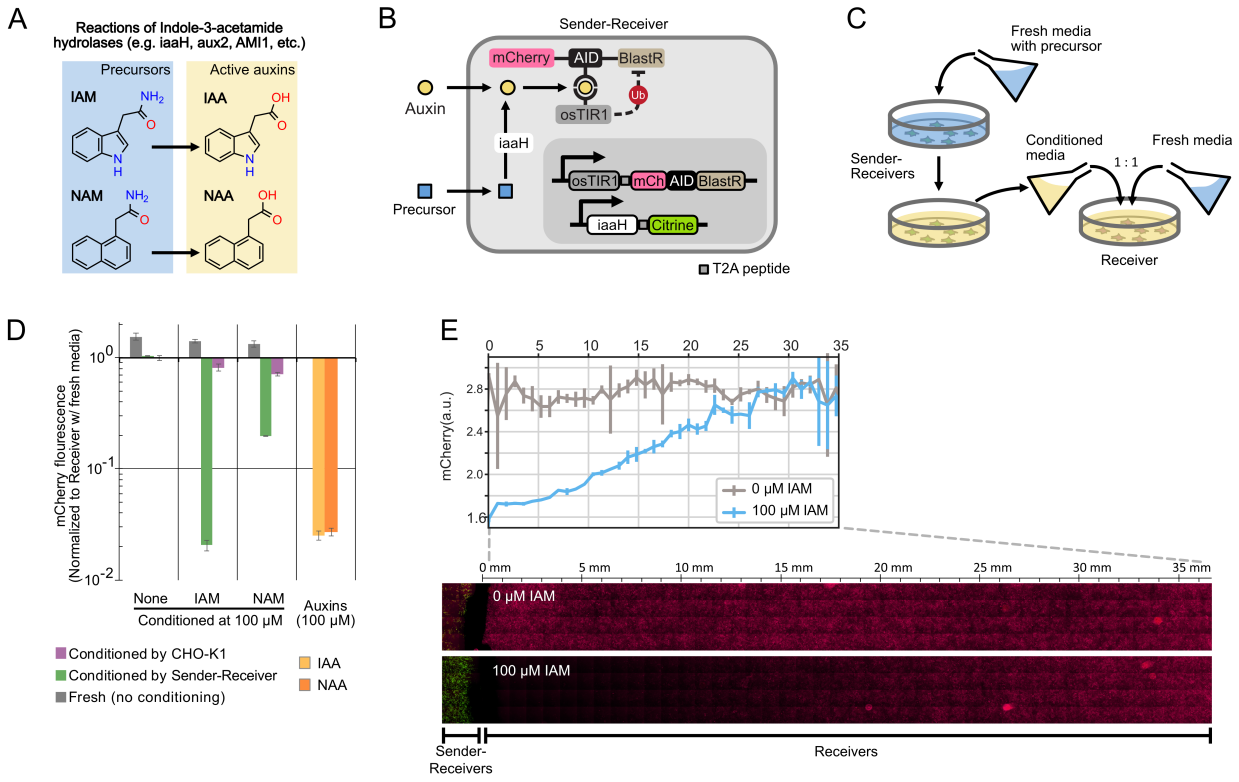


Figure 2.4: Sender-Receiver cells produce to auxin.
(A) Indole-3-acetic acid hydrolases such as iaaH, aux2, and AMI1 hydrolyze inactive auxin precursors (IAM and NAM) to their respective active form (IAA and NAA).
(B) Stable expression of iaaH in Receiver cells allows them to produce auxin from precursors (blue square).
(C) Conditioned media experiment (schematic). Fresh culture media with or without precursors was added to plated sender cells, collected, mixed at 1:1 ratio with standard fresh media, and then applied to receiver cells.
(D) iaaH can produce IAA and NAA from IAM and NAM precursors, respectively. Media with or without precursors were conditioned by Sender-Receivers or WT CHO-K1 cells for 48 hours and applied to Receiver cells. Receivers cultured with fresh media were also assayed as controls. Data are normalized to Receiver cell fluorescence treated with media conditioned by WT cells. n=3.
(E) Auxin senders can generate an auxin gradient. Sender-Receivers (green) were seeded within a 7x7 mm square at the edge of a 60 mm dish, and Receivers were plated everywhere else (Receiver region). One day after plating, the media was replaced with fresh media containing low-melting-point agarose, with or without IAM (Methods, section A.3). Plates were imaged after 48 hours. Inset on top: Quantification of the average pixel intensity of mCherry expression in cells. Error bars denote standard deviation of the four images of each column.

or IAA (Figure 2.5D). These results demonstrate density-dependent control of cell survival.

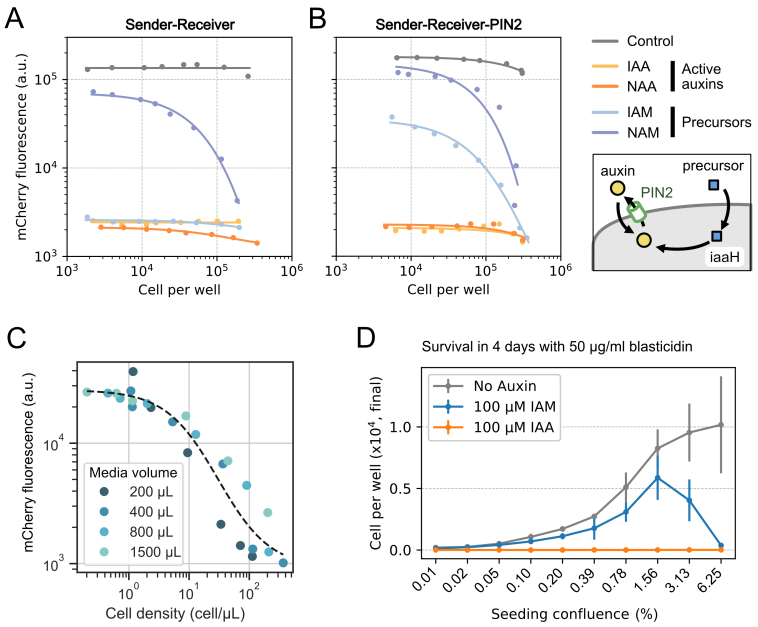
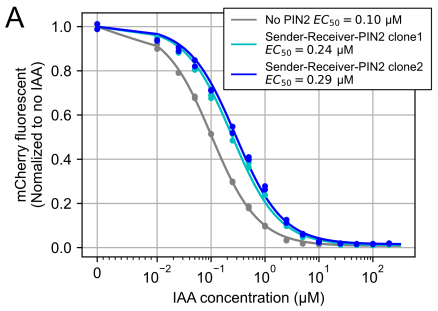


Figure 2.5: Sender-Receiver cells sense their population density and regulate survival accordingly.
(A-B) Sender-Receiver and Sender-Receiver-PIN2 cells perform quorum sensing. For both plots, cells were seeded at different densities and induction conditions, with either of the two auxins (saturating signaling) or their precursors (to allow quorum sensing). mCherry fluorescence was assayed after two days as a reporter of auxin sensing. Inset: Cells in B express the transporter PIN2, which actively exports auxin. Data were fitted onto an inverted Michaelis-Menten's function on log scale (see Figure 2.6B for fitted parameters, Supplementary text).
(C) Sender-Receiver-PIN2 cells sense population size per unit volume. Cells were grown for two days at 6 different densities for each media volume, and cultured on a rocker for better mixing.
(D) Sender-Receiver-PIN2 triggers cell death at high population density. Cells were seeded at different confluence levels, grown for four days in media with 50 µg/ml blasticidin and IAM, IAA, or no auxin. Error bars indicate standard deviation from triplicates.

2.3 Multicellular paradoxical circuit achieve robust population control

The circuit with quorum sensing linked to cell survival (section 2.2) presents a negative autoregulatory feedback loop, resembling a similar design in bacterial systems,



		Dynamic range (fold)			
		IAA/IAM		NAA/NAM	
	PIN2	-	+	-	+
a	Ideal dynamic range	54.9	86.2	63.6	78.7
b	Loss to self-sensing	51.3	5.2	2.0	1.2
c	Loss to saturation	0.8	1.2	0.9	2.5
d	Actual dynamic range	1.3	16.8	52.0	34.7

Figure 2.6: Sender-Receiver cells sense their population density.
(A) The right-shifted auxin response curve of PIN2 expressing cells is consistent with active export of IAA from the cells. Two PIN2-expressing cell lines were incubated with IAA for 48 hours and then assayed for mCherry fluorescence. The data were fitted with the same method in Figure 2.2B.
(B) Quantitative analysis of Figure 2.5A and B shows that the expression of PIN2 rescues dynamic range (indicated by the double-sided arrows in the left panel) lost to self-sensing (Supplementary text).

where a quorum sensing signal induces cell killing. The bacterial circuit was found to be susceptible to "cheater" mutations that allowed cells to escape control and grow to the limit of environmental capacity (Balagadde, 2005). In this section, we will demonstrate that similar problem was also encountered in our mammalian circuit, and we constructed a circuit with paradoxical architecture with improved robustness against cheaters.

Cells with simple negative feedback circuit are susceptible to cheaters

Slightly different from bacteria, for adherent mammalian cells, environmental capacity is limited by the surface area of the culture plate. We therefore define an escape event as cells growing to confluency (covering the full surface of the plate) and remaining confluent for the duration of the experiment. When we cultured the Sender-Receiver-PIN2 cell line in media containing blasticidin and IAM to activate the circuit, we observed escape after 16 days of culture (Figure 2.7A, Movie S1[1]), which reflects the acquisition of a "cheating" phenotype, with which cells are able to proliferate in high auxin (IAA) and blasticidin (Figure 2.7B). These results show that simple negative feedback population control circuits in mammalian cells, like their bacterial counterparts, are susceptible to selection pressure for cheater mutations.

[1]see section 2.5

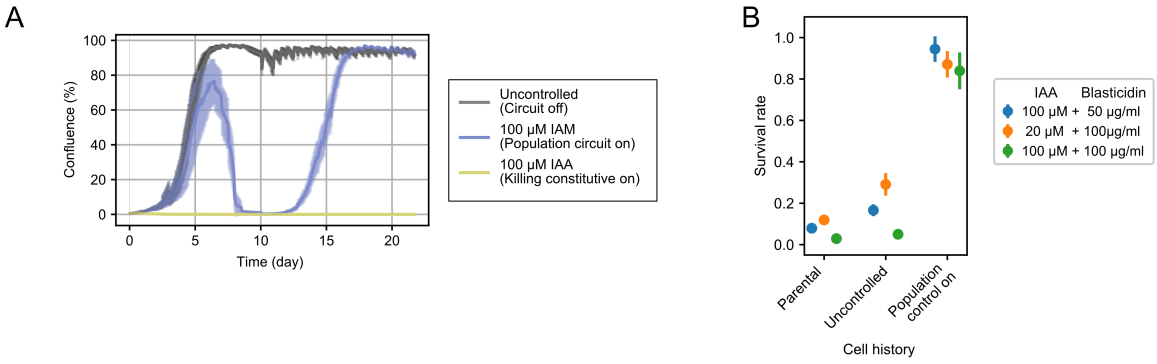


Figure 2.7: Cells with simple negative feedback circuit is susceptible to cheaters. **(A)** A pilot experiment showed that the Sender-Receiver-PIN2 cells escape regulation after 15 days of continuous culture. Cells were seeded into a 24-well imaging plate with 50 µg/ml of blasticidin solely (Circuit off) or also with 100 µM IAA or IAM (Circuit on). Each trace reflects an average of 12 positions in a sample well. Shaded envelopes represent the standard deviation of the averages.
(B) Cheater cells collected from the end of the movie in Figure 2.6A have high survival rates when grown in IAA with blasticidin for 4 days as compared to parental cells. Cell counts measured with flow cytometry were normalized to matched cells grown without drugs. Bars represent standard deviation from triplicates.

The "Paradaux" cells implement paradoxical population control

A "paradoxical" circuit architecture, in which a quorum sensing signal either stimulates or inhibits both cell proliferation and cell death, can make population control more mutationally robust by counterselecting against cheaters that lose signal sensing (Karin and Alon, 2017). In the paradoxical circuits, positive net growth can occur only at signal concentrations lying between two non-zero bounds, similar to the Allee effect in ecology (Courchamp, Berec, and Gascoigne, 2008) (Figure 2.8A, blue regions). By comparison, the simpler, non-paradoxical negative feedback architecture, in which signals only down-regulate growth, exhibits positive net growth at all signal concentrations below a single upper bound (Figure 2.8A, lower left). While both circuits exhibit stable fixed points at the maximum of their positive growth zones, they respond differently to cheater mutations that reduce auxin sensitivity (the predominant cheater phenotype observed in Figure 3.5C and D). Specifically, with negative feedback alone, such cheater mutations extend the regime of positive net growth to higher signal concentrations and cell densities, providing a growth advantage over non-mutant cells (Figure 2.8A, lower right). By contrast, in the paradoxical circuit, the same mutation would cause signal sensing to drop be-

low the lower bound, activating the cell death arm of the circuit (Figure 2.8A upper right). In this way, the paradoxical circuit design should suppress escapes by cheater mutations that reduce or eliminate signal sensing.

We designed a paradoxical circuit in which auxin represses both proliferation and death, a configuration that is well-suited to the inhibitory nature of auxin-dependent protein degradation (Figure 2.8A, upper left). In addition to auxin regulation of BlastR, we added a parallel regulatory pathway in which auxin negatively regulates cell death by inducing degradation of iCasp9, an ectopically expressed master regulator of apoptosis (caspase 9) activated by the small molecular dimerizer AP1903 (Figure 2.8B) (Straathof et al., 2005; Di Stasi et al., 2011). We added an AID domain in iCasp9 to provide auxin regulation, and fused it to the monomeric GFP variant mGFPmut3, to allow direct readout of its concentration (Landgraf et al., 2012) (Figure 2.8B). This design allows one to operate the same cell line in three regimes, depending on what combinations of blasticidin and AP1903 are added to the medium. With only blasticidin, the circuit operates in the pure negative feedback regime; with blasticidin and AP1903, it operates in the paradoxical regime; and with neither inducer it has unregulated growth. Because it implements paradoxical regulation through auxin, we dubbed this circuit "Paradaux."

To encode the Paradaux circuit, we designed a single multi-protein construct expressing osTIR1, the auxin-regulated iCasp9 system, and the auxin-regulated BlastR construct described above (Figure 2.8C). To eliminate one potential mechanism for evolutionary escape, we positioned BlastR at the C-terminal end of the construct, so that premature stop codon mutations would deactivate BlastR, decreasing survival. We integrated the construct to create stable monoclonal CHO-K1 cell lines for further analysis (Methods, section A.1). Because this line lacks PIN2, we used NAM/NAA rather than IAM/IAA for all Paradaux experiments. Survival of these cells increased monotonically with auxin in the presence of AP1903 and decreased with auxin in the presence of blasticidin, demonstrating that both branches of the Paradaux circuit were individually functional (Figure 2.9A, red and green lines). (An additional monoclonal line with an independent integration of the same circuit is shown in Figure 2.10A). Further, including both blasticidin and AP1903 produced a biphasic survival curve (Figure 2.9A, blue dots), the key requirement for paradoxical population control.

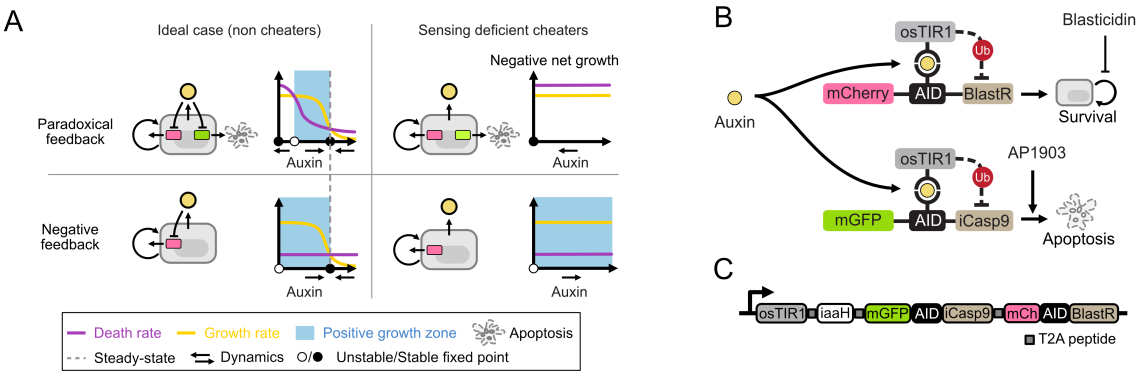


Figure 2.8: Paradoxical architecture reduces susceptibility to cheater mutations. **(A)** In the paradoxical architecture, the same signal inhibits growth (red pathway) and death (green pathway). This can produce a window of auxin concentrations leading to positive net growth (light blue region). Without mutation, the paradoxical and negative feedback circuits operate similarly around a stable equilibrium point of large population size (solid black dots, left panels). Mutations that eliminate sensing make both death and growth independent of auxin concentration (right panels), which selects against mutations in the paradoxical circuit due to negative net growth. **(B)** In the paradoxical circuit implementation, auxin regulates growth through BlastR (upper path), and also regulates apoptosis via iCasp9 (lower path), each with distinct fluorescent protein readouts and a small molecule (blasticidin and AP1903) as a control switch. **(C)** The full paradoxical circuit can be encoded as a single open reading frame, with distinct proteins separated by T2A peptides (grey squares).

Mathematical modeling identifies parameter regimes required for paradoxical control

To identify parameter regimes, including concentrations of blasticidin and AP1903, that optimize its population control capabilities, we developed a mathematical model of the Paradaux circuit. Assuming rapid intracellular-extracellular auxin equilibration, as observed for NAA (Figure 3.6A), and timescale separation between fast intracellular dynamics and the slower cell population dynamics (Figure 2.10B, Supplementary text), we derived an approximate model based on two differential equations. The first represents the extracellular auxin concentration shared by all cells, denoted A. The second describes the cell population size, denoted N, which ranges from 0 to the environmental carrying capacity, normalized as 1 (Supplementary text).

We assume auxin is produced at a constant rate per cell, λ_A, and diluted by periodic media changes, approximated as a continuous process with rate constant δ_A:

$$\frac{dA}{dt} = \lambda_A \cdot N - \delta_A \cdot A \tag{2.1}$$

We also assume cell growth can be described by a generalized logistic function (Richards 1959) (Figure 2.10C, Supplementary text), modified to incorporate the effects of blasticidin, B, and iCasp9, I, on growth:

$$\frac{dN}{dt} = R_g(B, I, A) \cdot N \cdot (1 - N^{\nu}) \tag{2.2}$$

Here, R_g denotes the cellular growth rate, a function that can be written as a sum of two Hill-like terms representing the combined auxin dependent effects of blasticidin and iCasp9 on cell survival (See Equation 2.15, 2.16 and 2.17for the exact form). The exponent ν is the non-linear correction parameter in the generalized logistic growth function (Richards, 1959).

To constrain the effective biochemical parameter values in the model, we incorporated experimentally measured values of the sensitivity of AID-tagged proteins to auxin, the unperturbed cell growth rate, the auxin secretion rate, and the non-linear growth correction parameter (Figure 2.2B; Figure 2.10C and 2.10D). Remaining parameters were fit using the auxin-dependent survival rates measured with either AP1903 or blasticidin (Figure 2.9A, red and green line). The model initially overestimated actual growth rates when both arms of the circuit were simultaneously active (Figure 2.9A, dashed purple lines), possibly due to previously reported synergy between apoptosis and blasticidin-dependent translational inhibition (Holcik and Sonenberg, 2005). We therefore added a phenomenological synergistic interaction term to the growth rate expression (Equation 2.19 and Supplementary text; Table 2.1, parameter set 1; Figure 2.9A, dotted line). Finally, we checked that experimental parameter values expected to be independent of integration site, such as maximum cell death and growth rates, as well as synergy and Hill coefficients, agreed, within 2-fold, with those measured for a second Paradaux monoclonal cell line with an independent integration of the circuit construct (Figure 2.10A; Table 2.1, parameter set 2).

Before initiating challenging long-term analysis of population control, we used the model to systematically scan for AP1903 and blasticidin concentrations likely to

favor paradoxical population control. For each pair of AP1903 and blasticidin concentrations, we classified the dependence of cell survival on auxin into one of five qualitatively distinct behaviors: (1) positive net cell growth across all auxin concentrations (Figure 2.9B, purple background, "uncontrolled growth" in 2.11); (2) negative net cell growth across all auxin concentrations (2.9B, green background); (3) permissive growth, in which cells proliferate only beyond a minimum auxin concentration (2.9B, pink background); (4) simple negative feedback control, analogous to the Sender-Receiver behavior (2.9B, yellow background, "negative feedback in 2.11"); and (5) paradoxical control (2.9B, blue background, "paradoxical feedback in 2.11"). Only the negative and paradoxical feedback regimes produce a non-zero, stable fixed point, allowing population control. The desired paradoxical regime occurred in a window of blasticidin and AP1903 concentrations centered around 50 μg/ml and 50 nM, respectively (2.9B). Further, this paradoxical window could be enlarged by optimizing the expression levels of iCasp9 and BlastR (2.9C). Within the paradoxical regime, the circuit produced the expected bistability of population size (Figure 2.10E) and robustness to mutations that eliminate auxin sensing, simulated by setting sensed auxin to zero (2.9D).

Additionally, we explored the effect of the 2-3 day time delay required for blasticidin to kill sensitive cells (Sato et al., 2012) (Figure 2.10F). Time delays in negative feedback loops are known to produce oscillations under some conditions (Potvin-Trottier et al., 2016; Elowitz and Leibler, 2000). In simulations of the paradoxical circuit, inclusion of a time delay led to oscillations of population density with periods of 2 weeks or more depending on the value of the delay parameter, τ (Equation 2.22). A bifurcation (Martinez-Corral et al., 2018) between damped and sustained oscillations occurred at approximately $\tau = 48\ hrs$ (Movie S2[2]). Taken together, these results provided insight into parameter dependence and expected dynamics of the circuit, and identified specific AP1903 and blasticidin concentrations for long timescale analysis of the experimental circuit.

Paradoxical circuits extend the duration of population control

To experimentally analyze population control, we continuously monitored cultures of the Paradaux cell line using time-lapse movies (Figure 2.11B). We performed three sets of time-lapse experiments, the longest of which lasted for 42 days (an arbitrary time scale constrained by technical limitations). Cells were cultured in regimes of uncontrolled growth (no blasticin or AP1903, the "always positive" regime in

[2]see section 2.5

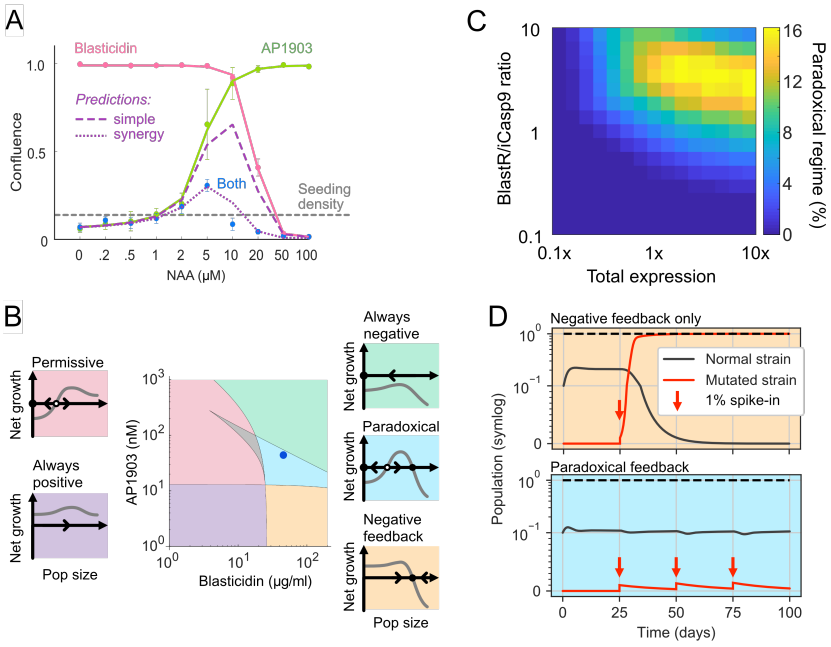


Figure 2.9: Paradaux cells are predicted to be robust to cheaters.
(A) Paradaux cells respond to auxin in a biphasic manner. Cells were seeded at 1/8 confluence (grey dashed line) and pretreated with NAA (24 hours), then treated with combination of NAA, blasticidin (20 μg/ml) and AP1903 (50 nM) (72 hours). Pink, green, and blue dots show the mean and standard deviation in the presence of AP1903, blasticidin, or both, respectively (n=3). Solid red and green lines: fits to the model. Purple dashed lines: predictions for the fully circuit based on the green and red curves. Dotted purple line: prediction with the synergy term.
(B) Different classes of behavior can occur in different parameter regimes. We simulated auxin-dependent growth in different regimes and identified five distinct regimes (insets). We numerically analyzed and sorted growth curves for each concentration of blasticidin and AP1903 (central plot). Blue dot indicates the inducer concentrations used in the time-lapse movie analysis (2.11). The grey region indicates curves that not classified into one of these categories (0.68% of total, see Figure 2.9)C and Supplementary text).
(C) Optimizing the expression and ratio of BlastR and iCasp9 widen the paradoxical regime. For each expression level, we analyzed the percent of blasticidin-AP1903 concentrations that generate paradoxical behavior, similar to panel E.
(D) Dynamic simulations show the Paradoxical Control circuit provides evolutionary robustness. For negative feedback system, $\beta_C = \beta_{syn} = 0$. Mutated strains were simulated with A fixed to zero, as sensing deficient mutations. On day 25, 50, 75 and 100, mutants were introduced into the system at 1%. These mutants take over in the negative feedback circuit (top) but not the paradoxical circuit (bottom). Dashed line: carrying capacity.

