

Design and Engineering of Synthetic Gene Networks

by

Fuqing Wu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2017 by the
Graduate Supervisory Committee:

Xiao Wang, Chair
Karmella Haynes
Pamela Marshall
David Nielsen
David Brafman

ARIZONA STATE UNIVERSITY

August 2017

ABSTRACT

Synthetic gene networks have evolved from simple proof-of-concept circuits to complex therapy-oriented networks over the past fifteen years. This advancement has greatly facilitated expansion of the emerging field of synthetic biology. Multistability is a mechanism that cells use to achieve a discrete number of mutually exclusive states in response to environmental inputs. However, complex contextual connections of gene regulatory networks in natural settings often impede the experimental establishment of the function and dynamics of each specific gene network.

In this work, diverse synthetic gene networks are rationally designed and constructed using well-characterized biological components to approach the cell fate determination and state transition dynamics in multistable systems. Results show that unimodality and bimodality and trimodality can be achieved through manipulation of the signal and promoter crosstalk in quorum-sensing systems, which enables bacterial cells to communicate with each other.

Moreover, a synthetic quadrastable circuit is also built and experimentally demonstrated to have four stable steady states. Experiments, guided by mathematical modeling predictions, reveal that sequential inductions