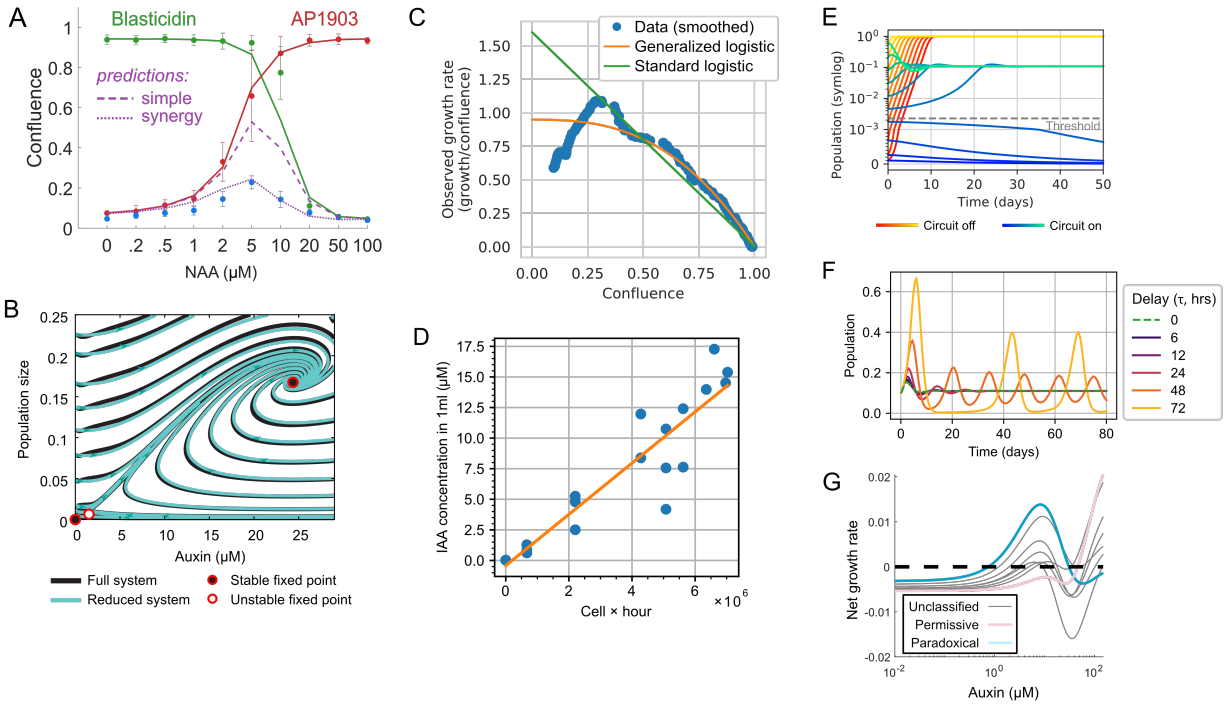


Figure 2.10: Model and parameter fitting of the paradoxical circuit.
(A) A second Paradaux line responds to auxin in a biphasic manner. Condition is same as Figure 2.9A, except 4 nM of AP1903 and/or 50 μg/ml blasticidin were used. Curves are fitted directly from the survival data due to in-availability of the maximum growth rate (Table 2.1).
(B) The complete and reduced (approximate) versions of the model show similar dynamics, (overlap of black and cyan trajectories, Supplementary text).
(C) Generalized logistic growth (Equation 2.2) fits PC1's growth curve better. Data are from the first 120 hours of the uncontrolled growth sample in movie set 1 (Figure 2.11B). Smoothed with a Gaussian filter. $\alpha = 0.0395/hr$; $\nu = 2.26$.
(D) Auxin secretion rate ($\lambda_A = 2.09\mu M/10^6 cell \cdot hr$. Sender-Receivers were seeded at 10,000 cells per well. After different durations (0 to 73.5 hours), conditioned media was collected and cells were counted to estimate exponential growth rate. The conditioned media was applied to Receivers at 1X, 0.5X, or 0.25X dilution, and then Receiver fluorescence was compared to a standard curve.
(E) Dynamic simulation of the system shows biphasic growth. For each initial cell density, the initial auxin concentration was determined by setting Equation 2.1 to equal zero (equilibrium for auxin).
(F) Simulation with delay (τ, Equation 2.22) incorporated show oscillated behaviours. Cells were seeded at 0.1, with delay applied to the blasticidin related response.
(G) Sub-sampling of the "unclassified" curves from Figure 2.9B shows that they are intermediate between "permissive" and "paradoxical" categories.

Parameter	Description	Fitted value	Unit
A	Production rate of auxin at 100% confluence	4.18	μM/hr
δ_A	Dilution rate of auxin	0.0289	hr^{-1}
ν	Generalized logistic growth coefficient	2.69	DL
α	Maximum natural cell proliferation rate	0.0395(1); 0.0327(2)	hr^{-1}
β_B	Maximum net growth reduction by blasticidin	0.0874(1); 0.147(2)	hr^{-1}
β_C	Maximum death rate induced by iCasp9	0.0451(1); 0.0375(2)	hr^{-1}
β_{syn}	The synergetic coefficient	1.12(1); 1.70(2)	hr^{-1}
κ_A	Effective strength of auxin	1.316	μM^{-1}
κ_B	Effective strength of blasticidin	0.045(1); 0.069(2)	$(\mu g/ml)^{-1}$
P_R	Effective expression rate of BlastR	1.77(1) ;0.716(2)	DL
$\kappa_I^{-n_2} \cdot P_C$	Effective expression rate of iCasp9	8.54e-4(1);1.23(2)	DL
n_1	Cooperativity of the blasticidin regulation	2.26(1); 2.28(2)	DL
n_2	Cooperativity of the iCasp9 regulation	1.41(1); 1.67(2)	DL

Table 2.1: List of parameters and fitted values for model of the Paradaux cells. Fitted values are labeled as (1) or (2), in cases where the major Paradaux line (labeled as 1, Figure 2.9A), and the one used from confirming the model (labeled as 2, Figure 2.10C) were fit with different values. Dimensionless values are labeled as "DL" in the "Unit" column.

Figure 2.9B), negative feedback (blasticidin added), or paradoxical feedback (both blasticidin and AP1903). In the uncontrolled regime, cells grew to and remained at full confluence for the duration of the movie (Figure 2.11A, Movie S3[3]). In the latter two regimes, population dynamics exhibited oscillations, consistent with models incorporating time delays for blasticidin-dependent cell killing (Figure 2.10F).

The negative feedback regime limited population size for 1 or 2 oscillatory periods (20-30 days), after which cells escaped control (exceeded >95% confluence) for the remainder of the movie (at least 10 subsequent days, 2.11B). This behav-

[3]see section 2.5

ior is consistent with the experimentally observed susceptibility of simple negative feedback control to escape mutants (Figure 2.13C and D), and simulated responses to introduction of sensing mutants (Figure 2.9D). By contrast, in the paradoxical regime, cultures exhibited oscillations in population density, but remained at sub-saturating densities for significantly longer ($p<0.005$, Kaplan-Meier estimate, log-rank test (Kishore, Goel, and Khanna, 2010)). In fact, over half of the cultures remained sub-confluent for the full duration of the movies (Figure 2.11A-C; Movie S3[4]). More specifically, across the three movie sets and the 11 individual cultures in the paradoxical group, only one culture escaped control before the 32 day mark (Figure 2.11A, Movie set 3). Together, these results demonstrate that the paradoxical control circuit reproducibly extends the duration of population control.

Phenotypical and expression analysis indicates paradoxical circuit prevent cheating

To gain insight into how cells evolved during long term culture, we isolated cells at the end of each of the 30 individual movies, passaged them in standard media, and assayed their behaviors under different conditions (Figure 2.12A-E). The fluorescence of mCherry and mGFP, which are co-expressed with BlastR and iCasp9, respectively, varied substantially across all isolates, reflecting independently acquired changes in the expression of circuit components, rather than amplification of a common founder mutation already present in the parental cell line. However, variation of the two reporters was strongly correlated expression across isolates, suggesting that mutations or other adaptations occurred upstream of both circuit arms (Figure 2.12A).

The two population control conditions produced different effects on circuit component expression and phenotypic behavior. In cells selected under negative feedback conditions, basal BlastR expression was upregulated, on average, by 73%, and showed 49% less responsiveness to auxin compared to the uncontrolled group (Figure 2.12B and C; $p = 0.014$ and $p<0.001$, respectively). By contrast, isolates from the paradoxical feedback conditions showed significantly lower basal BlastR expression ($p = 0.008$) and retained a larger dynamic range of BlastR regulation compared to negative feedback isolates ($p = 0.007$). These differences in BlastR regulation were also reflected in cell survival (Figure 2.12B and C). When cultured in a combination of blasticidin and NAA, mimicking high cell density, negative feedback isolates exhibited increased survival compared to the uncontrolled group ($p<0.001$

[4]section 2.5

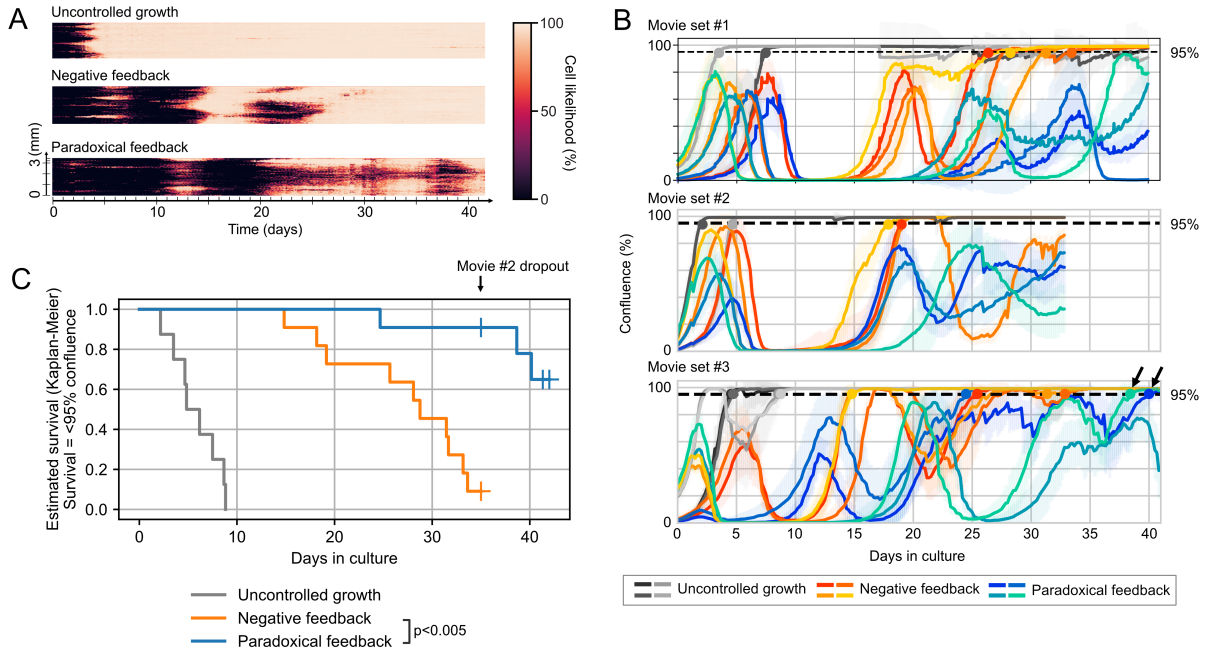


Figure 2.11: The paradoxical circuit allows mutationally robust population control. **(A)** Composite kymograph of long-term cultures with no control (100 µM NAM; upper panel), negative feedback (100 µM NAM and 50 µg/ml blasticidin; middle panel), or paradoxical feedback (100 µM NAM, 50 µg/ml blasticidin, and 50 nM AP1903; lower panel), from movie set 3. For visualization, the images were analyzed using ilastik and 30-pixel-wide strips from each timepoint were combined to make the kymograph.
(B) Population dynamics for the three movie sets conditions reveal delayed mutational escape for the paradoxical circuit (shaded envelopes represent the standard deviation across 12, 25, and 36 stage positions, for each movie set respectively). Solid dots indicate escape events, defined by cells exceeding 95% confluency and not returning below that threshold for the duration of the movie. Black arrows indicate late cheating isolates that are similarly denoted with arrows in Figure 2.12D.
(C) Kaplan-Meier estimate (Kishore, Goel, and Khanna, 2010) of survival (no mutant escape) for movies in (B). Samples in movie set 2 that ended earlier were treated as dropouts. The samples under paradoxical feedback retain population control significantly longer than those under negative feedback ($p<0.005$, log-rank test).

Figure 2.12D). While individual paradoxical isolates also exhibited elevated survival, the increase compared to the uncontrolled condition was not significant at the group level ($p=0.085$). We also note that the two isolates which were classified as cheaters in the final days of movie set 3 (Figure 2.11B, black arrows) were unable to survive in this condition (Figure 2.12D, pink dots with arrows), suggesting they could potentially have returned to lower cell density had the movie duration been

longer. Together, these results suggest that paradoxical conditions preserved more of the original BlastR regulation and function compared to negative feedback conditions.

The paradoxical architecture is designed to suppress sensing deficient mutations that reduce the responsiveness of both arms of the circuit to auxin. More specifically, cells that lose auxin sensing should no longer degrade BlastR (Figure 2.12I, lower right), and therefore obtain a growth advantage. By contrast, in the paradoxical circuit, such sensing deficient cells would also be unable to degrade iCasp9, leading to their elimination (Figure 2.12F, upper right). Thus, sensing deficient isolates from the negative feedback condition should become susceptible to killing through the iCasp9 arm, even in the presence of auxin. In fact, isolates from negative feedback conditions were sensitive to the combination of AP1903 and auxin, indicating that they had acquired the potential to be counter-selected under paradoxical conditions (Figure 2.12E, $p<0.001$). Thus, cells with reduced sensing gain a growth advantage in negative feedback conditions but are counter-selected under paradoxical conditions, consistent with the principle of paradoxical control.

To better understand the heritable changes that drive adaptation during long term culture, we sequenced the circuit from endpoint isolates from the first two movie sets. Several mutations appeared at elevated frequencies, occurring either in the AID domain fused to BlastR, consistent with positive selection through reduction of AID activity, or in osTIR1, near the 5' end of the circuit transcript, potentially reducing expression of all circuit components (Figure 2.13A; Table 2.2). However, in most isolates, no mutations occurred at significantly elevated frequencies, suggesting that most relevant adaptations (observed in Figure 2.12A-E) lie outside the circuit itself.

Therefore, we performed RNA-Seq expression analysis to investigate global transcriptional changes. Expression profiles from isolates cultured in the same conditions generally clustered together, suggesting that each operating mode of the circuit selects for a distinct set of gene expression changes (Figure 2.13B). In addition, solates from the uncontrolled growth conditions clustered with the parental cells, as expected. Pathway enrichment analysis further revealed that ribosome components, the primary targets of blasticidin, were significantly upregulated in both negative feedback and paradoxical conditions (Figure 2.13C). Proteasome components were up-regulated only in negative feedback conditions (Figure 2.13C and D), consistent with opposite selection pressures on proteasome activity generated by auxin regulation of BlastR and iCasp9 in the paradoxical condition.

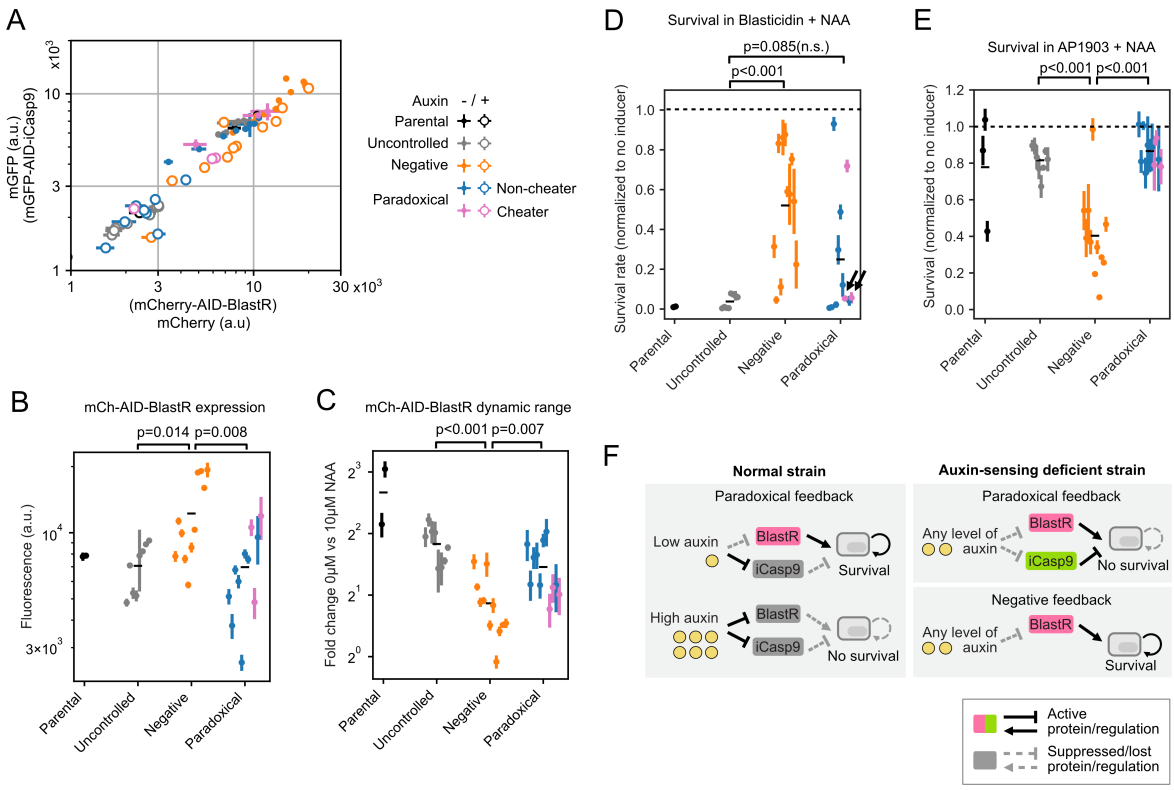


Figure 2.12: Phenotypes of isolates indicate paradoxical circuit prevent cheating.
(A-C) Isolates were treated with 10 μM NAA, or nothing, for two days prior to flow cytometry assay for mCherry and mGFP fluorescence, co-expressed with BlastR and iCasp9 respectively. Bars represent standard deviation from triplicates (A and B), or bootstrapping (C, n=9). (A) Isolates from all long term cultures show correlated BlastR and iCasp9 expressions. (B) Isolates from negative feedback, but not paradoxical feedback, showed upregulation in BlastR, compared to control. (C) Isolates from negative feedback showed diminished dynamic range compared to both control and paradoxical groups.
(D-E) BlastR and iCasp9 expression levels affect survival rates across different conditions. Cells were seeded in 96-well imaging plates with IAA or standard media for one day, and the second drug, blasticidin (D) or AP1903 (E), was added. Samples were then imaged to estimate confluency at day 4. For each isolate, survival rates were normalized to the group with no second drug. The black arrows in (D) highlight the isolates that cheated at the final days in movie set 3, black arrows in Figure 2.11B. Values and errors were calculated by bootstrapping (n=6, Methods, section A.6).
(F) How the paradoxical architecture, but not negative feedback, eliminates cells that lose the ability to sense auxin (schematic).

2.4 Conclusions and discussions

Natural cytokine-based control circuits allow cells to regulate their own population dynamics, as well as those of other cell types. Synthetic circuits could provide anal-

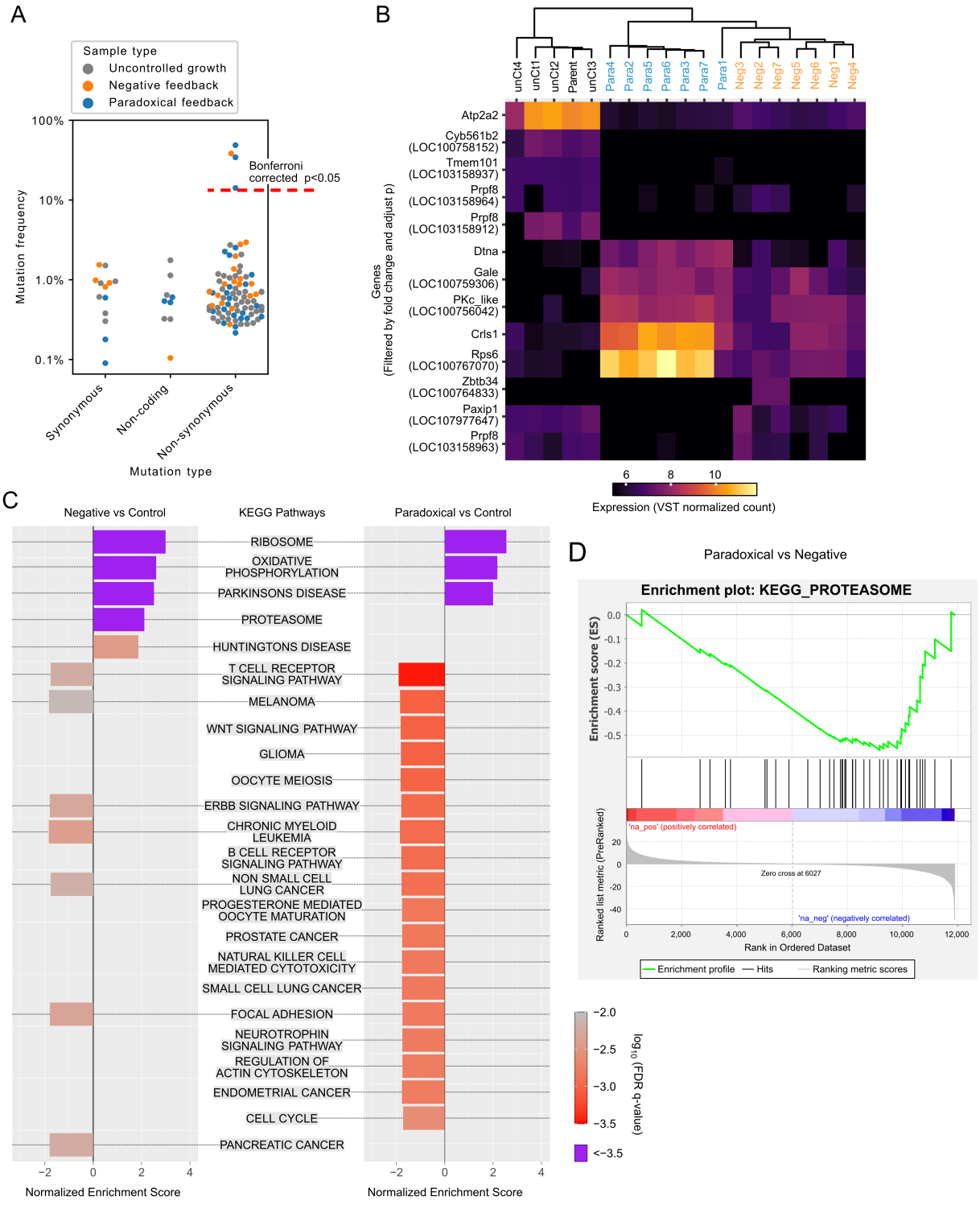


Figure 2.13: Targeted DNA and whole-genomic RNA sequencing reveal mutations and gene expression changes during long term culture. (continue on next page)

Figure 2.13: Targeted DNA and whole-genomic RNA sequencing reveal mutations and gene expression changes during long term culture. (continued from last page)
(A) Four mutations of significant positive selection were detected in the circuit component sequencing (details listed in Table 2.2). The integrated transgene was amplified and sequenced (Illumina). Mutations were detected by comparing reads of isolates to those of the parental line (chi-square test, Bonforroni corrected $p<0.05$, Method, section A.6). The frequency threshold for positive selection (red dashed line) was determined as the 5th percentile (Bonferroni corrected) of the distribution of synonymous mutations, assumed to follow normal distribution.
(B) Isolates that evolved in the same conditions clustered together, based on variance-stabilizing transformed (VST) read count across the whole transcriptome. Genes that are significantly differentially expressed between the groups (fold change > 32, adjusted $p<0.01$) are shown here (DESeq2, Love, Huber, and Anders, 2014). Genes that lack annotation in RefSeq were manually annotated (Method, section A.5), with the RefSeq number in parentheses.
(C) Pathway enrichment analysis shows that the negative feedback and paradoxical feedback affect different sets of pathways. Enrichment analysis was performed with Gene Set Enrichment Analysis' (GSEA) ranked method against MSigDB's curated version of KEGG canonical pathways (Kanehisa et al., 2021).
(D) The negative and paradoxical circuit design exert different evolutionary pressures on the proteasome pathway. Direct comparison of the KEGG proteasome pathway between the paradoxical and negative group was analyzed by GSEA, as shown with a standard GSEA plot. Here, blue (red) regions indicate genes whose expression is significantly greater (lower) in the negative feedback condition compared to the paradoxical condition.

ogous capabilities. To this end, we engineered simultaneous production and sensing of the plant hormone auxin in mammalian cells, and coupled it to genes controlling cell proliferation and death. The enzymes iaaH and iaaM, together with PIN2, allow cells to produce and export auxin. osTIR1 together with AID domains provide a simple, direct means of sensing auxin and coupling it to arbitrary protein targets. These components thus provide a long-range private communication channel, and enable the foundational property of quorum sensing (section 2.2). Coupling quorum sensing to cell survival opens up the possibility of creating population control circuits, provoking the question of what circuit architectures can provide robust, long-term control. Consistent with previous work (Karin and Alon, 2017), mathematical modeling showed that a paradoxical architecture, in which auxin inhibits survival mediated by BlastR and killing by iCasp9, can generate a range of qualitatively different behaviors and, in some regimes, suppress cheaters (Figure 2.8 and 2.9). To experimentally realize this capability, we constructed the "Paradaux" cir-

Isolate	Movie set	Mutations detected (FDR p<0.05)	Positively selected mutation (FDR p<0.05)	Position/Substitution/Frequency
Parental	N/A	N/A	N/A	
unCt1	1	1	0	
unCt2	1	3	0	
unCt3	2	34	0	
unCt4	2	18	0	
Neg1	1	9	0	
Neg2	1	2	0	
Neg3	1	0	0	
Neg4	1	1	0	
Neg5	2	6	1	in AID(BlastR) R29G (Fr=0.38)
Neg6	2	0	0	
Neg7	2	6	0	
Para1	1	4	0	
Para2	1	1	1	in osTIR1 R22X (Fr=0.48)
Para3	1	5	1	in AID(BlastR) Y28C (Fr=0.34)
Para4	1	0	0	
Para5	2	5	1	in AID(BlastR) M33R (Fr=0.14)
Para6	2	15	0	
Para7	2	3	0	

Table 2.2: Details of the mutations detected in DNA sequencing. (continue on next page)

cuit, and compared its operation in three distinct regimes—uncontrolled growth, negative feedback, and paradoxical—by using media with different combinations of blasticidin and AP1903. Long-term culturing for up to 43 days revealed that while both negative and paradoxical feedback architectures can limit cell population size

Table 2.2: Details of the mutations detected in DNA sequencing. (continued from last page)
Circuit components of isolates from movies set 1 and 2 (Figure 2.11) were amplified and sequenced. For simplicity, uncontrolled growth, negative feedback and paradoxical feedback conditions are noted as unCt, Neg and Para, respectively, in the table. Four out of 18 sequenced isolates exhibited non-synonymous mutations at frequencies that were significantly elevated compared to those of synonymous mutations, consistent with positive selection (Bonferroni corrected, $p<0.05$, Figure 2.13A). Three of these (one in a negative feedback cheater isolate, the other two from non-cheating paradoxical isolates) squired mutations in the AID domain of the AID-BlastR fusion protein. Reduced activity of this AID domain would be expected to increase fitness under both paradoxical and negative feedback conditions. The fourth mutation, in a non-cheating paradoxical isolate, was a premature stop codon in osTIR1, near the 5' end of the circuit transcript, which could potentially eliminate expression of all circuit components from the integrated copy in which it appears. (Because there are multiple genomic integrations of the circuit, any individual mutation could have a limited beneficial effect.)

for weeks, the former is more susceptible to escape by sensing-deficient cheaters. By contrast, the paradoxical design suppressed these cheaters, as predicted theoretically, and provided more robust population control for the conditions and timescales explored here.

The auxin system can be applied cell therapies

With these developments, private communication channels, population sensing, and population control could improve engineered cell therapies by allowing cells to coordinate their responses and localize activities at target sites. Previously, several studies attempted to activate latent natural signaling abilities to trigger coordinated multicellular responses. For example, to locally stimulate immune function, T cells were engineered to secrete IL-12 and IL-18 upon tumor infiltration (Chmielewski et al., 2011; Hu et al., 2017). Similarly, secretion of bispecific T-cell engagers (BiTEs) by engineered T cells was shown to guide bystander T-cells to attack at the cancer site, successfully improving infiltration and reducing toxicity caused by normal tissue expressing the target (Choi et al., 2019). Private auxin-based communication channels would complement these approaches by allowing engineered cells to not only specifically sense and limit their own local population size but also to enable conditional activation only beyond a minimum density.

We therefore anticipate the incorporation of synthetic communication, or population control systems in future generations of engineered cell therapies. In the later

part of the thesis (Chapter 4), I will present our ongoing effort in this direction, demonstrating the potential of one possible multicellular circuit in engineered cell therapy.

Limitation of this study

The Paradaux cell line requires addition of the auxin precursors IAM or NAM for auxin synthesis. Though this provides a good external control method for our experimental setup, it might be suboptimal in therapeutic scenarios. For these applications, complete biosynthesis of IAA can be achieved by expressing tryptophan 2-monooxygenase (iaaM), in conjunction with iaaH (Figure 2.3).

The use of blasticidin and its auxin-regulated resistance gene for growth control is not ideal, since antibiotics have complex effects on the cell, must be added to the media, and broadly affect all cells exposed to the drug, making them inappropriate for cell therapy applications. Enzyme-prodrug systems could provide an alternative approach with fewer non-specific effects on host cells (Trask et al., 2000; Sharrock et al., n.d.). Alternatively, a more generalizable, cell-autonomous system could be achieved by coupling auxin to a cell-cycle regulator, such as Cdk1, 2 or 3 (Satyanarayana and Kaldis, 2009), or other genes (Harborth et al., 2001) essential for survival and/or cell cycle progression.

Due to the inherent time delays within the feedback loop, the Paradaux circuit exhibits oscillatory behaviors (Figure 2.10F), similar to earlier synthetic population control circuits in bacteria Balagadde, 2005. Reducing feedback delays or implementing more sophisticated control systems (Chevalier et al., 2019) should facilitate non-oscillatory homeostatic dynamics.

Finally, although the paradoxical control system successfully extended the duration of population control, cells nevertheless accumulated adaptive changes (Figure 2.11, 2.12 and 2.13). Future work exploring even longer timescales should reveal how long the paradoxical design can extend the duration of control in the presence of strong selection pressure to subvert it.

2.5 Supplementary text

In this section, I will provide the detailed mathematical deductions that are omitted in sections 2.2 and 2.3, as well as the captions for the related movies.

Estimation of auxin diffusion coefficient

To estimate the auxin diffusion coefficient, we modeled the system as a fixed concentration source, with the standard result for one-dimensional diffusion:

$$n(x,t) = n_0 \cdot \text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{2.3}$$

in which, n is the concentration as a function of time (t) and space (x), n_0 is the fixed concentration source, erfc the complementary error function and D is the diffusion coefficient. The diffusion coefficient of auxin has been previously measured to be $5.58\times10^{-6}cm^2/s$ (Robinson, J. E. Anderson, and J. L. Lin, 1990). Therefore after 48 hours, the diffusion will expand to approximately 3 to 4 centimeters (Figure 2.3D). Compared to this ideal fixed concentration source model, the in vitro experimental gradient (Figure 2.4E) is expected to exhibit delays in both auxin production and auxin sensing. Nevertheless, the experimentally observed gradient length scale was similar to the theoretical model prediction (tens of millimeters).

Fitting for the auxin and population responsive curves

To fit the mCherry's response curve to auxin and population (Figure 2.2B, 2.5A-B and 2.6), we assumed the log of fluorescence (F) follows a inverted Michaelis-Menten's equation:

$$log(F) = log(F_{max}) - Amp \cdot \frac{x}{EC_{50} + x} \tag{2.4}$$

Here, F_{max} represents the basal level of fluorescence; Amp represents the max amplitude the population or auxin (represented by x) could reduce the fluorescent; and EC_{50} represents the concentration, or population number when the reduction is half of the Amp.

In Figure 2.6B, to avoid the problem that the above function (Equation 2.4) does not extrapolate well when fitted with flat lines (controls and samples treated with auxin), we used $x = 10^3$ and $x = 3 \cdot 10^5$ for the extreme values. More specifically, the "ideal dynamic range" was defined as $\frac{F_{control}(x=10^3)}{F_{auxin}(x=10^3)}$, the "loss to self-sensing" was defined as $\frac{F_{control}(x=10^3)}{F_{precursor}(x=10^3)}$, the "loss to saturation" was defined as $\frac{F_{precursor}(x=3\cdot10^5)}{F_{auxin}(x=3\cdot10^5)}$, and the "actual dynamic range" was defined as $\frac{F_{precursor}(x=10^3)}{F_{precursor}(x=3\cdot10^5)}$.

Model of the paradoxical population control circuit

Here we describe a dynamical model of the paradoxical population control circuit. The model is based on the following biochemical reactions, interactions, and assumptions:

- Auxin, denoted A, is synthesized from its precursor through an iaaH-catalyzed hydrolysis reaction at a constant synthesis rate per cell λ_A, eliminated at a rate δ_A, which is dominated by dilution due to media changes. Auxin diffuses rapidly in and out of the cell compared to the timescales of cell population dynamics, and its concentration is therefore assumed to be at quasi-steady state (Equation 2.1).
- iCasp9, denoted C, and BlastR, denoted R, are produced at rates λ_C and λ_R, respectively.
- Auxin binds reversibly to osTIR1 to form an auxin-osTIR1 complex, which ubiquitylates and degrades iCasp9 and BlastR via their attached AID domains. These reactions are described using classical enzyme kinetics with the auxin-osTIR1 complex as the activating enzyme, described as a constant rate ν_{ub}. In addition to auxin-induced active degradation, iCasp9 and BlastR are also eliminated at rate δ due to dilution (Equations S3 and S4).
- The concentrations of iaaH and osTIR1 are assumed to be constant. Auxin precursor is assumed to remain at excess, saturating concentration.
- osTIR1 is assumed to be present at excess concentration compared to the iCasp9-auxin-osTIR1 and BlastR-auxin-osTIR1 complexes, and therefore potential competition between iCasp9 and BlastR for osTIR1 can be neglected.

With these assumptions, we can describe the dynamics of iCasp9 and BlastR with the following differential equations:

$$\frac{dR}{dt} = \lambda_R - \nu_{ub} \cdot R \cdot A - \delta \cdot R \tag{2.5}$$

$$\frac{dC}{dt} = \lambda_C - \nu_{ub} \cdot C \cdot A - \delta \cdot C \tag{2.6}$$

The model represents blasticidin and AP1903 interactions as follows:

- Extracellular Blasticidin, denoted B, diffuses into the cell, where it is denoted B_{int}, and undergoes subsequent enzymatic inactivation by BlastR, with a threshold concentration of K_B, and a Hill coefficient n_1:

$$B_{int} = B \cdot \frac{(K_B)^{n_1}}{(K_B)^{n_1} + R^{n_1}} \tag{2.7}$$

- AP1903 forms an active caspase complex, $[I : C_{n_2}]$ with iCasp9, with a threshold concentration of K_d^C and a Hill coefficient n_2:

$$[I : C_{n_2}] = \frac{1}{(K_d^C)^{n_2}} \cdot I \cdot C^{n_2} \tag{2.8}$$

In these two cases, we allow the more general Hill kinetics to account for potential intermediate reaction mechanisms that could influence the effective cooperativity in the final expressions. Additionally, we assumed that both inactivation of Blasticidin and iCasp9 binding to AP1903 are rapid and have reached steady state.

As described in the main text the overall population dynamics can be described using a generalized logistic function, with the growth rate represented as a linear combination of blasticidin-dependent and iCasp9-dependent terms (Equation 2.2). The Blasticidin-dependent growth rate, F_G, is a sum of two terms. The first describes attenuation of the maximum natural cell proliferation rate, α, with increasing blasticidin, while the second represents an increase in the cell death rate, β, with increased blasticidin. These terms are associated with half-maximal blasticidin concentrations of K_g and K_d, respectively:

$$F_G = \alpha \frac{\kappa_g}{\kappa_g + B_{int}} - \beta \frac{B_{int}}{\kappa_d + B_{int}} \tag{2.9}$$

For simplicity, we assume $\kappa_g = \kappa_d = \kappa$. Thus, Equation 2.9 can be reduced to the following form:

$$F_G = (\alpha + \beta) \frac{\kappa}{\kappa + B_{int}} - \beta \tag{2.10}$$

We similarly describe the iCasp9-dependent cell death rate, F_D, with a Hill function dependence on the concentration of the AP1903-iCasp9 ($I : C_{n_2}$) complex:

$$F_D = \beta_C \frac{[I : C_{n_2}]}{[I : C_{n_2}] + \kappa_{[I:C]}} \tag{2.11}$$

Adding Equations 2.10 and 2.11 together and substituting the corresponding terms from Equations 2.7 and 2.8, generate the complete form of the growth rate function $R(B, I, A)$ in Equation 2.2:

$$R_g(B, I, A) = (\alpha + \beta)\frac{\kappa\left[(K_B)^{n_1} + R^{n_1}\right]}{\kappa\left[(K_B)^{n_1} + R^{n_1}\right] + B(K_B)^{n_1}} - \beta - \beta_C \frac{\frac{1}{(K_d^C)^{n_2}} \cdot I \cdot C^{n_2}}{\frac{1}{(K_d^C)^{n_2}} \cdot I \cdot C^{n_2} + \kappa_{[I:C]}} \tag{2.12}$$

To simplify this description, we assumed a time-scale separation between the faster auxin-population dynamics and the slower intracellular reactions involving R and C. Using singular perturbation theory (Del Vecchio and Murray, 2014), the system can then be approximated by a simpler system that retains only the slower dynamics (Equation 2.1 and 2.2), while the faster dynamics (Equation 2.5 and 2.6) are considered to be at equilibrium: $\frac{dR}{dt} = \frac{dC}{dt} \approx 0$. With this approximation, we can write R and C in terms of the auxin concentration, A.

$$R = \frac{\lambda_R}{\delta + \nu_{ub} A} \tag{2.13}$$

$$C = \frac{\lambda_C}{\delta + \nu_{ub} A} \tag{2.14}$$

We also defined the following additional parameter combinations for simplicity (see Table 2.1):

$$\kappa_A = \frac{\nu_{ub}}{\delta};\ \kappa_B = \frac{1}{\kappa};\ P_R = \frac{\lambda_R}{K_B};\ P_C = \frac{\lambda_C}{K_d^C};\ \kappa_I = \frac{1}{\kappa_{[I:C]}} \tag{2.15}$$

We can then substitute R and C in Equation 2.12 to obtain the following:

$$R_g(B, I, A) = (\alpha + \beta) \cdot G(B, A) - \beta - \beta_C \cdot D(I, A) \tag{2.16}$$

in which:

$$G(B, A) = \frac{(\kappa_A A + 1)^{n_1} + \left(\frac{P_R}{\delta}\right)^{n_1}}{(\kappa_A A + 1)^{n_1}(\kappa_B B + 1) + \left(\frac{P_R}{\delta}\right)^{n_1}} \tag{2.17}$$

$$D(I,A) = \frac{\kappa_I I \cdot \left(\frac{P_C}{\delta}\right)^{n_2}}{\kappa_I I \cdot \left(\frac{P_C}{\delta}\right)^{n_2} + (\kappa_A A + 1)^{n_2}} \tag{2.18}$$

To verify the validity of the approximation, we compared the full model (Equation 2.1, 2.2, 2.5 and 2.6) to the approximate model, where Equation 2.5 and 2.6 are set to 0. We simulated both models using the parameter values in Table 2.1 (the Paradaux set). The two sets of traces closely followed each other (Figure 2.10B), indicating that the approximate system accurately reproduced the dynamics of the full model.

Representing synergy between iCasp9 and blasticidin control of cell survival

As discussed in the main text, the model above overestimated actual survival rates when both arms of the circuit were simultaneously active (Figures 2.9A and 2.10A, dashed purple lines), likely due to synergy between apoptosis and blasticidin-dependent translational inhibition. We therefore added a phenomenological synergistic interaction term: $-\beta_{syn} \cdot [1 - G(B,A)] \cdot D(I,A)$ to the growth rate expression (R_g in Equation 2.12):

$$R_g(B,I,A) = (\alpha + \beta_B) \cdot G(B,A) - \beta - \beta_C \cdot D(I,A) - \beta_{syn} \cdot [1 - G(B,A)] \cdot D(I,A) \tag{2.19}$$

Equation 2.19 gives the final form of the paradoxical growth curve with the synergistic correction, and is used to improve data fitting (Figures 2.9A and 2.10A, dotted purple lines). Together with the fitted parameters, the reduced system (Equations 2.1 and 2.2, with R_g defined as Equation 2.19) was used to run parameter screens and dynamic simulations (Figures 2.9B-D and 2.10E-F).

Parameter screening and stability analysis

For numerical parameter screening and stability analysis, we computed some terms analytically to make the process faster and more efficient. The stability at equilibrium points of the reduced dynamical system (Equations 2.1 and 2.2), is determined by its Jacobian matrix (J_{eq}, see below) and its eigenvalues.

$$\boldsymbol{J}_{eq} = \begin{pmatrix} -\delta_A & \lambda_A \\ R'(A) \cdot L(N) & R(A) \cdot L'(N) \end{pmatrix} \tag{2.20}$$

Here, $L(N) = N(1 - N^{\nu})$.

Note that in the operating range ($0 < N < 1$), $L(N) > 0$ and $L'(N) > 0$; and at any equilibrium points, $dN/dt = 0$. Together with the explicit form of Equation 2.1, the points above leads to $R(A) = 0$ at equilibrium. Therefore, the eigenvalues ($\lambda_1 \, and \, \lambda_2$) of J_{eq} at equilibrium are the roots of the quadratic characteristic equation:

$$\lambda \cdot (\delta_A + \lambda) - \lambda_A \cdot R' \cdot L' = 0 \tag{2.21}$$

As noted above that $L' > 0$, the sign of roots (eigenvalues $\lambda_1 \, and \, \lambda_2$) can be solely determined by the sign of R'. For $R' > 0$, at least one of the eigenvalues has a real part greater than zero, making the associated equilibrium point unstable. For If $R' > 0$, we can deduce that the real parts of both eigenvalues are less than zero, making the equilibrium point stable. Based on this analysis, we screened the equilibrium points of the system and the corresponding sign of R' to determine the stability and type of each parameter set (Figure 2.9B-C).

Besides the five major behavior categories described above (Figure 2.9B, surrounding plots), a small but significant portion (0.68% of total) of conditions appearing at the border between "permissive" and "paradoxical" types, could not be classified into any of the five types. To further investigate these cases we down-sampled the space from 201 x 201 to 21 x 21 conditions and plotted all the six (0.62% or total) unclassified curves, as well the permissive and paradoxical types next to this region (Figure 2.10G; grey, pink, and light blue, respectively). Inspection of these curves revealed a transitional behavior between permissive and paradoxical, with the unusual equilibrium points caused by the nonlinearity of the introduced synergetic term. Note that the range of net growth rate of the system is around -0.09 0.04 defined by α, $\alpha - \beta_B$ and $\alpha - \beta_C$. The net growth rates of these curves around the unusual equilibrium points is significantly lower (-0.01 to 0.02), indicating those points are unstable. Therefore, those curves' dynamics, although not mathematically classified, will behave similarly to either permissive or paradoxical types.

Delayed blasticidin killing affects oscillations

To simulate the delay effect of blasticidin killing cells (Sato et al., 2012), we added a time delay τ to the blasticidin related growth function (Equation 2.17) at time t, resulting in a delayed blasticidin growth term, $G(B, A)_{t-\tau}$. We then replaced blasticidin related growth term $G(B, A)$ in Equation 2.19 with this new term with delay:

$$R_g(B, I, A, \tau) = (\alpha + \beta_B) \cdot G_{t-\tau} - \beta - \beta_C \cdot D_t - \beta_{syn} \cdot [1 - G_{t-\tau}] \cdot D_t \quad (2.22)$$

In the above equation, we omitted the arguments of the growth functions for simplicity (e.g. $G(B, A)_t$ is simplified to G_t). Additionally, for simulations in Figure 2.10F and Movie S2 we assumed $G_{t-\tau} = G_0; \forall t < \tau$. The results of this delay are the oscillations shown in Figure 2.10F and Movie S2.

Captions for the movies

- Movie S1. Sender-Receiver-PIN2 cells escape after 15 days of continuous culture. Available at: `cell.com/cms/10.1016/j.cell.2022.01.026/attachment/4d93513d-0e2b-4c05-915d-d9758e6a59fc/mmc2.mp4`

  Cells were seeded into a 24-well imaging plate with 50 μg/ml of blasticidin solely ("Uncontrolled") or with 100 μM IAM ("Population circuit on") or IAA ("Killing constitutive on"). Images of constitutively expressed NLS-citrine were taken with a 20x inverted microscope once per hour.

- Movie S2. Numerical simulation reveals delay bifurcation between damped and limit cycle oscillations. Available at: `cell.com/cms/10.1016/j.cell.2022.01.026/attachment/d5136ed3-b14b-4d23-959d-b05fd2ddcc98/mmc3.mp4`

  This movie shows simulated dynamics of the paradoxical feedback model for different values of the delay parameter, τ. For each value of τ, the simulation shows two trajectories starting from different initial conditions. Note the transition from damped to limit cycle oscillations between 42 and 54 hours. Initial conditions were held fixed at a cell seeding density of 0.1 and auxin concentration of 40 μM and 16 μM for trajectory 1 and trajectory 2, respectively.

- Movie S3. Time-lapse imaging from Figure 5A reveals robustness of the paradoxical population control architecture. Available at: `cell.com/cms/10.1016/j.cell.2022.01.026/attachment/c8b7534d-3f50-47f9-b41c-7e952a8f3b7f/mmc4.mp4`

  10,000 cells were seeded per well into a 24-well imaging plate (Figure 2.11A), with no control (100μM NAM; left), negative feedback (100 μM NAM and 50 μg/ml blasticidin; middle), or paradoxical feedback (100 μM NAM, 50 μg/ml

blasticidin, and 50 nM AP1903; right). Images of constitutively expressed mTagBFP2 were taken with a 20x inverted microscope every 4 hours, in a 6x6 grid, and stitched together. Bar=500micron. This movie is from movie set 3.

Chapter 3

EPIGENETIC-BASED SYNTHETIC GENE REGULATION

3.1 Introduction

The field of synthetic epigenetics seeks to harness epigenetic regulatory mechanisms to control gene expression on different timescales (Kungulovski and Jeltsch, 2016; Nakamura et al., 2021). The inherent memory-like property of epigenetical expression regulation make it idea to create, or at least simulate, multiple cell types without the required additional circuit designs to ”lock” the cells in a certain state (Gardner, Cantor, and Collins, 2000; R. Zhu et al., 2022).

Regulation of gene silencing by DNA methylation in natural context

One particular type of synthetic regulation is utilizing DNA methylation to regulate desired gene target. In natural mammalian context, DNA methylation occurs exclusively at CG dinucleotides (CpGs) site, and plays critical roles in development, tumor progression, and aging (Ehrlich, 2002; Greenberg and Bourc'his, 2019; McCabe, Brandes, and Vertino, 2009; Smith and Meissner, 2013). These functions result mainly from the ability of CpG methylation to induce and stabilize gene silencing in mammals, through multiple mechanisms (Attwood, Yung, and Richardson, 2002; H. Zhu, G. Wang, and Qian, 2016). Control of DNA methylation and further gene silencing depends on both trans-acting factors and the DNA sequence itself. Trans-factors include methylation “writers,” such as DNA methyl-transferases (DNMTs) (Lyko, 2018) and “erasers” such as TET1 (Morita et al., 2016) that establish and alter methylation marks, as well as “readers,” such as MeCP2 and histone deacetylases (Choudhury et al., 2016) that link methylation to regulation of gene transcription (Moore, T. Le, and Fan, 2013; H. Zhu, G. Wang, and Qian, 2016).

In mammals, DNA methylation occurs almost exclusively at CpGs, it has been unclear how the dynamics of gene silencing of these system depends on the DNA sequence of the regulated target gene. Previously, various evidence that suggest CpG content could play pivot role in gene regulated has been discovered in natural context, at genome level. For example, regions with different CpG content exhibit distinct methylation patterns (Lövkvist et al., 2016; Weber et al., 2007), potentially due to cooperativity between nearby CpGs (Haerter et al., 2014; Bruno, Williams, and Del Vecchio, 2022). High CpG-density regions (CpG islands) from a human

chromosome largely maintain their methylation state when hosted in a transchromosomic mouse model (Long et al., 2016), suggesting that DNA sequence composition plays a strong role in establishing stable methylation states. Conversely, insertion of several hundred base pairs of CpG-free DNA can disrupt these patterns, permitting de novo methylation of the surrounding CpG island (Takahashi et al., 2017). These studies suggest that DNA sequence plays an important role in methylation.

However, the precise role of CpG sequence context can be difficult to discern at natural loci, where regulation is also affected by many other cis- and trans-acting factors, including cell-type specific methylation writer and reader profiles, neighboring (non-CpG) motifs that recruit epigenetic modifiers, pre-existing chromatin states, etc.

Synthetic gene silencing by DNMT recruitment

Recent work demonstrated the ability to regulate synthetic or endogenous gene expression by recruiting DNMTs to specific target genes (Bintu et al., 2016; X. S. Liu et al., 2016; Nuñez et al., 2021; Van, Fujimori, and Bintu, 2021), and even create fully synthetic DNA methylation based systems for synthetic epigenetic memory (Park et al., 2019).

These synthetic systems grant us full control over the sequence of the regulated target, and timing of the DNMT recruitment, thus provide us a unique opportunity to gain insight into how CpG content can affect the dynamics of the silencing. Additionally, understanding the effects of sequence composition on silencing rates could provide insight into gene regulation by DNA methylation and also expand the synthetic epigenetic toolbox, for fine tuning circuits.

In this chapter, we adapted a previously established DNMT recruitment system to analyze the effects of DNA sequence on methylation-dependent silencing (Bintu et al., 2016). We derived a library of promoters with different CpG densities from a synthetic promoter and observed the relationship between CpG density and the silencing dynamics occurring after DNMT recruitment (Figure 3.1). We also identified several specific CpG elements that appear to play disproportionate roles in silencing dynamics and confirmed that one of them (near the TATA box and transcription start site), causes significant changes in methylation dynamics. Our results reveal how CpG density influences silencing dynamics, and provide a library of promoters with different silencing rates for synthetic applications.

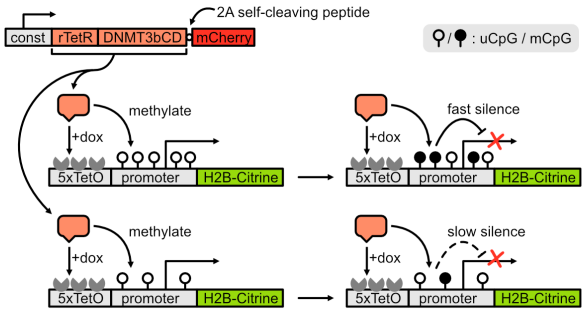


Figure 3.1: rTetR fused DNMT3b catalytic domain methylate and silence a reporter with of different CpG content upon recruitment.
Schematic of the synthetic methylation-silencing system: rTetR fused DNMT3bCD is expressed constitutively, and upon induction of dox, recruited to the promoter region of a site-specifically integrated Citrine reporter. The recruitment methylates the promoter and further silences the gene expression, with dynamics depending on the promoter's CpG content.

3.2 Promoter CpG content regulates DNMT-dependent silencing dynamics

In this section, we will investigate how CpG content, or more specifically density, on promoters can influence the rate of silencing when targeted by DNMT. We will construct a library of promoter with different CpG density follow their silencing dynamics inside a well controlled, fully synthetic, DNMT recruitment system.

Construction of a promoter library with varying CpG content

To investigate the relation between promoter CpG content, and its DNMT-dependent silencing rate, we adopted a previously described synthetic methylation-silencing system (Bintu et al., 2016). In this system, the catalytic domain of DNMT3b (named DNMT3bCD) is fused with a reversed tetracycline repressor (rTetR), allowing precise temporal control of the recruitment to a target gene by adding doxycycline (dox) to the culture media (Urlinger et al., 2000). The construct also incorporated a co-expressed H2B-mCherry fluorescent protein fusion. We stably integrated this construct using the piggyBac transposon system, and selected a single stable clone. This enabled direct comparison of different promoters (see below) with the same DNMT3bCD expression context.

As the target gene, we used an H2B-mCitrine (Zacharias et al., 2002) fluorescent fusion protein. This target was driven by one of a set of promoters containing varying densities of CpG (see below). In each promoter, an array of 5 rTetR binding

sites (TetO) was fused upstream of the promoter, allowing recruitment of rTetR-DNMT3bCD (Figure 3.1). To enable direct comparison between target promoters at the same genomic context, all reporter cassettes were site-specifically integrated at single copy into a landing pad previously engineered within an artificial chromosome (S. Yamaguchi et al., 2011).

We constructed a library of synthetic promoters that differed in their CpG densities. We started with a synthetic version of the human elongation factor 1α promoter (pEF1s(orig), with 18% CpG density), a 544bp fusion of promoter fragments from the human EF1α promoter and human T-cell virus (HTLV) (Argentova et al., 2017) that has been commercially available (InvivoGen inc.) as well as used for antibody expression and gene therapies (Fu et al., 2021; Oliveira et al., 2022). To identify conserved CpG elements, we compared both the EF1α fragment and HTLV fragments of this promoter to there natural orthologs, respectively (Sayers et al., 2022). We then removed or added CG pairs into the promoter at non-conserved sites. With this procedure, we generated promoters with varying CpG densities at 9.6% and 24% (pEF1s(low) and pEF1s(high), respectively, Figure 3.2, middle).

Next, to broaden the range of CpG densities, we designed an additional DNA segment, inserted upstream of the promoter, containing high (60%) CpG density (Figure 3.2, high CpG insert). We altered this CpG insert by swapping out CG with GC dinucleotides, or by replacing C with T, to create a lower CpG density (5.4%) insert, while otherwise preserving its sequence similarity with the high CpG insert (Figure 3.2, low CpG insert). Altogether, we combined the three pEF1s promoters with the two inserts, or with no insert, to produce a library of 7 sequences whose overall CpG density ranged from 8.0% to 36% (Figure 3.2 right). Despite their variation in CpG density, all 7 promoters drove strong expression of the fluorescent protein reporter (Figure 3.3A and B).

DNMT-dependent silencing rate correlates with promoter CpG density

Previous analysis of silencing dynamics by DNA methylation in a similar system revealed that silencing occurs through stochastic, all-or-none, irreversible events in individual cells (Bintu et al., 2016). In order to assess the fraction of cells silenced under different conditions, we used the unsilenced (actively expressing) expression distribution to define a silencing threshold, set at 2 standard deviations below the mean fluorescence levels of actively expressing cells (Figure 3.4A). Once silenced by recruitment of DNMT3b, cells remained off for at least ten cell generations, as

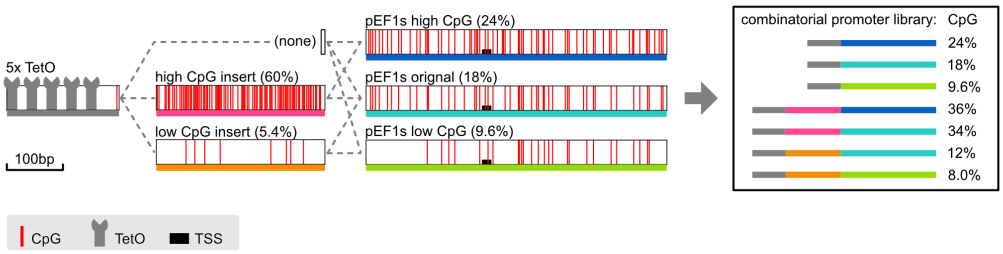


Figure 3.2: Design of the library of promoters with varying CpG content.
5x tandem TetO binding sites were fused with a insert (or no insert) and a pEF1s synthetic promoter to make the library of the promoter. Red lines represent CpG dinucleotides, and the pEF1s promoters are vertically aligned to show sequence homology.

previously described (Bintu et al., 2016) (Figure 3.3C).

To quantify silencing dynamics across the library, we analyzed the dynamics of silent cell accumulation over a 25 day time course after addition of dox (Figure 3.4B, right). The stability of the H2B-Citrine protein makes it challenging to capture the dynamics occurring at a similar or faster scale compared to cell division (approx. 22 hours). Therefore, we repeated a similar time course for the four fastest promoters, except that before analyzing cells, we cultured them for 2 additional days without dox, allowing time for fluorescent protein levels to dilute out in transcriptionally silent cells, and therefore better reflect the transcription state of the gene (Figure 3.4B, left). This procedure allowed us to quantify the fraction of silent cells for up to 25d after dox addition.

The time-course dynamics for each promoter could be summarized by a stochastic silencing rate. We defined this empirical silencing rate as the inverse of the time required to silence 50% of the population ($t_{1/2}^{-1}$). Silencing rates varied over nearly an order of magnitude across promoters with various CpG densities (Figure 3.4C). Further, the logarithm of the silencing rate correlated linearly with CpG density over this range (Figure 3.4C). These results show that across varying CpG densities, DNMT-dependent silencing dynamics are broadly consistent with a stochastic, all-or-none silencing process, occurring at a rate that depends on CpG density.

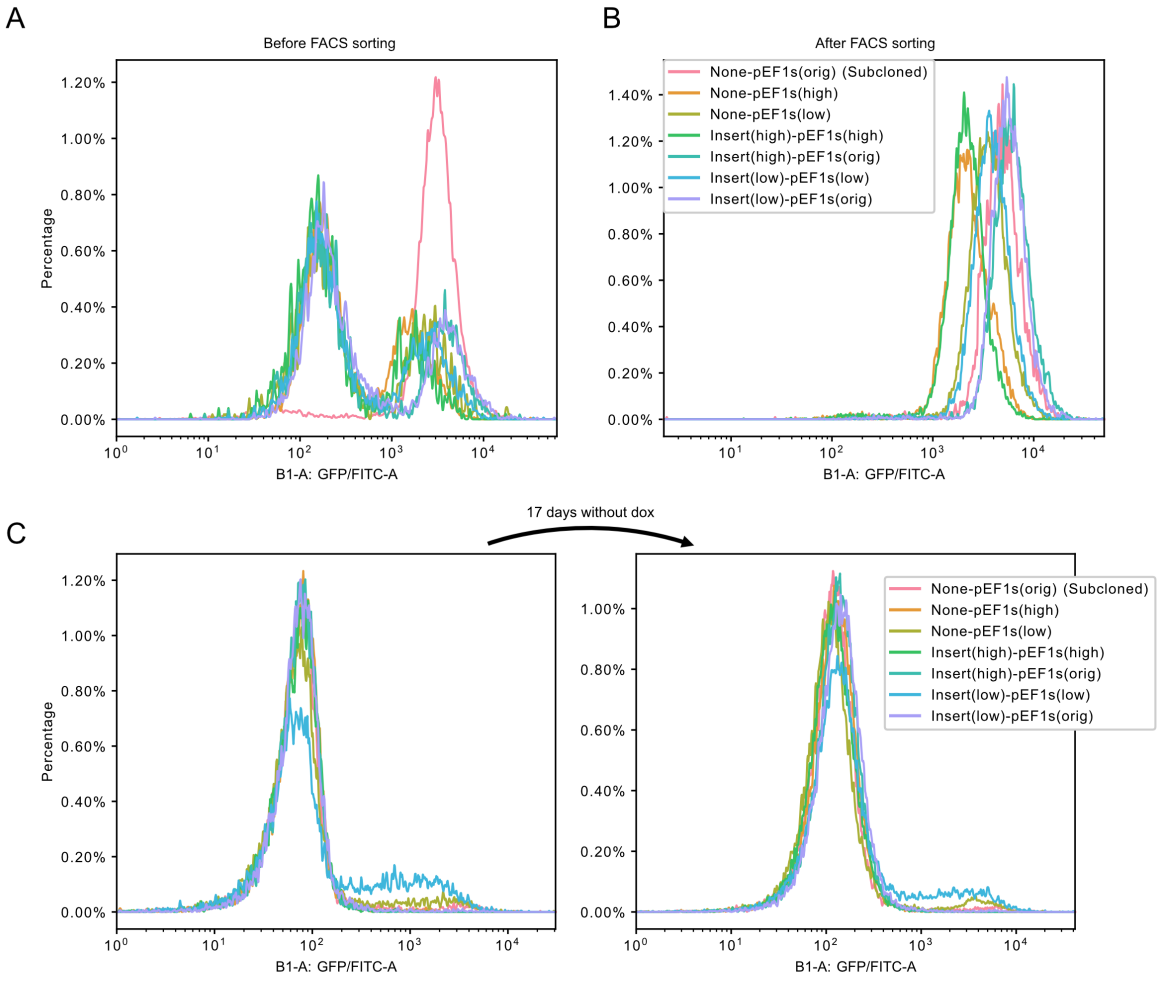


Figure 3.3: Generation of the cell line and their stability after silencing.
(A-B) Site-specific integration of the reporters generates cell populations with uniform expression after FACS sorting: Cells are transfected and selected with geneticin for 14 days (A). And then subcloned (None-pEF1s(orig)) or FACS sorted (all others) for further analysis (B).
(C) Cells that are silenced remain stably silenced: The silenced population were taken from the end of time trace of Figure 3.4B, no release group (left). These cells were further cultured for 17 days without recruitment (no dox) (right).

A phenomenological model explains the dependence of silencing rate on CpG content

Next, to explore potential mechanisms for the control of silencing rate by CpG density, we constructed a phenomenological mathematical model. We assume that each cell can be in either an active or a silent state. Under DNMT recruitment condi-

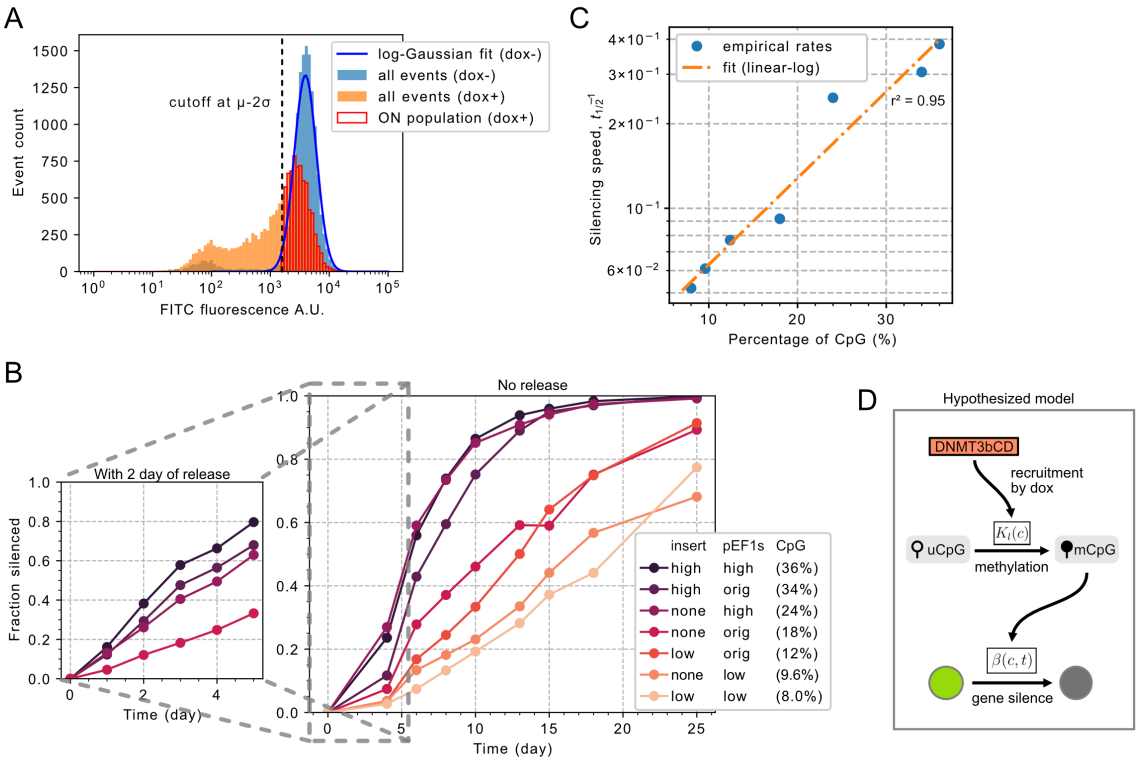


Figure 3.4: Promoters' silencing rate correlates tightly with their CpG content.
(A) Example of one time points' flow data: Cells with reporters driven by the "None-pEF1s(orig)" promoter were induced with dox (or no dox for control) for 10 days, and analyzed by flow cytometry. The "log-Gaussian fit" is estimated by fitting data higher than 400 FITC fluorescence A.U. with a Gaussian distribution, after logarithmic transformation.
(B) Time course of the silenced fraction of different promoters: For the long time scale experiment (right), cells were treated with or without dox, and the fraction of silencing is determined as described in A: cells with lower fluorescence than $\mu - 2\sigma$ of the no dox control group were determined as silenced. The silencing rate is further normalized to the no dox group (Method, see section A.6). For the shorter time scale (left), the same method is used except the flow cytometry analysis is performed two days after the withdrawal of the dox.
(C) Summary of the silencing rates in B: The silencing rate ${t_{1/2}}^{-1}$ is the inverse of time at which point 50% of the population got silenced. More specifically, the time is derived from linear interpolation of the neighboring data points before and after the 50% mark. If silencing rates of the cell lines is available in the faster scale plot (B, left), it is used instead.
(D) Hypothesis of the two-step phenotypical model: In this model, DNA are methylated first by the recruited DNMT3bCD at rate K_I. Subsequently the methylated CpGs leads to gene silencing at rate β.

tions, cells in the active state transition to the silent state at a rate β , whose value can depend on the CpG density, c, and the time since the start of DNMT recruitment, t. With these assumptions the size of the active population fraction, A, can be described by a simple differential equation:

$$\frac{dA}{dt} = -\beta(c,t) \cdot A \tag{3.1}$$

Note that cell proliferation does not need to be explicitly incorporated due to the heritability of the expression state.

To identify a minimal model capable of explaining the data, we first assumed that the silencing rate β is time independent, and determined only by c:

$$\beta(c,t) = K_1(c) \tag{3.2}$$

Empirically, this means the promoters' silencing is not affected by the dox induction history, and is a constant only determined by its own CpG content, c. This model fits (see supporting information) the behavior of the faster silencing (higher CpG density) promoters. However, it failed to capture an initial delay in silencing of the slower (lower CpG density) promoters (Figure 3.5A).

To account for the delay in silencing of lower CpG promoters, we next tried the simplest time-dependent model, in which the silencing rate β grows linearly with time:

$$\beta(c,t) = K_2(c) \cdot t \tag{3.3}$$

In contrast to the time-independent model (Equation 3.2), this model fit slower promoters, but overestimated silencing in the faster promoters (Figure 3.5B, Supplementary text).

These results suggested a third model, in which time-dependent effects saturate at higher silencing rates:

$$\beta(c,t) = \frac{K_{max} \cdot K_l(c) \cdot t}{K_{max} + K_l(c) \cdot t} \tag{3.4}$$

Here, K_{max} represents the maximum silencing rate, which is approached when the product $K_l(c) \cdot t$ is high. Critically, we assume K_{max} is a constant independent of

the CpG density, and therefore shared between different promoters, increasing the total number of parameters across all promoters by one, from 7 (one for each of the $K(c)$ values) to 8. This model generally captured the dynamics of both the faster and slower promoters (Figure 3.5C). However, some early dynamics of one intermediate density promoter deviated significantly. The log of the parameters $K_I(c)$ generally correlated with c, similar to the empirically determined ${t_{1/2}}^{-1}$ (Figures 2D, S2D). Taken together, these results provide a phenomenological parameterization of the dependence of silencing dynamics on CpG density. To explain how this model (described by Equation 3.4) could arise from underlying biochemical interactions, we considered an underlying model in which the process of DNMT-silencing occurs through two distinct steps (Figure 2D). First, recruitment of DNMT3b leads to slow and independent methylation of individual CpG dinucleotides ($K_I(c) \cdot t$), such that the total rate of methylation is proportional to the number of unmethylated CpG in the promoter proximal region. Second, transcriptional silencing occurs at a rate that depends in a saturating manner on the total number of methylated CpGs in the promoter. These assumptions provide a hypothesis for the empirical model.

3.3 Methylation-specific sequencing revealed accumulation of methylation and a potentially significant CpG

In this section, we performed methylation specific sequencing (EM-Seq, NEB) to test the model we proposed above. Additionally, the sequencing results also revealed a specific CpG locating between TATA box and TSS, that can explain a significant portion of the difference in silencing rate between promoters.

Methylation accumulates after DNMT recruitment

The model predicts a linear accumulation of methylation over time after dox addition at a rate dependent on the CpG density of the promoter. To test this prediction, we used FACS to isolate the transcriptionally active cell fraction (A in the model), at different times after dox addition (Figure 3.6A). We then measured promoter CpG methylation profiles using methylation specific sequencing (EM-seq, NEB. Vaisvila et al., 2021) (Figure 3.6A).

As expected, we observed constant rates of methylation accumulation in the transcriptionally active populations (Figure 3.6B). Contrary to the model, however, the rate of methylation accumulation was independent of the CpG density of the promoter. In fact, the rate of methylation per CpG was greater at promoters with lower CpG densities (Figure 3.6A), but the total number of methylated CpG in the pro-

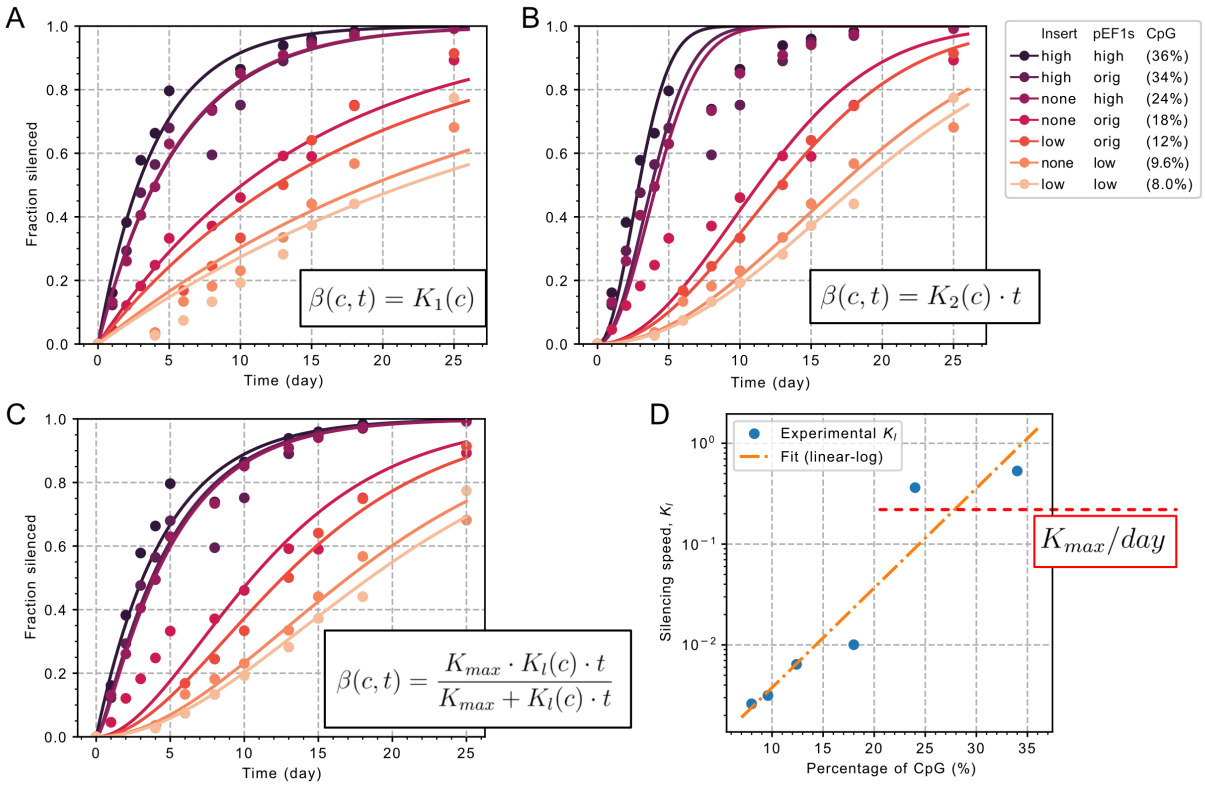


Figure 3.5: Comparison of the different models.
(A-C) Fitting of the data using different models (described by Equation 3.2-3.4): Data points are from the same set from Figure 3.4B, combined. All fitting is done with the least square method (see SUPPORTING INFORMATION).
(D) Silencing rates from the fitting of the model in C are mostly consistent with our experimental finding (Figure 3.4C). Fitting of K_l from C are plotted against the promoters' CpG content. Red line implies where K_{max} is when divided by 1 day. Note the K_ls of the promoter with highest CpG content is omitted here as it is over 105. This is likely due to the artifact of the fitting algorithm: As the initial empirical rate ($\beta(c,t)_{t=0}$) get closer to the maximum possible rate K_{max}, the fitting static (sum of the variations) became insensitive to the time dependent term ($K_l(c) \cdot t$), as long as it is big enough. So that the algorithm will try to increase its value for trivial gain in the static for "infinitely fast" saturating kinetics.

moter region is similar across different promoters. This behavior is compatible with saturation of methylation capacity of the locally recruited DNMT3bCD. Alternatively, it could also reflect an effective interaction, in which unmethylated CpGs inhibit methylation at nearby CpG sites (Haerter et al., 2014).

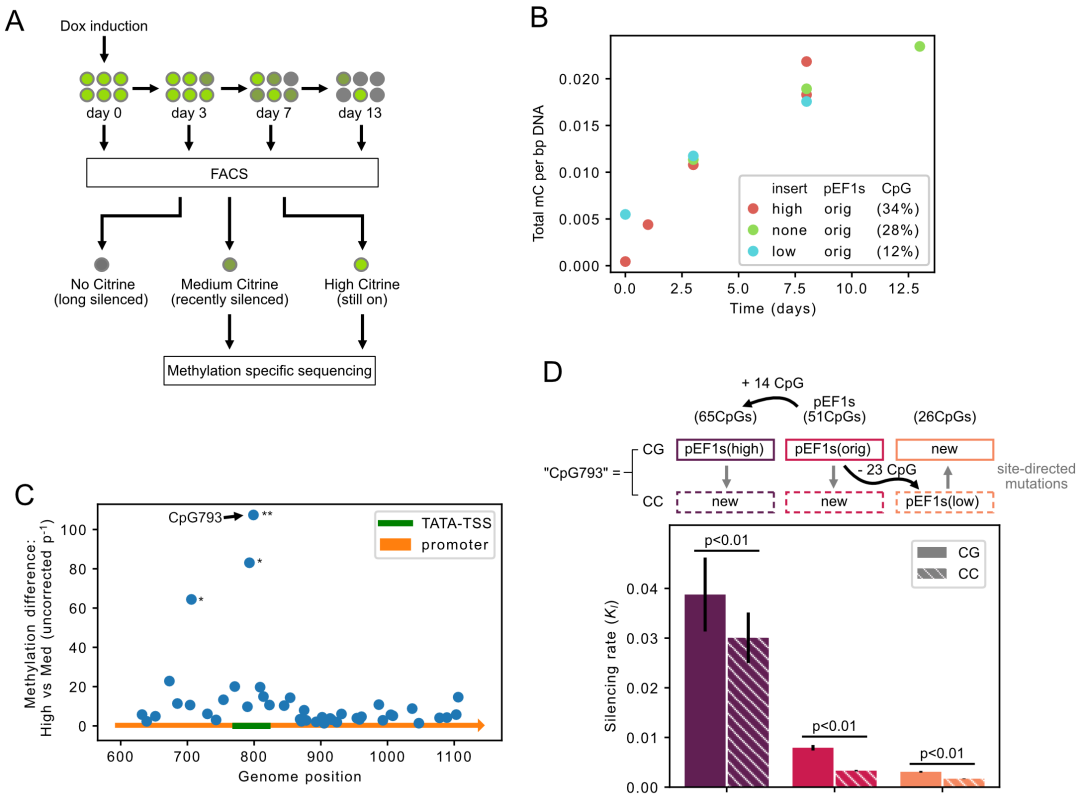


Figure 3.6: Sequencing reveals the constant accumulation of methylation, and potentially master CpGs.

(A) Schematics of the FACS-Sequencing experiment. Cells are treated with dox, and then FACS-sorted to three bins based on the Citrine brightness (high, med and low), consisting cells that are "still ON", "recently silenced" and "long silenced" respectively. The first two groups were analyzed by methylation specific sequencing.

(B) Methylation accumulates with time in the "still ON" cell population, regardless of promoter sequence: "still ON" populations are sorted out as indicated in A at intended dates, and subsequently analyzed by methylation specific sequencing (EM-seq), targeting the integrated gene promoter.

(C) CpGs around TATA-box and TSS (highlighted in green) show significant difference in methylation between the "still ON" and "recently silenced" group: Methylation percentages of different samples at different days were pooled together for comparison (a total of 10 from "still ON" vs. 6 from "recently silenced". P-values are from student t-test, and stars are given with the step-down multi-hypothesis correction (** corrected p<0.01; * corrected p<0.05).

(D) Mutation at CpG793 changes promoters' silencing rate significantly: Quantification of the silencing rate K_l of time course (Figure 3.6). K_l is estimated based on the model described by Equation 3.4 (Supplementary text). Error bars are generated by bootstrapping of the data points.

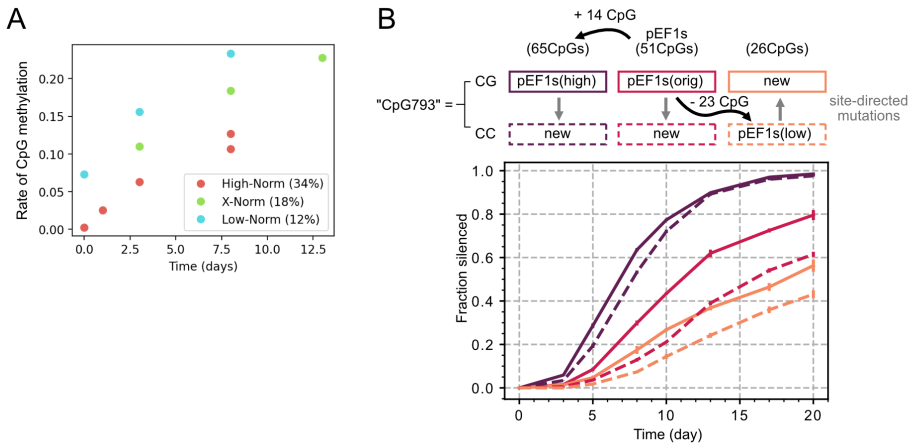


Figure 3.7: Mutation on CpG793 change silencing rate significantly.
(A) Methylation rates (number of methylated CpG on each individual CpG accumulate overtime, but are slower on CpG denser promoters. Same method as Figure 3.6B, except the y-axis is plotted as the number of methylated CpG over total CpG.
(B) Time course is performed similarly to Figure 3.4B. Error bar represents standard deviation from bootstrapping of triplicates.

A single CpG partially explains differences in silencing rate

The apparent discrepancy between the CpG density-dependent silencing rate and the density-independent methylation rate provoked the question of whether certain individual CpGs might play disproportionate roles in controlling silencing. Such CpGs would be expected to exhibit a significant difference between cell populations containing active versus recently silenced promoters.

To discover such CpGs, we pooled all available sequencing results from different days. Within the pEF1s region, where all three promoters overlap (90% of the pEF1s region) (Figure 3.6C), we identified three CpGs with significantly different methylation levels between the two expression groups ($p<0.05$, with step-down multi-hypothesis correction Romano and Shaikh, 2006). Interestingly, two of the most significant CpGs, including the top ranked one (CpG at position 793, or CpG793 for short), are located between the TATA box and the transcription start site (TSS) (arrow in Figure 3.6C), consistent with previous reports suggesting functionally important CpG islands around the TSS (Fenouil et al., 2012). CpG793 was among the CpGs that were eliminated in the construction of the low CpG pEF1s(low) promoter, consistent with the lower silencing rate observed for this promoter (Figure 3.4B).

To test for a functional role of CpG793, we mutated it to CpC in pEF1s(orig) and in pEF1s(high). Conversely, we also reverted this position back to CpG in pEF1s(low), where all 22 other CpGs including for CpG793 were mutated previously. Together, these constructs provided a set of controlled comparisons in which position 793 was either CC or CG in pEF1s(high), pEF1s(orig), and pEF1s(low) (Figure 3.6D, top).

We analyzed silencing rates for each of these promoters. These rates, K_l, were significantly ($p<0.01$) lowered in all three "CC" variants of promoters, compared to the CG variants (Figure 3.6D lower, and Figure 3.7B). Further, the CpG793 variant of pEFs1(low) and the CpC793 variant of pEFs1(orig) exhibited similar silencing rates despite differing systematically at 22 sites other than position 793. This suggests that a single CpG position could compensate for 22 other mutations.

3.4 Conclusion and discussion

While effects of sequence on DNMT-dependent gene silencing have long been observed, a controlled system in which to directly analyze the effects of sequence on silencing has not been available. Here, we constructed a library of synthetic promoters, featuring varying CpG content and DNMT-dependent silencing kinetics (Figure 3.1 and 3.2). Strikingly, silencing rate correlates directly with CpG content (Figure 3.4C). However, this correlation could not be explained by a corresponding effect of CpG content on methylation, as methylation accumulated at similar rates in all promoter variants (Figure 3.6B). Finally, we observed evidence that a certain CpG (CpG793), located between the TATA box and the TSS, can play a disproportionate role in control of silencing rate (Figure 3.6D). Together, these results should provide a versatile set of components for engineering synthetic epigenetic circuits with desired silencing behaviors, as well as a foundation for future investigations of the mechanisms of DNMT-dependent silencing. Finally, our observation that the DNA sequence-based substrate of epigenetic modifications could alter the regulation dynamics, might also apply into fully synthetic epigenetic circuits (Park et al., 2019).

A remaining mystery is why the rate of methylation accumulation is not correlated with the rate of silencing. One possible explanation is that silencing is affected by sequence-dependent, but methylation-independent, processes. For example, RNA secondary structure can induce silencing independent of DNA methylation (Ajjugal, Kolimi, and Rathinavelan, 2021). Modifying CpG content could also alter the binding of sequence specific DNA binding proteins and thereby affect silencing (Bird, 2002; Scala et al., 2020). It will also be interesting to see how these observations can

be generalized to non-constitutive promoters that are naturally regulated by methylation, especially since several previous reports suggested DNA methylation machinery can have different effects on regulated and constitutive promoters (Blackledge, Thomson, and Skene, 2013; Neri et al., 2013). Additionally, the sequencing results obtained here indicate that methylation accumulation is remarkably independent of CpG density (Figure 3.6B). This suggests possible saturation of methylation capacity of the locally recruited DNMT3bCD. In the future, however, one could test saturation of methylation capacity, by increasing the amount of local DNMT3b recruitment. Furthermore, it will be interesting to see if the observations made here can be extended to fully synthetic methylation-based gene regulation systems, both mammalian (Park et al., 2019) and bacterial (Maier, Möhrle, and Jeltsch, 2017). These and other directions should help to understanding of how sequence and regulatory proteins interact to control gene silencing, to enable predictive design of synthetic epigenetic circuits.

3.5 Supplementary text

To more efficiently fit the observed time traces to the mathematical model, we solved the first two proposed models analytically (time-independent β, and linearly time-dependent β, described in Equation 3.2 and 3.3) before fitting. Specifically, by combining Equation 3.1 and 3.2, we have:

$$\frac{dA}{dt} = -K_1(c) \cdot A \tag{3.5}$$

This can be solved by integrating the time-independent right hand part, with the assumption that the cells are 100% in "ON" state at beginning $A_{t=0} = 1$. We have the following for the time-independent rate model:

$$A(t) = e^{-K_1(c) \cdot t} \tag{3.6}$$

Similarly, the time-dependent model (described by Equation 3.3) can be solved to:

$$A(t) = \frac{1}{2} e^{-K_2(c) \cdot t^2} \tag{3.7}$$

In Figures 3.5A and B, Equation 3.6 and 3.7 are used respectively to fit the experimental time traces (Figure 3.4B). Each trace of different genotypes are fitted individually, resulting in 7 $K_1(c)$s and 7 $K_2(c)$s, respectively. For the model with

the saturation relationship (Equation 3.4), since there is no analytical solution to the model, we fitted the model with a single K_{max} shared by all genotypes, and 7 independent $K_l(c)$s for each genotype. All data points are pooled together and fitted with a least-square method (scipy package, Figure 3.5C and D).

Chapter 4

MULTICELLULAR CIRCUIT FOR IMMUNOTHERAPIES

4.1 Introduction

Engineered cell therapies represent one of the most exciting frontiers of biomedicine. In current cell-based immunotherapy, patient immune cells are isolated, stimulated, expanded, genetically modified to express synthetic receptors (e.g. chimeric antigen receptor, or CAR), and then reintroduced into patients, where they can kill targeted cell populations, such as tumors (Feldman et al., 2015; Rohaan, Wilgenhof, and Haanen, 2019). Currently, even the most successful cell therapies (Ramos, Heslop, and Brenner, 2016) focus on engineering a single modified cell type. By contrast, the natural immune system typically uses an army of specialized cell types that regulate one another's activity, proliferation, and death with great temporal and spatial precision. Here we ask whether a similar principle of multicellular specialization could enhance engineered cell therapy.

Here we propose a multicellular circuit consisting of 1) macrophages specialized to detect tumor cells ("Seekers") and 2) T-cells bearing auxin-activatable chimeric antigen receptors (CARs) that could kill cells when activated ("Killers"). More specifically, the Seeker cell will release auxin upon detection of tumor antigens and label the region, while the Killers only activate in the labeled tumor region. This first-of-its-kind multicellular circuit design should provide key advantages over single cell-type designs, including specialization of cell function, and reduced off-site cytotoxicity and T-cell exhaustion. More generally, it should demonstrate the potential of multi-cell type strategies for cell therapy.

More specifically, we will prototype our circuit based on the regime where the macrophage seekers will "seek" EGFRvIII and produce auxin to locally activate CAR bearing T cells for killing of Her2 positive cells (Figure 4.1). We chose EGFRvIII as our target for seeker, because it is a neoantigen that are only expressed by tumor cells, but is hard to target directly as tumor cells with lower or no EGFRvIII expression escapes (An et al., 2018). While Her2 is highly expressed by several types of tumor cells (including in breast cancer, lung cancer, etc. (Oh and Bang, 2020)), it is also broadly expressed by healthy epithelial cells. This makes Her2 dangerous for targeting systemically with high affinity antibody or CARs. With our multicellular

design, we hope to demonstrate that we can eliminate a Her2 positive tumors, with sparse expression of EGFRvIII, while not affecting Her2 cells elsewhere.

In the following sections, we will present advances of this ongoing project, demonstrating the auxin can be produced and sensed by engineered immune cells, and can locally label a tumor in mouse model.

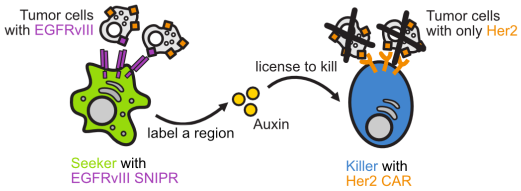


Figure 4.1: Seeker-Killer multicellular circuit for spatially localized tumor elimination.
Seeker (monocyte or macrophage) express a SNIPR synthetic receptor that can detect tumor specific EGFRvIII antigen, and produce auxin to label the region upon engagement. Then the auxin will activate the Killers (CAR bearing T-cells) for elimination of Her2 positive cells without EGFRvIII expression.

4.2 Auxin is produced and sensed by immune cells

To demonstrate that the auxin cell-cell communication system can be implemented in immune cells, we first implemented both the auxin sending and receiving components into cell lines derived from lymphocytes.

THP-1 cells produce auxin upon detection of cancer specific antigen

To construct the seeker, we start with THP-1 monocyte cells as they are easy to engineer and can be induced to differentiate to macrophages (Daigneault et al., 2010). We first test if THP-1 seekers with constitutively expressed auxin synthesis cassette, consisting of the two enzymes, iaaM and iaaH (Figure 4.2A), can produce auxin. Additional to the original construct, we designed a new optimized version with rearranged gene order (Figure 4.2B). The rearranged gene order should allow higher expression of the two proteins, as the T2A peptide is known to decrease translation efficiency of the downstream gene (Z. Liu et al., 2017).

The auxin production was measured by coculturing the senders with CHO receivers (Figure 2.2). As expected, both the original and optimized cassette drove the produc-

tion of auxin, as they decrease the mCherry in the receivers cells in a cell number-dependent manner (Figure 4.2C). Notably, our optimized cassette produced auxin at a similar level in THP-1 cells to the CHO senders from the original publication.

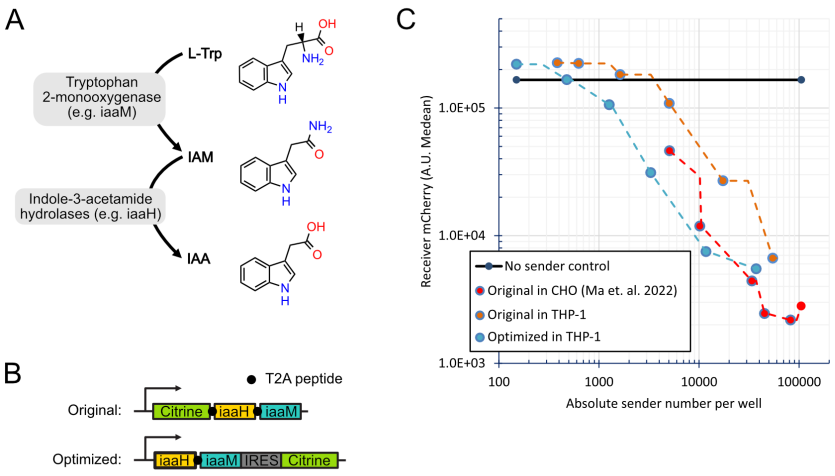


Figure 4.2: THP-1 cells produces auxin with optimized auxin synthesis module.
(A) Schematic of the auxin synthesis pathway, same as Figure 2.3A.
(B) The auxin synthesis plasmid is optimized by rearranging the component position on the transcribed mRNA. The expressions are all driven by a constitutive pEF1α promoter.
(C) CHO senders with the original synthesis module (from Ma et al., 2022), and THP-1 senders integrated with both the original and optimized module via lentiviral transduction are coculuted with CHO receivers (from Ma et al., 2022 with additional mTagBFP2 labeling). Sender cells are seeded with 3x titration series starting from 10,000 cell per well. Both senders and receivers are assayed by flowcytometry after 4 days, and are differentiated via BFP brightness in analysis.

THP-1 seekers produce auxin upon tumor specific antigen recognition

To construct seekers that could release auxin in response to detection of tumor specific antigen, we adopted the SNIPR receptor system (I. Zhu et al., 2022) with single chain antibody against EGFRvIII (R. A. Morgan et al., 2012) as extracellular domain. The EGFRvIII-recognizing SNIPR receptor will release Gal4-VP16 fusion transcriptional factor into the nucleus, inducing the expression of genes driven by a UAS fused minimal promoter (Figure 4.3A). We chose EGFRvIII as our target as it is a neoantigen that is only expressed by tumor cells, but is hard to target directly as tumor cells with lower or no EGFRvIII expression escape (An et al., 2018). We also engineered SKOV3 cells, a tumorigenic cancer cell line to express EGFRvIII as a target.

The THP-1 seeker cells successfully detected the EGFRvIII target and expressed the auxin cassette when cultured with target cells (Figure 4.3B and C). More importantly, the seeker cells produce a significant amount of auxin with the target cells (Figure 4.3D).

Auxin regulates CAR activity in Jurkat cells

For prototyping the killer circuit, we adopted a Jurkat cell (T-cell lymphoma-based cell line) based assay (Bloemberg et al., 2020). In order to make auxin able to activate CAR activity, we modified a 2nd generation CAR against Her2 (X. Liu et al., 2015), a tumor associated antigen expressed on SKOV3 cells. We fused a TEV protease cleavage site (tevcs) between the co-stimulation domain (4-1BB), and effective domain (CD3ζ). With this modification, the CAR should be rendered non-functional with the presence of TEV protease (TEVp). The TEVp is then tagged with AID, making it degradable with the presence of auxin. Therefore the CAR functionality will be activated when the cells sense auxin (Figure 4.4A). We integrated these components, together with the osTIR1 auxin receptor into Jurkat cell via lentiviral transduction, co-incubated the engineered cells with SKOV3 cells, and stained for the early T-cell activation mark CD69 (Figure 4.4B). We demonstrated that the CAR activity is almost suppressed to basal level with the TEV strategy, and can be recovered when auxin is added.

4.3 Auxin labels tumor locally in vivo

Another critical question about this system is whether auxin can label a tumor locally, or it defuses too fast and becomes a global signal inside an animal. To test this, we take advantage of the tumorigenic capability of CHO cells used in Chapter 2. More specifically, we took the full sender-receivers (Figure 2.3C) and receivers (Figure 2.4E, with additional label of mTagBFP2 expression), and implemented the mixture of the two subcutaneously into one flank of immuno-compromised NSG mouse and let it form a tumor (the proximal side). Simultaneously we implemented a receiver-only tumor on the other flank of the same mouse (the distal side) (Figure 4.5A). After about two week (exact time depends on the humane endpoint criteria), tumors are dissected out and assayed by flowcytometry. If the auxin signal is local, we should see only the proximal receivers showing down-regulated mCherry fluorescence. Indeed, we observed pair-wise, significant down-regulation of the mCherry only in the proximal side, and the distal receivers shows almost identical mCherry to that of receivers in the mouse without implemented senders (Figure 4.5B).

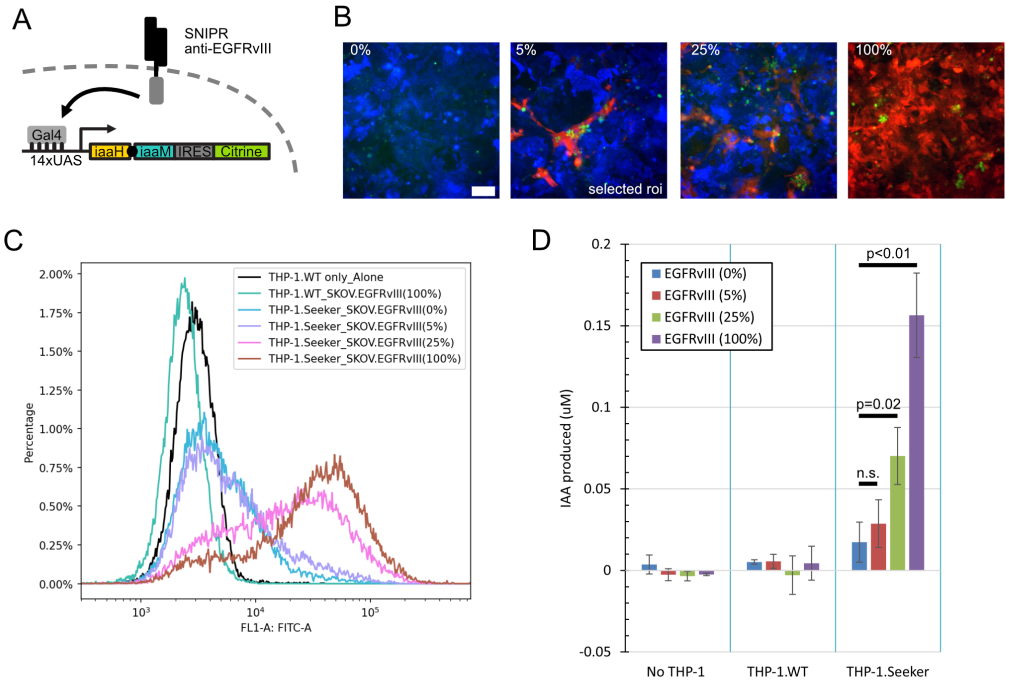


Figure 4.3: THP-1 seekers produce auxin upon tumor specific antigen recognition.
(A) Schematic of the SNIPR system. The SNIPR receptor detects surface expression EGFRvIII and release a Gal4-VP16 fusion transcriptional activator, triggering expression of auxin synthesis pathway.
(B) THP-1 seekers express citrine when coculture with EGFRvIII+ targets. THP-1 seeker cells coculutred with mixtures of two type of SKOV3 cells, that express either a control plasmid (mTagBFP2, blue), or EGFRvIII (mScarlet, red). Cells were imaged after 5 days of or coculture. Numbers on the upper left corner of each image indicate the percentage of EGFRvIII+ SKOV3 cells. Note the second image (from the right) is specially selected as most field of view of that group does not have EGFRvIII positive (red) cells. Initial effector (Seeker) to target (SKOV3) seeding ratio = 1:5. BAR = 100 micron.
(C) THP-1 seekers express citrine when cocultured with EGFRvIII+ target cells. Same setup as (B) except the cells were assayed by flowcytometry at the end.
(D) THP-1 seekers produced when cocultured with EGFRvIII+ target cells. Conditioned media was collected at the end of experiment (same setup as (B) and (C)), and diluted by 2x and applied to receiver CHO cells. A parallel experiment with receivers cultured with different auxin (IAA) was performed to calculated the standard curve, which is used for back-calculating the IAA concentration in the conditioned media. n = 3.

The above result demonstrated that auxin senders, when spatially restricted, can signal to their immediately neighboring cells, but not distal cells on the whole organism level.

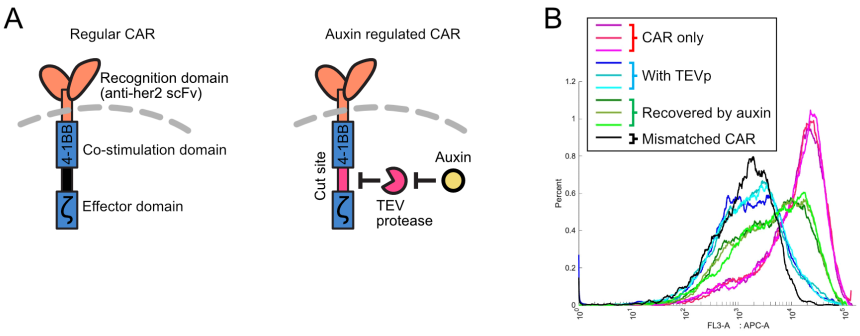


Figure 4.4: Auxin regulate CAR activity in Jurkat cells.
(A) Schematic of the auxin-TEVp regulated CAR. The left shows the regular 2nd generation CAR, consisting of a external antigen recognation domain (anti-her2 scFv), a co-stimulation domain (4-1BB) and a effector domain (CD3ζ). The right shows auxin regulated CAR, in which a TEV protease cleavage site (tevcs) is inserted between 4-1BB and CD3ζ. The auxin-regulated CAR is rendered non-functional with the presence of TEV protease (TEVp), which is degraded by auxin.
(B) Jurkat cell bearing the auxin-TEV regulated CAR is cultivatable by auxin. Jurkat cells that express the CAR, TEVp-AID and osTIR1 via lentiviral transduction are co-incubated with SKOV3 cells, and are stained for surface CD69 expression after 48 hours. Mismatched CAR recognize CD19, which is absent on target cells.

4.4 Conclusion and future directions

Our results above have demonstrated that we could implement both the auxin sending and receiving machinery into immune related cell types, and wire them to regulate related functions. Further experiments are needed to realize the full potential of this multicellular circuit.

Demonstration of auxin regulated killing in primary T-cells

The auxin regulated CAR system can be directly implemented and test in T-cell killing assay. Though Jurkat cells are derived from T-cell lymphoma, they do not have cytotoxicity against target cells, partially also due to they are more similar to CD4 T-cells. Furthermore, even in primary T-cell, early activation marker CD69 expression does not necessarily correlated with killing capacity (Bloemberg et al., 2020). Critically, engineered primary T-cells can be used for further in vivo testing in NSG mice with xenografted tumors, as a well-established model (see Choe et al., 2021; I. Zhu et al., 2022 etc.).

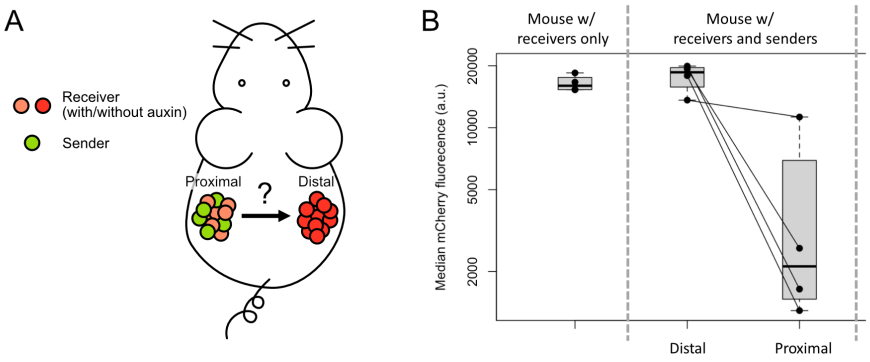


Figure 4.5: Auxin labels tumor locally in vivo.
(A) Schematic of the in vivo experiment. CHO sender-receivers (citrine labeled) and pure receivers (mTagBFP labeled) are mixed in 1:10 ratio, and then injected subcutaneously on one (proximal) flank of NSG mouse, while pure receivers were injected on the other (distal) flank. 5 million cells in total are injected on each flank. Another group of mice received only the distal flank injection as control.
(B) Mice were sacrificed upon humane endpoint, and tumors were dissected out for downstream analysis. The tumors were first homogenized by incubation with collagenase (2 mg/ml) and DNase I (0.1 mg/ml), and analyzed by flowcytometry. In flowcytometry analysis, We always gated on the mTagBFP2 high population for the receiver.

In vitro and in vivo demonstration of the circuit

The multicellular circuit needs to be tested in vitro first. In vitro test can be performed by co-incubation of seekers, killers (Jurkat or primary T cell based) and target cells with different EGFRvIII expression sparsity (similarly to Figures 4.3 and 4.4). The in vitro experiment should provide important info about how sparse can our seeker detect EGFRvIII bearing cells, and what is the optimal ratio (or absolute number) between seeker and killer. It is also interesting to see if we can induce the differentiation of THP-1 cells to macrophages effectively and also retain our circuit functionality.

In vivo experiments should be carried similarly to established CAR-T testing against xenografted tumor in mice. Initially, the seeker will be co-injected with the tumor as there is no active seeking component (more discussed below). However, it would also be interesting to see if the monocyte cells, or differentiated macrophages can ”seek” the tumor naturally.

Engineering the seeker function into primary macrophage

Our approach majorly focuses on engineering immortalized monocytes, due to the fact that the more clinically relevant PBMC derived monocytes and macrophages cannot be engineered effectively. However, recent studies have shown advances in this field, either by using a specially designed adenovirus (N. R. Anderson et al., 2021), or by engineering iPSCs and differentiating them into macrophages (Zhang et al., 2020). It would be interesting to see if we can use the same the strategy to engineer our seekers.

Engineering active seeking mechanism

Another missing part of our circuit is that our seekers are not engineered to be ”tumor-homing.” Though in our design, the auxin is not produced until the detection of the cancer specific target, preventing ”leaking” of auxin at off-tumor places, we still might face challenges in getting enough seeker cells into the tumor. Two potential strategies can be applied here: 1) Promoting macrophage proliferation after the engagement of the target. Though this is technically different from homing, an inplace proliferation can also enrich the macrophage number on the tumor site. This can potentially be achieve by expressing the macrophage survival factors (e.g. CSF-1 or its receptor, Hume and MacDonald, 2012). 2) Implementation of active seeking receptors. Previous studies have employed similar strategies such as expressing chemotaxis-enabling chemokine receptor (Craddock et al., 2010), or by expressing synthetic receptors that change the surface adhesive property of the seekers (Stevens et al., 2022).

Chapter 5

CONCLUSIONS AND DISCUSSIONS

In this thesis, I presented an auxin-based diffusive signaling module for mammalian cell-cell communication and based on this communication module, we achieved quorum-sensing, as well as robust population control in mammalian cells. We also demonstrated a DNA methylation-based, tunable gene switch that can be fine tuned by changing promoter CpG density and potentially drive stable cell-type switching and differentiation. Finally, I presented our ongoing effort on constructing auxin-based multicellular circuit in immune cells, demonstrating the potential of multicellular circuits in clinically relevant applications.

In this chapter, I will further discuss the potentials of using these components for multicellular circuits, and remaining biological questions about our discoveries.

5.1 Potential applications of the auxin communication module in synthetic development biology

Besides applications in immunotherapies (demonstrated in Chapter 4), the auxin communication system opens up the possibility of engineering other more complex multicellular circuits for developmental biology. For example, to record the relative distance between two groups of cells, one could engineer receivers that permanently activate in response to auxin secreted by a second cell type. Alternatively, one could take advantage of the two-step nature of auxin biosynthesis (Figure 2.3) by separating the steps into distinct cells, thereby enabling a proximity-dependent "AND gate", that enables recording of three cell groups' proximity information simultaneously.

More complex circuits for synthetic development could be achieved by combining the auxin with additional diffusible or contact dependent, orthogonal signals, as well as other synthetic gene circuit. This approach could enable synthetic bidirectional signaling (X. Zhou et al., 2018) or even Turing-like spatial pattern formation (Turing, 1990). Another possibility is to construct a synthetic segmentation clock circuit, as described in section 1.1. In natural context, segmentation clock is driven by a combination of gene oscillation, contact-dependent signaling, moving gradients formed by diffusive signals, and cell type determination (differentiation). With the auxin cell-cell communication module presented here, we now have synthetic version of

all the modules mentioned above, including gene oscillator (Tigges et al., 2009), synthetic surface receptor (SynNotch and SNIPR (Morsut et al., 2016; I. Zhu et al., 2022)), and various modules to define cell states with memory. It would be interesting to understand the fundamental mechanisms of segmentation by constructed a synthetic circuit with these components ”from bottom-up”.

5.2 Other molecules potentially serve as additional communication channels

Natural circuits (e.g. immune system (Boyman and Sprent, 2012)) rarely use only one channel for communications between cells. In fact, even a simple two-cell circuit for maintaining cell type ratio and population size requires a two-signal, bidirectional communication (X. Zhou et al., 2018). Therefore, it is critical to develop additional synthetic and orthogonal signal modules in mammalian system. Candidate molecules for additional channels include the plant hormones abscisic acid and gibberellin, both of which have partially mapped biosynthetic pathways (Mashiguchi et al., 2011; Finkelstein, 2013; Early and Martin, 1990) and were previously engineered to induce protein-protein interactions, enabling them to regulate transcription or protein localization in mammalian cells (Liang, Ho, and Crabtree, 2011; Miyamoto et al., 2012). The plant IP-CRE1 cytokinin system, which has been successfully ported to yeast, uses a plant-specific hormone as a diffusible signal to induce phosphorylation and further transcriptional activation, providing an additional potential signaling system (M.-T. Chen and Weiss, 2005). Bacterial autoinducers and synthetic proteins also provide attractive candidates for additional orthogonal signaling channels (Daringer et al., 2014; Morsut et al., 2016; Hong et al., 2012).

5.3 Remaining questions about how CpG density regulate silencing rate

Our methylation specific sequencing results (Figure 3.4B) demonstrated that regardless of the CpG densities on the promoter, the speed of CpG methylation (measured by total methylation) seems to be similar among promoters. Even more puzzling part is that, methylation rates, measured by methylation normalized to CpG number, are anti-correlated with promoter CpG density (Figure 3.7A). Considering the CpG density is tightly correlated with silencing speeds (Figure 3.4), the above observation provokes the question why we observed different silencing speeds when the underlying methylation are the same, or sometime suggests the opposite.

Here we will discuss the following possible explanations that can be further explored:

- The change of sequence by adding or removing CpGs could alter the transcription factors' (TFs) binding motifs. These changes on binding motif could have changed the composition of factors surrounding the promoter, and further change how CpG methylation silence the promoter. To test this hypothesis, one could query the surrounding sequences of each altered CpG against known TF binding motifs and ask if the change of CpG will disrupt existing, or introduce new motifs. However, this requires a high throughput approach, as most known motifs are defined as a short (normally 5 to 8 base pair) consensus sequence with ambiguous nucleotides (commonly defined as position probability matrix) (Castro-Mondragon et al., 2022). Due to this ambiguity, any sequence that is longer than several hundreds base pairs will likely have hundreds of potential known TF binding sites. Indeed, comparison between our modified promoters in the CIS-BP database (Weirauch et al., 2014) identified around one hundred binding motifs that are potentially altered (109 between pEF1s-Orig and pEF1s-Low; 99 between pEF1s-Orig and pEF1s-High). To fully delineate which TF binding motif is responsive for the difference in silencing speed we observed, a screening in the lists are likely to be needed.

- DNMT3b-CD we recruited to the promoter can silenced the gene through a methylation independent, but CpG dependent pathway. It is well documented that the protein DNMT3b, even with its catalytic domain deactivated, has significant function in regulating epigenetics and thus gene expression through its protein and heterochromatin interacting domain (Nowialis et al., 2019). Though we specifically used the "catalytic domain" version of the DNMT3b in this study, the ATRX domain that are shown to be associated with heterochromatin (T. Chen, Tsujimoto, and E. Li, 2004) still remains in this version. To further distinguish between methylation dependent and independent effect, one could use a catalytic inactivated version of DNMT3b as control.

- Besides alternating TF motifs, adding CpGs also create artificial CpG islands in and around the promoter, and CpG islands (regardless of methylated or not) can recruit TFs or epigenetic moderators. Though proteins that binds to unmethylated CpG islands are commonly considered to be associated with gene activation (Yin et al., 2017; Thomson et al., 2010), it is possible that the additional CpG island we introduced can recruit other proteins that are insensitive to methylation, and regulate gene silencing subsequently. For example, Tet1, a known protein that bind to CpG rich region but are insensitive to methylation,

can both maintain hypomethylation of a promoter (activating role), as well as interact with polycomb repressive complex 2 (PRC2) to induce gene silencing (H. Wu et al., 2011). Another similar example is KDM2B, that binds to CpG islands and recruit polycomb repressive complex 1 (PRC1), potentially inducing gene silencing (Farcas et al., 2012).

- Phenotypically, our finding is also very similar to the genetic mechanism of fragile X syndrome (FRX), in which an increased CGG repeat number upstream of the promoter leads to the abnormal hypermethylation and gene silencing during development (Garber, Visootsak, and Warren, 2008). The exact molecular mechanism leading to silencing in FRX is not yet fully understood, but various hypothesis, including toxic secondary RNA structure (Ajjugal, Kolimi, and Rathinavelan, 2021) and aberrant histone deacetylation (Coffee et al., 1999) has been proposed. It will be interesting to cross compare the promoters and inserts we created in this study to the promoter involved in FRX, and investigate if there is any similarity.

We also want to note that it is very likely that the above proposed hypothesises are not mutually exclusive, as ”interplay” between epigenetic marks and factors is very common (Atlasi and Stunnenberg, 2017). Therefore to solve this mystery, multiple aspects of the system, including but not limited to histone modification, TF interaction with specific DNA sequence, RNA secondary structure and post-transcriptional modifications might be required to be investigated further.

BIBLIOGRAPHY

Aalipour, Amin et al. (May 2019). "Engineered immune cells as highly sensitive cancer diagnostics." en. In: *Nature Biotechnology* 37.5, pp. 531–539.

Abedi, Mohamad H et al. (2022). "Ultrasound-controllable engineered bacteria for cancer immunotherapy." In: *Nature Communications* 13.1, p. 1585.

Ajjugal, Yogeeshwar, Narendar Kolimi, and Thenmalarchelvi Rathinavelan (Apr. 2021). "Secondary structural choice of DNA and RNA associated with CGG/CCG trinucleotide repeat expansion rationalizes the RNA misprocessing in FXTAS." en. In: *Science Reports* 11.1, p. 8163.

Allen, Greg M et al. (2022). "Synthetic cytokine circuits that drive T cells into immune-excluded tumors." In: *Science* 378.6625, eaba1624.

Amendola, Mario et al. (2005). "Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters." In: *Nature biotechnology* 23.1, pp. 108–116.

An, Zhenyi et al. (Mar. 2018). "Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies." en. In: *Oncogene* 37.12, pp. 1561–1575.

Anderson, Nicholas R et al. (2021). "Macrophage-Based Approaches for Cancer ImmunotherapyMacrophage-Based Approaches for Cancer Immunotherapy." In: *Cancer research* 81.5, pp. 1201–1208.

Antebi, Yaron E et al. (2017). "Combinatorial signal perception in the BMP pathway." In: *Cell* 170.6, pp. 1184–1196.

Argentova, V V et al. (2017). "Studies on the influence of different designs of eukaryotic vectors on the expression of recombinant IgA." In: *Moscow University Biological Sciences Bulletin* 72.2, pp. 63–68.

Atlasi, Yaser and Hendrik G Stunnenberg (2017). "The interplay of epigenetic marks during stem cell differentiation and development." In: *Nature Reviews Genetics* 18.11, pp. 643–658.

Attwood, J T, R L Yung, and B C Richardson (2002). *DNA methylation and the regulation of gene transcription*.

Bacchus, William et al. (Oct. 2012). "Synthetic two-way communication between mammalian cells." en. In: *Nature Biotechnology* 30.10, pp. 991–996.

Balagadde, F K (2005). "Long-Term Monitoring of Bacteria Undergoing Programmed Population Control in a Microchemostat." In: *Science* 309.5731, pp. 137–140.

Barnea, Gilad et al. (2008). "The genetic design of signaling cascades to record receptor activation." In: *Proceedings of the National Academy of Sciences* 105.1, pp. 64–69.

Berg, Stuart et al. (2019). "Ilastik: interactive machine learning for (bio) image analysis." In: *Nature methods* 16.12, pp. 1226–1232.

Berman, Helen M. et al. (Jan. 2000). "The Protein Data Bank." In: *Nucleic Acids Research* 28.1, pp. 235–242. ISSN: 0305-1048. DOI: 10.1093/nar/28.1.235. eprint: https://academic.oup.com/nar/article-pdf/28/1/235/9895144/280235.pdf. URL: https://doi.org/10.1093/nar/28.1.235.

Bier, E and E M De Robertis (2015). "BMP gradients: A paradigm for morphogen-mediated developmental patterning." In: *Science* 348.6242, aaa5838–aaa5838.

Bintu, Lacramioara et al. (Feb. 2016). "Dynamics of epigenetic regulation at the single-cell level." en. In: *Science* 351.6274, pp. 720–724.

Bird, Adrian (2002). "DNA methylation patterns and epigenetic memory." In: *Genes & Development* 16.1, pp. 6–21.

Blackledge, Neil P, John P Thomson, and Peter J Skene (Nov. 2013). "CpG island chromatin is shaped by recruitment of ZF-CxxC proteins." en. In: *Cold Spring Harbor Perspectives in Biology* 5.11, a018648.

Bloemberg, Darin et al. (2020). "A high-throughput method for characterizing novel chimeric antigen receptors in Jurkat cells." In: *Molecular Therapy-Methods & Clinical Development* 16, pp. 238–254.

Boyman, Onur and Jonathan Sprent (2012). "The role of interleukin-2 during homeostasis and activation of the immune system." In: *Nature Reviews Immunology* 12.3, pp. 180–190.

Bray, Nicolas L et al. (May 2016). "Near-optimal probabilistic RNA-seq quantification." en. In: *Nature Biotechnology* 34.5, pp. 525–527.

Bruno, Simone, Ruth J Williams, and Domitilla Del Vecchio (Apr. 2022). "Epigenetic cell memory: The gene's inner chromatin modification circuit." en. In: *PLoS Comput. Biology* 18.4, e1009961.

Cai, Long, Nir Friedman, and X Sunney Xie (2006). "Stochastic protein expression in individual cells at the single molecule level." In: *Nature* 440.7082, pp. 358–362.

Cao, Feng et al. (2010). "Comparison of gene-transfer efficiency in human embryonic stem cells." In: *Molecular imaging and biology* 12, pp. 15–24.

Castro-Mondragon, Jaime A et al. (2022). "JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles." In: *Nucleic acids research* 50.D1, pp. D165–D173.

Chen, Chih-Chiang et al. (Apr. 2015). "Organ-level quorum sensing directs regeneration in hair stem cell populations." en. In: *Cell* 161.2, pp. 277–290.

Chen, Ming-Tang and Ron Weiss (Dec. 2005). "Artificial cell-cell communication in yeast Saccharomyces cerevisiae using signaling elements from Arabidopsis thaliana." en. In: *Nature Biotechnology* 23.12, pp. 1551–1555.

Chen, Taiping, Naomi Tsujimoto, and En Li (2004). "The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin." In: *Molecular and cellular biology* 24.20, pp. 9048–9058.

Chevalier, Michael et al. (Oct. 2019). "Design and Analysis of a Proportional-Integral-Derivative Controller with Biological Molecules." en. In: *Cell System* 9.4, 338–353.e10.

Chmielewski, Markus et al. (Sept. 2011). "IL-12 release by engineered T cells expressing chimeric antigen receptors can effectively Muster an antigen-independent macrophage response on tumor cells that have shut down tumor antigen expression." en. In: *Cancer Res.* 71.17, pp. 5697–5706.

Cho, Jang Hwan, James J Collins, and Wilson W Wong (2018). "Universal chimeric antigen receptors for multiplexed and logical control of T cell responses." In: *Cell* 173.6, pp. 1426–1438.

Choe, Joseph H et al. (2021). "SynNotch-CAR T cells overcome challenges of specificity, heterogeneity, and persistence in treating glioblastoma." In: *Science translational medicine* 13.591, eabe7378.

Choi, Bryan D et al. (Sept. 2019). "CAR-T cells secreting BiTEs circumvent antigen escape without detectable toxicity." en. In: *Nature Biotechnology* 37.9, pp. 1049–1058.

Choudhury, Samrat Roy et al. (July 2016). "CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter." en. In: *Oncotarget* 7.29, pp. 46545–46556.

Chtarto, Abdelwahed et al. (2003). "Tetracycline-inducible transgene expression mediated by a single AAV vector." In: *Gene therapy* 10.1, pp. 84–94.

Coffee, Bradford et al. (1999). "Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells." In: *Nature genetics* 22.1, pp. 98–101.

Counsell, John R et al. (2017). "Lentiviral vectors can be used for full-length dystrophin gene therapy." In: *Scientific reports* 7.1, pp. 1–10.

Courchamp, Franck, Ludek Berec, and Joanna Gascoigne (2008). *Allee Effects in Ecology and Conservation*.

Craddock, John A et al. (2010). "Enhanced tumor trafficking of GD2 chimeric antigen receptor T cells by expression of the chemokine receptor CCR2b." In: *Journal of immunotherapy (Hagerstown, Md.: 1997)* 33.8, p. 780.

Curtin, JA et al. (2008). "Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct." In: *Gene therapy* 15.5, pp. 384–390.

Daigneault, Marc et al. (2010). "The identification of markers of macrophage differentiation in PMA-stimulated THP-1 cells and monocyte-derived macrophages." In: *PloS one* 5.1, e8668.

Danecek, Petr et al. (Feb. 2021). "Twelve years of SAMtools and BCFtools." en. In: *Gigascience* 10.2.

Daringer, Nichole M et al. (Dec. 2014). "Modular extracellular sensor architecture for engineering mammalian cell-based devices." en. In: *ACS Synthetic Biology* 3.12, pp. 892–902.

Davidson-Pilon, Cameron (2019). "lifelines: survival analysis in Python." In: *Journal of Open Source Software* 4.40, p. 1317. DOI: 10.21105/joss.01317. URL: https://doi.org/10.21105/joss.01317.

Del Vecchio, Domitilla and Richard M Murray (Oct. 2014). *Biomolecular Feedback Systems*. en. Princeton University Press.

Di Stasi, Antonio et al. (Nov. 2011). "Inducible apoptosis as a safety switch for adoptive cell therapy." en. In: *The New England Journal of Medicine* 365.18, pp. 1673–1683.

Dietrich, Lars EP et al. (2013). "Bacterial community morphogenesis is intimately linked to the intracellular redox state." In: *Journal of bacteriology* 195.7, pp. 1371–1380.

Din, M Omar et al. (Aug. 2016). "Synchronized cycles of bacterial lysis for in vivo delivery." en. In: *Nature* 536.7614, pp. 81–85.

Dullaers, Melissa et al. (2004). "Side-by-side comparison of lentivirally transduced and mRNA-electroporated dendritic cells: implications for cancer immunotherapy protocols." In: *Molecular Therapy* 10.4, pp. 768–779.

Early, Jack D and George C Martin (1990). "Photoperiodic regulation of vegetative growth and gibberellin metabolism in strawberry." In: *HortScience* 25.9, 1118e–1118.

Ehrlich, Melanie (Aug. 2002). "DNA methylation in cancer: too much, but also too little." en. In: *Oncogene* 21.35, pp. 5400–5413.

Elowitz, Michael B and Stanislas Leibler (Jan. 2000). "A synthetic oscillatory network of transcriptional regulators." en. In: *Nature* 403.6767, pp. 335–338.

Elowitz, Michael B, Arnold J Levine, et al. (2002). "Stochastic gene expression in a single cell." In: *Science* 297.5584, pp. 1183–1186.

Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet (2008). "A one pot, one step, precision cloning method with high throughput capability." In: *PloS one* 3.11, e3647.

Farcas, Anca M et al. (2012). "KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands." In: *elife* 1, e00205.

Feldman, Steven A et al. (2015). "Adoptive Cell Therapy—Tumor-Infiltrating Lymphocytes, T-Cell Receptors, and Chimeric Antigen Receptors." In: *Seminars in Oncology* 42.4, pp. 626–639.

Fenouil, Romain et al. (2012). "CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters." In: *Genome Research* 22.12, pp. 2399–2408.

Fernández, Lucía et al. (2019). "GMP-compliant manufacturing of NKG2D CAR memory T cells using CliniMACS prodigy." In: *Frontiers in Immunology* 10, p. 2361.

Finkelstein, Ruth (2013). *Abscisic Acid Synthesis and Response*. Vol. 11, e0166.

Fu, Xin et al. (2021). "CRISPR/Cas9 mediated somatic gene therapy for insertional mutations: the vibrator mouse model." In: *Precision Clinical Medicine* 4.3, pp. 168–175.

Gao, Xiaojing J et al. (Sept. 2018). "Programmable protein circuits in living cells." en. In: *Science* 361.6408, pp. 1252–1258.

Garber, Kathryn B, Jeannie Visootsak, and Stephen T Warren (2008). "Fragile X syndrome." In: *European journal of human genetics* 16.6, pp. 666–672.

Gardner, Timothy S, Charles R Cantor, and James J Collins (2000). "Construction of a genetic toggle switch in Escherichia coli." In: *Nature* 403.6767, pp. 339–342.

Ghassemi, Saba et al. (2022). "Rapid manufacturing of non-activated potent CAR T cells." In: *Nature biomedical engineering* 6.2, pp. 118–128.

Gibb, Sarah, Miguel Maroto, and J Kim Dale (Oct. 2010). "The segmentation clock mechanism moves up a notch." en. In: *Trends Cell Biology* 20.10, pp. 593–600.

Gibson, Daniel G et al. (May 2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." en. In: *Nature Methods* 6.5, pp. 343–345.

Greenberg, Maxim V C and Deborah Bourc'his (Oct. 2019). "The diverse roles of DNA methylation in mammalian development and disease." en. In: *Nature Review Mol. Cell Biology* 20.10, pp. 590–607.

Guindon, Stéphane et al. (May 2010). "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." en. In: *System Biology* 59.3, pp. 307–321.

Guye, Patrick et al. (2016). "Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6." In: *Nature communications* 7.1, p. 10243.

Haerter, Jan O et al. (Feb. 2014). "Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states." en. In: *Nucleic Acids Res.* 42.4, pp. 2235–2244.

Harborth, Jens et al. (2001). “Identification of essential genes in cultured mammalian cells using small interfering RNAs.” In: *Journal of cell science* 114.24, pp. 4557–4565.

Hart, Yuval et al. (Aug. 2014). “Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels.” en. In: *Cell* 158.5, pp. 1022–1032.

Holcik, Martin and Nahum Sonenberg (2005). “Translational control in stress and apoptosis.” In: *Nature Reviews Molecular Cell Biology* 6.4, pp. 318–327.

Hong, Seok Hoon et al. (Jan. 2012). “Synthetic quorum-sensing circuit to control consortial biofilm formation and dispersal in a microfluidic device.” en. In: *Nature Communications* 3, p. 613.

Hu, Biliang et al. (2017). “Augmentation of Antitumor Immunity by Human and Mouse CAR T Cells Secreting IL-18.” In: *Cell Reports* 20.13, pp. 3025–3033.

Hume, David A and Kelli PA MacDonald (2012). “Therapeutic applications of macrophage colony-stimulating factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling.” In: *Blood, The Journal of the American Society of Hematology* 119.8, pp. 1810–1820.

Jin, Bilian and Keith D Robertson (2013). “DNA methyltransferases, DNA damage repair, and cancer.” In: *Epigenetic alterations in oncogenesis*, pp. 3–29.

Kanehisa, Minoru et al. (Jan. 2021). “KEGG: integrating viruses and cellular organisms.” en. In: *Nucleic Acids Res.* 49.D1, pp. D545–D551.

Karin, Omer and Uri Alon (June 2017). “Biphasic response as a mechanism against mutant takeover in tissue homeostasis circuits.” en. In: *Molecular System Biology* 13.6, p. 933.

Kawaguchi, Masayoshi et al. (Mar. 1991). “The Presence of an Enzyme that Converts IndoIe-3-acetamide into IAA in Wild and Cultivated Rice.” In: *Plant Cell Physiology* 32.2, pp. 143–149.

Keener, J and J Sneyd (2009). *Mathematical physiology: systems physiology, vol. 2*.

Khakhar, Arjun et al. (2016). “Cell–Cell Communication in Yeast Using Auxin Biosynthesis and Auxin Responsive CRISPR Transcription Factors.” In: *ACS Synthetic Biology* 5.4, pp. 279–286.

Khalil, Ahmad S et al. (2012). “A synthetic biology framework for programming eukaryotic transcription functions.” In: *Cell* 150.3, pp. 647–658.

Kimura, Makoto, Akira Takatsuki, and Isamu Yamaguchi (Nov. 1994). “Blasticidin S deaminase gene from Aspergillus terreus (BSD): a new drug resistance gene for transfection of mammalian cells.” en. In: *Biochim. Biophys. Acta* 1219.3, pp. 653–659.

Kishore, Jugal, Manishkumar Goel, and Pardeep Khanna (2010). “Understanding survival analysis: Kaplan-Meier estimate.” In: *International Journal of Ayurveda Research* 1.4, p. 274.

Kitada, Tasuku et al. (2018). "Programming gene and engineered-cell therapies with synthetic biology." In: *Science* 359.6376, eaad1067.

Klichinsky, Michael et al. (Aug. 2020). "Human chimeric antigen receptor macrophages for cancer immunotherapy." en. In: *Nature Biotechnology* 38.8, pp. 947–953.

Kojima, Ryosuke, Leo Scheller, and Martin Fussenegger (2018). "Nonimmune cells equipped with T-cell-receptor-like signaling for cancer cell ablation." In: *Nature chemical biology* 14.1, pp. 42–49.

Krueger, Felix and Simon R Andrews (2011). "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications." In: *Bioinformatics* 27.11, pp. 1571–1572.

Kungulovski, Goran and Albert Jeltsch (Feb. 2016). "Epigenome Editing: State of the Art, Concepts, and Perspectives." en. In: *Trends Genetics* 32.2, pp. 101–113.

Landgraf, Dirk et al. (2012). "Segregation of molecules at cell division reveals native protein localization." In: *Nature Methods* 9.5, pp. 480–482.

Langmead, Ben and Steven L Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." In: *Nature Methods* 9.4, pp. 357–359.

Li, Pulin and Michael B Elowitz (2019). "Communication codes in developmental signaling pathways." In: *Development* 146.12, dev170977.

Liang, Fu-Sen, Wen Qi Ho, and Gerald R Crabtree (Mar. 2011). "Engineering the ABA plant stress pathway for regulation of induced proximity." en. In: *Science Signal.* 4.164, rs2.

Liao, Michael J et al. (Sept. 2019). "Rock-paper-scissors: Engineered population dynamics increase genetic stability." en. In: *Science* 365.6457, pp. 1045–1049.

Lin, Paul et al. (2012). "Efficient lentiviral transduction of human mesenchymal stem cells that preserves proliferation and differentiation capabilities." In: *Stem cells translational medicine* 1.12, pp. 886–897.

Liu, X Shawn et al. (2016). "Editing DNA Methylation in the Mammalian Genome." In: *Cell* 167.1, 233–247.e17.

Liu, Xiaojun et al. (2015). "Affinity-tuned ErbB2 or EGFR chimeric antigen receptor T cells exhibit an increased therapeutic index against tumors in mice." In: *Cancer research* 75.17, pp. 3596–3607.

Liu, Yue et al. (2021). "Chimeric STAR receptors using TCR machinery mediate robust responses against solid tumors." In: *Science translational medicine* 13.586, eabb5191.

Liu, Ziqing et al. (2017). "Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector." In: *Scientific reports* 7.1, p. 2193.

Long, Hannah K et al. (Aug. 2016). “Protection of CpG islands from DNA methylation is DNA-encoded and evolutionarily conserved.” en. In: *Nucleic Acids Res.* 44.14, pp. 6693–6706.

Love, Michael I, Wolfgang Huber, and Simon Anders (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” en. In: *Genome Biology* 15.12, p. 550.

Lövkvist, Cecilia et al. (June 2016). “DNA methylation in human epigenomes depends on local topology of CpG sites.” en. In: *Nucleic Acids Res.* 44.11, pp. 5123–5132.

Lyko, Frank (2018). *The DNA methyltransferase family: a versatile toolkit for epigenetic regulation*.

Ma, Yitong et al. (2022). “Synthetic mammalian signaling circuits for robust cell population control.” In: *Cell* 185.6, pp. 967–979. DOI: https://doi.org/10.1016/j.cell.2022.01.026.

Maier, Johannes A H, Raphael Möhrle, and Albert Jeltsch (May 2017). “Design of synthetic epigenetic circuits featuring memory effects and reversible switching based on DNA methylation.” en. In: *Nature Communications* 8, p. 15336.

Manhas, Janvie et al. (2022). “The evolution of synthetic receptor systems.” In: *Nature Chemical Biology* 18.3, pp. 244–255.

Marasini, Subash et al. (2017). “Effects of adenoviral gene transduction on the stemness of human bone marrow mesenchymal stem cells.” In: *Molecules and Cells* 40.8, p. 598.

Martinez-Corral, Rosa et al. (2018). “Bistable emergence of oscillations in growing Bacillus subtilis biofilms.” In: *Proceedings of the National Academy of Sciences* 115.36, E8333–E8340.

Mashiguchi, K et al. (2011). “The main auxin biosynthesis pathway in Arabidopsis.” In: *Proceedings of the National Academy of Sciences* 108.45, pp. 18512–18517.

Matsuda, Mitsuhiro et al. (Apr. 2012). “Synthetic signal propagation through direct cell-cell interaction.” en. In: *Science Signal.* 5.220, ra31.

Matsuura, Satoshi et al. (2018). “Synthetic RNA-based logic computation in mammalian cells.” In: *Nature communications* 9.1, p. 4847.

McCabe, Michael T, Johann C Brandes, and Paula M Vertino (2009). *Cancer DNA Methylation: Molecular Mechanisms and Clinical Implications*.

Mishra, Pradeep et al. (July 2016). “Characterization of an Indole-3-Acetamide Hydrolase from Alcaligenes faecalis subsp. parafaecalis and Its Application in Efficient Preparation of Both Enantiomers of Chiral Building Block 2,3-Dihydro-1,4-Benzodioxin-2-Carboxylic Acid.” en. In: *PLoS One* 11.7, e0159009.

Miyamoto, Takafumi et al. (Mar. 2012). “Rapid and orthogonal logic gating with a gibberellin-induced dimerization system.” en. In: *Nature Chemistry Biology* 8.5, pp. 465–470.

Moore, Lisa D, Thuc Le, and Guoping Fan (Jan. 2013). “DNA methylation and its basic function.” en. In: *Neuropsychopharmacology* 38.1, pp. 23–38.

Morgan, Richard A et al. (2012). “Recognition of glioma stem cells by genetically modified T cells targeting EGFRvIII and development of adoptive cell therapy for glioma.” In: *Human gene therapy* 23.10, pp. 1043–1053.

Morita, Sumiyo et al. (2016). “Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions.” In: *Nature Biotechnology* 34.10, pp. 1060–1065.

Morsut, Leonardo et al. (Feb. 2016). “Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors.” en. In: *Cell* 164.4, pp. 780–791.

Muldoon, Joseph J et al. (2020). “Macrophages employ quorum licensing to regulate collective activation.” In: *Nature communications* 11.1, p. 878.

Nakamura, Muneaki et al. (2021). “CRISPR technologies for precise epigenome editing.” In: *Nature Cell Biology* 23.1, pp. 11–22.

Natsume, Toyoaki et al. (Apr. 2016). “Rapid Protein Depletion in Human Cells by Auxin-Inducible Degron Tagging with Short Homology Donors.” en. In: *Cell Reports* 15.1, pp. 210–218.

Neri, Francesco et al. (Sept. 2013). “Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs.” en. In: *Cell* 155.1, pp. 121–134.

Nishimura, Kohei et al. (Dec. 2009). “An auxin-based degron system for the rapid depletion of proteins in nonplant cells.” en. In: *Nature Methods* 6.12, pp. 917–922.

Noll, Julia Han and Joseph A. Fraietta (2022). “Synthetic biology-based optimization of T cell immunotherapies for cancer.” In: *Current Opinion in Biomedical Engineering* 22, p. 100372. ISSN: 2468-4511. DOI: https://doi.org/10.1016/j.cobme.2022.100372. URL: https://www.sciencedirect.com/science/article/pii/S2468451122000058.

North, Khrystyna et al. (2022). “Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells.” In: *Nature biotechnology* 40.7, pp. 1103–1113.

Nowialis, Pawel et al. (2019). “Catalytically inactive Dnmt3b rescues mouse embryonic development by accessory and repressive functions.” In: *Nature communications* 10.1, p. 4374.

Nuñez, James K et al. (Apr. 2021). “Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing.” en. In: *Cell* 184.9, 2503–2519.e17.

Oh, Do-Youn and Yung-Jue Bang (2020). “HER2-targeted therapies—a role beyond breast cancer.” In: *Nature Reviews Clinical Oncology* 17.1, pp. 33–48.

Oliveira, Diane Sthefany Lima de et al. (2022). “Hinge influences in murine IgG binding to Cryptococcus neoformans capsule.” In: *Immunology* 165.1, pp. 110–121.

Papenfort, Kai and Bonnie L Bassler (Aug. 2016). “Quorum sensing signal-response systems in Gram-negative bacteria.” en. In: *Nature Review Microbiology* 14.9, pp. 576–588.

Park, Minhee et al. (Jan. 2019). “Engineering Epigenetic Regulation Using Synthetic Read-Write Modules.” en. In: *Cell* 176.1-2, 227–238.e20.

Petrásek, Jan et al. (May 2006). “PIN proteins perform a rate-limiting function in cellular auxin efflux.” en. In: *Science* 312.5775, pp. 914–918.

Potvin-Trottier, Laurent et al. (2016). “Synchronous long-term oscillations in a synthetic gene circuit.” In: *Nature* 538.7626, pp. 514–517.

Preibisch, Stephan, Stephan Saalfeld, and Pavel Tomancak (June 2009). “Globally optimal stitching of tiled 3D microscopic image acquisitions.” en. In: *Bioinformatics* 25.11, pp. 1463–1465.

Ramos, Carlos A, Helen E Heslop, and Malcolm K Brenner (2016). “CAR-T Cell Therapy for Lymphoma.” en. In: *Annual Review Medicine* 67, pp. 165–183.

Raven, JA (1975). “Transport of indoleacetic acid in plant cells in relation to pH and electrical potential gradients, and its significance for polar IAA transport.” In: *New Phytologist* 74.2, pp. 163–172.

Reitsma, Justin M et al. (Nov. 2017). “Composition and Regulation of the Cellular Repertoire of SCF Ubiquitin Ligases.” en. In: *Cell* 171.6, 1326–1339.e14.

Richards, Francis J (1959). “A flexible growth function for empirical use.” In: *Journal of experimental Botany* 10.2, pp. 290–301.

Robinson, David, James E. Anderson, and Jeong Long Lin (1990). “Measurement of diffusion coefficients of some indoles and ascorbic acid by flow injection analysis.” In: *The Journal of Physical Chemistry* 94.2, pp. 1003–1005.

Rohaan, Maartje W, Sofie Wilgenhof, and John BAG Haanen (2019). “Adoptive cellular therapies: the current landscape.” In: *Virchows Archiv* 474, pp. 449–461.

Romano, Joseph P and Azeem M Shaikh (2006). “On stepdown control of the false discovery proportion.” In: *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, pp. 33–50.

Ruder, Warren C, Ting Lu, and James J Collins (2011). "Synthetic biology moving into the clinic." In: *Science* 333.6047, pp. 1248–1252.

Rumbaugh, Kendra P and Karin Sauer (2020). "Biofilm dispersion." In: *Nature Reviews Microbiology* 18.10, pp. 571–586.

Sato, M et al. (Oct. 2012). "Determination of the optimal concentration of several selective drugs useful for generating multi-transgenic porcine embryonic fibroblasts." en. In: *Reprod. Domest. Anim.* 47.5, pp. 759–765.

Satyanarayana, A and P Kaldis (Aug. 2009). "Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms." en. In: *Oncogene* 28.33, pp. 2925–2939.

Sayers, Eric W et al. (Jan. 2022). "Database resources of the national center for biotechnology information." en. In: *Nucleic Acids Res.* 50.D1, pp. D20–D26.

Scala, Giovanni et al. (Feb. 2020). "DNA sequence context as a marker of CpG methylation instability in normal and cancer tissues." en. In: *Science Reports* 10.1, p. 1721.

Scarfò, Irene and Marcela V Maus (2020). "Unraveling the Signaling Balance of Activation and Exhaustion of CAR T Cells." In: *Cancer Cell* 37.2, pp. 143–144.

Scott, Spencer R and Jeff Hasty (Sept. 2016). "Quorum Sensing Communication Modules for Microbial Consortia." en. In: *ACS Synthetic Biology* 5.9, pp. 969–977.

Sekine, Ryoji, Tatsuo Shibata, and Miki Ebisuya (Dec. 2018). "Synthetic mammalian pattern formation driven by differential diffusivity of Nodal and Lefty." en. In: *Nature Communications* 9.1, p. 5456.

Sharrock, Abigail V et al. (n.d.). *NTR 2.0: a rationally-engineered prodrug converting enzyme with substantially enhanced efficacy for targeted cell ablation*.

Shih, Ryan M and Yvonne Y Chen (2022). "Engineering principles for synthetic biology circuits in cancer immunotherapy." In: *Cancer Immunology Research* 10.1, pp. 6–11.

Sievers, Fabian et al. (Oct. 2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." en. In: *Molecular System Biology* 7, p. 539.

Sitbon, F et al. (July 1992). "Transgenic Tobacco Plants Coexpressing the Agrobacterium tumefaciens iaaM and iaaH Genes Display Altered Growth and Indoleacetic Acid Metabolism." en. In: *Plant Physiology* 99.3, pp. 1062–1069.

Skylar-Scott, Mark A et al. (2022). "Orthogonally induced differentiation of stem cells for the programmatic patterning of vascularized organoids and bioprinted tissues." In: *Nature biomedical engineering* 6.4, pp. 449–462.

Smith, Zachary D and Alexander Meissner (2013). "DNA methylation: Roles in mammalian development." In: *Nature Reviews Genetics* 14.3, pp. 204–220.

Sockolosky, Jonathan T et al. (2018). “Selective targeting of engineered T cells using orthogonal IL-2 cytokine-receptor complexes.” In: *Science* 359.6379, pp. 1037–1042.

Sprinzak, David et al. (2010). “Cis-interactions between Notch and Delta generate mutually exclusive signalling states.” In: *Nature* 465.7294, pp. 86–90.

Stevens, Adam J et al. (2022). “Programming multicellular assembly with synthetic cell adhesion molecules.” In: *Nature*, pp. 1–3.

Stirling, Finn et al. (Oct. 2018). “Rational Design of Evolutionarily Stable Microbial Kill Switches.” en. In: *Mol. Cell* 72.2, p. 395.

Straathof, Karin C et al. (June 2005). “An inducible caspase 9 safety switch for T-cell therapy.” en. In: *Blood* 105.11, pp. 4247–4254.

Subramanian, Aravind et al. (Oct. 2005). “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.” en. In: *The Proceedings of the National Academy of Sciences* 102.43, pp. 15545–15550.

Szymczak, Andrea L et al. (May 2004). “Correction of multi-gene deficiency in vivo using a single ’self-cleaving’ 2A peptide-based retroviral vector.” en. In: *Nature Biotechnology* 22.5, pp. 589–594.

Takahashi, Yuta et al. (May 2017). “Integration of CpG-free DNA induces de novo methylation of CpG islands in pluripotent stem cells.” en. In: *Science* 356.6337, pp. 503–508.

Tanaka, H et al. (Dec. 2006). “Spatiotemporal asymmetric auxin distribution: a means to coordinate plant development.” en. In: *Cell. Mol. Life Science* 63.23, pp. 2738–2754.

Teague, Brian P, Patrick Guye, and Ron Weiss (2016). “Synthetic Morphogenesis.” In: *Cold Spring Harbor Perspectives in Biology* 8.9, a023929.

Thomson, John P et al. (2010). “CpG islands influence chromatin structure via the CpG-binding protein Cfp1.” In: *Nature* 464.7291, pp. 1082–1086.

Tigges, Marcel et al. (2009). “A tunable synthetic mammalian oscillator.” In: *Nature* 457.7227, pp. 309–312.

Toda, Satoshi, Lucas R Blauch, et al. (July 2018). “Programming self-organizing multicellular structures with synthetic cell-cell signaling.” en. In: *Science* 361.6398, pp. 156–162.

Toda, Satoshi, Wesley L McKeithan, et al. (Oct. 2020). “Engineering synthetic morphogen systems that can program multicellular patterning.” en. In: *Science* 370.6514, pp. 327–331.

Tornabene, Patrizia et al. (2019). “Intein-mediated protein trans-splicing expands adeno-associated virus transfer capacity in the retina.” In: *Science translational medicine* 11.492, eaav4523.

Trask, T W et al. (Feb. 2000). “Phase I study of adenoviral delivery of the HSV-tk gene and ganciclovir administration in patients with current malignant brain tumors.” en. In: *Mol. Ther.* 1.2, pp. 195–203.

Turing, Alan Mathison (1990). “The chemical basis of morphogenesis.” In: *Bulletin of mathematical biology* 52.1-2, pp. 153–197.

Urlinger, Stefanie et al. (2000). “Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity.” In: *Proceedings of the National Academy of Sciences* 97.14, pp. 7963–7968.

Vaisvila, Romualdas et al. (June 2021). “Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA.” en. In: *Genome Res.* 31.7, pp. 1280–1289.

Van, Mike V, Taihei Fujimori, and Lacramioara Bintu (Jan. 2021). “Nanobody-mediated control of gene expression and epigenetic memory.” en. In: *Nature Communications* 12.1, p. 537.

Waters, Christopher M and Bonnie L Bassler (2005). “Quorum sensing: cell-to-cell communication in bacteria.” In: *Annual Review of Cell and Developmental Biology* 21.1, pp. 319–346.

Weber, Michael et al. (Apr. 2007). “Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome.” en. In: *Nature Genetics* 39.4, pp. 457–466.

Weirauch, Matthew T et al. (2014). “Determination and inference of eukaryotic transcription factor sequence specificity.” In: *Cell* 158.6, pp. 1431–1443.

Wroblewska, Liliana et al. (2015). “Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery.” In: *Nature biotechnology* 33.8, pp. 839–841.

Wu, Hao et al. (2011). “Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells.” In: *Nature* 473.7347, pp. 389–393.

Yamaguchi, Shigeyuki et al. (Feb. 2011). “A method for producing transgenic cells using a multi-integrase system on a human artificial chromosome vector.” en. In: *PLoS One* 6.2, e17267.

Yin, Yimeng et al. (2017). “Impact of cytosine methylation on DNA binding specificities of human transcription factors.” In: *Science* 356.6337, eaaj2239.

You, Lingchong et al. (2004). “Programmed population control by cell–cell communication and regulated killing.” In: *Nature* 428.6985, pp. 868–871.

Zacharias, David A et al. (May 2002). “Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells.” en. In: *Science* 296.5569, pp. 913–916.

Zhang, Li et al. (2020). “Pluripotent stem cell-derived CAR-macrophage cells with antigen-dependent anti-cancer cell functions.” In: *Journal of hematology & oncology* 13.1, pp. 1–5.

Zhou, Xu et al. (Feb. 2018). “Circuit Design Features of a Stable Two-Cell System.” en. In: *Cell* 172.4, 744–757.e17.

Zhu, Heng, Guohua Wang, and Jiang Qian (Aug. 2016). “Transcription factors as readers and effectors of DNA methylation.” en. In: *Nature Review Genetics* 17.9, pp. 551–565.

Zhu, Iowis et al. (2022). “Modular design of synthetic receptors for programmed gene regulation in cell therapies.” In: *Cell* 185.8, pp. 1431–1443.

Zhu, Ronghui et al. (2022). “Synthetic multistability in mammalian cells.” In: *Science* 375.6578, eabg9765.

Appendix A

EXPERIMENTAL AND ANALYSIS METHODS IN THIS THESIS

A.1 Cell line handling and plasmid construction

Tissue culture and cell lines

CHO-K1 (Hamster cells, ATCC Catalog No. CCL-61) cells and their derivatives were grown on tissue culture grade plastic plates in Alpha MEM Earle's Salts, supplemented with 10% Tet System Approved FBS (ClonTech, or VWR), 100 U/ml penicillin, 100 µg/ml streptomycin, 0.292 mg/ml L-glutamine (GIBCO) or 1x GlutaMax (GIBCO). The complete media is filtered with 0.22 micron filters.

SKOV3 (ATCC Catlog No. HTB-77) cells are cultured in McCoy's 5A media (ThermoFisher), supplemented with 10% Tet System Approved FBS (ClonTech, or VWR), 100 U/ml penicillin, 100 µg/ml streptomycin, 0.292 mg/ml L-glutamine (GIBCO) or 1x GlutaMax (GIBCO). The complete media is filtered with 0.22 micron filters.

Both Jurkat (ATCC Catlog No. TIB-152) and THP-1 (ATCC Catlog No. TIB-202) cells are cultured in RPMI-1640 media (ThermoFisher, or ATCC), supplemented with 10% Tet System Approved FBS (ClonTech, or VWR), 100 U/ml penicillin, 100 µg/ml streptomycin, 0.292 mg/ml L-glutamine (GIBCO) or 1x GlutaMax (GIBCO), and 0.05 mM 2-mercaptoethanol. The complete media is filtered with 0.22 micron filters.

293FT (Thermofisher Catlog No. R70007) cells for lentiviral production are cultured in DMEM media (ThermoFisher), supplemented with 10% Tet System Approved FBS (ClonTech, or VWR), 100 U/ml penicillin, 100 µg/ml streptomycin, 1x NEAA (GIBCO Cat # 11140076), 1 mM sodium pyruvate, and 0.292 mg/ml L-glutamine (GIBCO) or 1x GlutaMax (GIBCO).

All cells above a maintained and passaged according to manufacture's instruction unless specified otherwise.

For long-term culturing demonstrated in chapter 2.3, cells were seeded in 24-well TC-treated plates (total media 500 µL per well) with imaging-grade plastic bottoms (ibidi inc. # 82406), and media was changed every 12 hours with one of the following methods:

- Adding and mixing 200 μL fresh media into the well, taking out the media, and putting back 500 μL (movie set 1 in Figure 2.11)
- Taking out the media. Centrifuge at around 2000 xg is applied for the old media to remove floating cells. Add back 350 μL and add 150 μL of fresh media (movie sets 2 and 3 in Figure 2.11)

Both methods simulate a media refreshing rate of 0.693/day (equivalent to media half-life = 1 day).

For the co-culturing experiments in Chapter 4 between THP-1, CHO, and SKOV3 cells, we use a mixture (1:1) of media of the involved cell types. For coculturing Jurkat and SKOV3 cells, Jurkat cell media is used.

Lentivirus production

To produce lentivirus for tranducing Jurkat, THP-1 and SKOV3 cells, we used the ViraPower system (Invitrogen Cat # K497500). More specifically, we followed the "Forward Transfection Procedure" in the manual, except virus are either produced in T75 flasks instead of 10 cm dishes suggested in the manual due to our lab's biosafty regulation. We also produced virus in T25 flasks. In that case, everything is scaled down to 1/3 of protocol for T75/10 cm protocol.

For transducing THP-1 cells, an additional step of concentrating the virus is added after the virus production. We used the lenti-X kit (Takara Bio, Cat # 631232), and concentrated the virus 5 folds. The final concentrated virus was re-suspended in complent THP-1 medea described above.

Cell line engineering via lipofection

To create each stable cell line for CHO cells, the following steps were used: 1) Cells were first transfected with 800-1000 ng of plasmid DNA using Lipofectamine 2000 or Lipofectamine LTX according to manufacturer's instructions. 2) 24-48 hours later, cells were transferred to selection media containing antibiotics (see below) as appropriate for 1-2 weeks. 3) If needed, single clones were isolated through the technique of limiting dilution.

For piggyBac constructs, the initial transfection consisted of the target plasmid along with the construct expressing the piggyBac transposase in a 1:4 mass ratio.

For integration of the reporters in Chapter 3, methods similar to previous literature (Bintu et al., 2016) were used. Briefly, we co-transfected 600 ng reporter plasmid

and 200 ng PhiC31 integrase plasmid using the method above.

Antibiotic concentration for selection media: Blasticidin 10-50 ng/ml; Geneticin 400 ng/ml

Cell line engineering via lentivirus

To establish stable cell lines via lentiviral transduction, 50k cells are seeded in 24 well plate, together with 400 μL of fresh media, 400 μL of media containing virus (see above) and polybrene (0.8 μg/ml final concentration, Sigma, Cat # TR-1003). Cells were incubated for 24 hours and put into fresh media after. For transduction with two virus, 200 μL of media with each virus is added.

THP-1 seeker cells are established by the viral transduction described above, and then sorted for positive citrine expression after coculture with EGFRvIII positive SKOV3 cell (labeled with mScarlet). After 96 to 120 hours of culculture, cells are harvested and sorted for mScarlt negative, Citrine positive. Sorted cells are then cultured and passaged for at least a week for resting before used for downstream assays.

Conditioned media

The process is described in Figure 2.4C. Cells were first seeded at about 20% confluence with fresh media for conditioning, and cultured for 2 days. The supernatant was collected as "conditioned media," and further filtered with 0.22 micron filter or centrifuged at 300 g for 3 minutes to remove any remaining cells. The conditioned media was then combined with fresh culture media at 1:1 ratio, and applied to receiving cells.

Plasmid construction

All plasmids are constructed using standard cloning techniques, including Gibson Assembly (NEB, Gibson et al., 2009) and GoldenGate Assembly (NEB, Engler, Kandzia, and Marillonnet, 2008). The plasmids and their maps are available by requests at Addgene (see Chapter B).

mAID and osTIR1 coding sequences were amplified from Addgene plasmids #72827 and #72834 (Natsume et al., 2016). iCasp9 coding sequence was amplified from addgene #15567 (Straathof et al., 2005). PIN2, mGFP, and iaaM coding sequences (CDS) were codon optimized for expressing in mice and synthesized as dsDNA at Integrated DNA Technology together with all oligos for cloning. Coding

sequences for screening indole-3-hydrolases were synthesized as cloning plasmid at Twist Bioscience.

A.2 Flowcytometry and FACS

Flowcytometry

Cells are washed with DPBS (GIBCO) first and then trypsinized with 0.25% trypsin-EDTA (GIBCO) for 1 minute at room temperature. The mixture was then neutralized by culture media and cells were resuspended in Hank's Balanced Salt Solution (GIBCO) with 2.5 mg/ml Bovine Serum Albumin (BSA). Cells were then filtered through a 40 μm cell strainer (Falcon) and analyzed by flow cytometry (MACSQuant VYB, Miltenyi or CytoFLEX, Beckman Coulter).

Analysis of data is done with the open source, in-house developed software, EasyFlow (`github.com/AntebiLab/easyflow`) or EasyFlowQ (`github.com/ym3141/EasyFlowQ`).

For counting cells, 1000 CountBright beads (Life Technologies) were spiked into the sample before filtering, gated out by their fluorescence in analysis, and used to estimate cell number in each sample.

FACS

For sorting in Chapter 2 and 3, FACS was performed with SY3200 Cell Sorter (Sony) at Caltech FLow Cytometry Facility.

For sorting in Chapter 4, cells were sorted on Sony MA900. Briefly, cells were collected by centrifuging at 500 g for 5 minutes, and are resuspended in Hank's Balanced Salt Solution (GIBCO) with 1% FBS. Cells were then filtered through a 40 μm cell strainer (Falcon) before sorting. THP-cells were sorted with chips with 130 μm nozzle and standard laser/filter configuration, and subsequently collected in full media.

A.3 Imaging setup and analysis

General imaging setup

For imaging experiments, cells were seeded at 24 or 96-well TC-treated plates with imaging grade plastic bottoms (ibidi inc.), as described above. For imaging in Figure 4.3B, media was replaced by FluoroBrite DMEM (Thermofisher, Cat # A1896701) right before imaging.

Snapshots were acquired on either 1) the EVOS imaging system (ThermoFisher)

with the EVOS AMG 4x objective (0.13 NA), or a 10x olympus objective (0.30 NA), with the system's default auto-scanning function; 2) the Olympus IX81 fluorescence microscope described below, but without the climate control system.

Time-lapse imaging

Time-lapse images were acquired on an inverted Olympus IX81 fluorescence microscope with Zero Drift Control (ZDC), an ASI 2000XY automated stage, iKon-M CCD camera (Andor, Belfast, NIR), and a 20x olympus objective (0.70 NA). Fluorophores were excited with an X-Cite XLED1 light source (Lumen Dynamics). Cells were kept in a custom-made environmental chamber enclosing the microscope, with humidified 5% CO2 flow at 37°C. Microscope and image acquisition were controlled by Metamorph software (Molecular Devices).

For results in Figure 2.11, 12, 25, and 36 positions (about 665um x 665um each) were imaged in each sample, for movie set 1, 2, and 3, respectively. For movie set 1, 12 imaging positions are distributed on a 3x4 grid with 3.5mm intervals. For movie sets 2 and 3, the positions are distributed in a 5x5 or 6x6 grid with 600μm intervals, and later stitched using Fiji's stitching plugin (Preibisch, Saalfeld, and Tomancak, 2009). Image of each position is acquired every hour for movie set 1, and every four hours for movie sets 2 and 3. Figure 2.7A were imaged with the same method as movie set 1.

Long-range gradient and analysis

For Figure 2.4, silicone-based inserts (ibidi inc. #80269) were first attached to the bottom of TC-treated 6cm dishes. Sender-receiver cells were seeded inside the inserts and allowed to settle down for 2 hours. The inserts are removed and the whole dish is washed with PBS twice to remove non-attached Sender-Receiver cells. Receiver cells were then seeded in the dish at approximately 20% confluence, and allowed to settle down for another 6 hours. To prepare agarose infused media, 2% low melting point agarose (EMD) was melted in alpha-MEM at 95°C for 10 minutes, and temperature was cooled to 42°C, before IAM and other ingredients of complete media (described above) were mixed in. Agarose infused media was applied to dishes with original media removed and allowed to solidify at room temperature for 20 minutes, before moved into the incubator.

For analysis, the images were background-subtracted, stitched together and masked by the constitutive mTagBFP2 fluorescent in blue channel (not shown) and quantified by summing up intensities of pixels that passed the mask. Error bars are standard

deviations of four images at the same distance.

Confluency estimation from images

Images, regardless of acquisition conditions, were converted to grey-scale for analysis in cases where pseudo-color was applied by the software automatically (a 2x2 binning is applied if acquired using the EVOS system). For each experiment, around 5 images were used in ilastik (Berg et al., 2019) pixel classification mode to train a classifier (decision tree-based), that classifies each pixel as cell or not cell (the trained models are available with the full data set). The classifier was then applied to the entire set of images, and output as probability masks. The masks were then analyzed to determine the fractions of "cell" pixels in the image. This value was then used to estimate confluence.

A.4 Animal experiment

Mouse line and handling

NSG mice used in Chapter 4 are from The Jackson Laboratory (Strain # 005557), and maintained by the Shapiro lab as described in Abedi et al., 2022.

Seeding of the tumors

Cells are grown to 80-100% confluence in the appropriate media and then trypsinized, spun down, and resuspended in sterile PBS for injection. For each subcutaneous tumor, 5 million cells in 100 μL are injected. Then the mouse grow for 12 days with tumor reaching approximately 1000 mm^3 in volume before sacrificed for analysis.

Tumor homogenization and analysis

Tumors were first dissected from mice scissors and crush into small pieces. Then they were digested with 2 mg/mL collagenase P and 0.1 mg/mL DNAse I in Leibovitz's L-15 medium (Thermofisher) at 37C for 2 hours. Cells were then spun down, and washed twice, and eventually resuspended with flowcytometry buffer described above. They were later analyzed by flowcytometry.

A.5 NGS and analysis

DNA-sequencing analysis

Cells were passaged for at least twice after the long term cultures, before harvested for genomic DNA extraction (Qiagen #69504, with on column RNA digestion). The extracted DNA was then amplified by PCR. Sequencing library were prepped and

indexed with Illumina Nextera XT kit (Illuminia #FC-131-1024) and sequenced on Illumina MiSeq, with 300 nt pair-ended setup.

The reads were first trimmed and filtered by Trim Galore! (Babraham Institute) and then aligned with Bowtie2 (Langmead and Salzberg, 2012) to the template sequence. For mutation calling (Figure 2.13 and Table 2.2), we calculated the distribution among four possible nucleotides at each position based on the result from the parental line, as our expected nucleotide distribution for each position (one pseudocount included to avoid significance calling due to low sampling size). We then compared the frequencies of such from other isolates to the expected distribution above to identify any significant deviation (Chi-Square test was used here as a good approximation for multinomial test. $p<0.05$ with Bonferroni correction).

All the significant results are plotted, with their corresponding isolates origin and frequency, in Figure 2.13A. Among these mutations, we assumed synonymous mutations follow normal distribution to set our expected mutation frequencies for neutral mutations, and determined any non-synonymous mutation that exceeds the 5th percentile (after Bonferroni correction) of this distribution is potentially positively selected (Figure 2.13A, red dotted line).

RNA-sequencing analysis

Cells were cultured for 2 days at under 80% confluence before harvested for bulk RNA extrusion (Qiagen #74104). The library preparation and sequencing was performed at Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech as service. Specifically, RNA integrity was assessed using Bioanalyzer (Agilent Technologies #5067-1513) and mRNA was isolated using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490). RNA-seq libraries were constructed using NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB #E7770) following manufacturer's instructions. Libraries were quantified with Qubit dsDNA HS Kit (ThermoFisher Scientific #Q32854) and the size distribution was confirmed with Bioanalyzer (Agilent Technologies #5067-4626). Libraries were sequenced on Illumina HiSeq2500 in single read mode with the read length of 50nt to the depth of 15 million reads per sample following manufacturer's instructions. Base calls were performed with RTA 1.13.48.0 followed by conversion to FASTQ with bcl2fastq 1.8.4.

The reads were first trimmed and filtered by Trim Galore! (Babraham Institute) and then counted with kallisto (Bray et al., 2016) with the CHO transcriptome (RefSeq

GCF_000223135.1) as a reference. Genes that are differentially expressed were determined with DESeq2 (Love, Huber, and Anders, 2014) and exported as ranked lists, with $sign(log2FoldChange) \cdot -log(padj)$ as weights. The top 30 genes in each of the three comparisons (negative vs control, paradoxical vs control and paradoxical vs negative), if not annotated with a symbol in the RefSeq file, are manually annotated by looking up their gene products in RSCB PDB (RSCB.org, Berman et al., 2000). The ranked gene list is then input into the GSEA (Subramanian et al., 2005) to perform pathway enrichment analysis, against the curated KEGG pathway annotation (MSigDB: c2.cp.kegg.v7.4.symbols), with mapping from mouse gene symbol to human orthologs.

Enzymatic methylation-specific sequencing

Cells were sorted as described above and immediately lysed for DNA extraction (DNeasy Blood & Tissue Kit, Qiagen). Total DNA is then converted with NEBNext Enzymatic Methyl-seq Conversion Module (NEB) according to the manufacturer's instructions, and further amplified (EpiMark Hot Start Taq DNA Polymerase, NEB) with primers targeting the promoter region and the gene body (nucleotide 1943-4438 on the none-pEF1s(orig) plasmid). The amplified targets are further prepared into library (Nextera XT Library Prep protocol, Illumina) and sequenced on the MiSeq (250bp pair ended, Illumina) platform.

The resulting reads were first trimmed and filtered by Trim Galore! (Babraham Institute) and then aligned and analyzed by Bismark (Krueger and S. R. Andrews, 2011) and SAMtools (Danecek et al., 2021) to generate the methylation calling statistics.

A.6 Miscellaneous data processing and statistical testing

Silenced fractions in Chapter 3

The background silenced fraction (S_{dox-}) from the no-recruitment control sample (no dox) is subtracted from the observed silenced fraction (S_{dox+}) from the with-recruitment group experiment, and further normalized by the "fraction still available for silencing" ("still ON" fraction in the control$1-S_{dox-}$. Consequently, the silenced rate of a given sample is calculated as $S = \frac{S_{dox+}-S_{dox-}}{1-S_{dox-}}$.

Note we observed that the highest background silencing is about 35% in the insert(high)-pEF1s(high) genotype at 25th day (last time points in Figure 3.4B).

Bootstrapping and significant tests

All statistical testings in this study are Student's t-test if not specified otherwise.

The curve and significance test in Figure 2.11C are based on the Kaplan-Meier estimation of the "survival rate." Between negative and paradoxical feedback, a log-rank test is performed with the null hypothesis that there is no significant difference between the estimated survival rate (Kishore, Goel, and Khanna, 2010), using the lifeline python package (Davidson-Pilon, 2019).

For expression ratios (Figure 2.12C) and survival rates (Figure 2.12D and E), the means and standard deviations were calculated by normalizing treatment replicates to control replicates with bootstrapping. Specifically, for each isolate, we perform replicates for both the treatment (n=3 for all) and control (n=3 for 2.12C, n=2 for 2.12D and E) group, then normalize the measurements from the treatment groups to the control groups, with all possible combinations (n=9 for 2.12C, n=6 for 2.12D and E). The means and standard deviations of all the normalized measurements were reported as a single data point with bars in the corresponding figures

Similarly, for each time point in Figure 3.6B, each of the three "with recruitment" samples are normalized to each of three "no recruitment" control samples, according to the method described above. Therefore a total of 9 data points are generated, and the error bars represent the standard deviation of these points. To generate standard deviation of the estimated silencing rate in Figure 3.6D, we bootstrapped the data points in Figure 3.6D. Specifically, in each round, 10 of the 64 data points (9 points for each of 7 time points and plus one at (0, 0)) are randomly selected and fitted to the model. We bootstrapped 6 times, and bars represent the standard deviation.

Appendix B

DATA AVAILABILITY AND DEPOSITS

The plasmids used in Chapters 2 and 3 have been deposited at addgene, under the article numbers #28212030 and #28233817, respectively.

Matlab code for mathematical modeling of the paradoxical circuit (section 2.3) is available at Github (`github.com/labowitz/auxin_paradox_matlab_code`). Imaging analysis code is available at Github (`github.com/labowitz/MovieTools`).

For section 2.3, the raw data of DNA sequencing for detection of mutations are available as ACCN. PRJNA784643 at BioProject (`ncbi.nlm.nih.gov/bioproject/`). Raw and processed data of RNA sequencing for pathway analysis are available at ACCN. GSE189948 at Gene Expression Omnibus (`ncbi.nlm.nih.gov/geo/`).

Raw and processed data of methylation specific sequencing of Chapter 3 is also available at Gene Expression Omnibus (`ncbi.nlm.nih.gov/geo/`), under ACCN. 224403.

The plasmid GenBank files, raw data, and processing/plotting scripts for generating the figures shown in Chapters 2 and 3 are available at data.caltech.edu, `doi.org/10.22002/D1.1613` and `doi.org/10.22002/ct5kt-cv878`, respectively.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Elowitz lab.

INDEX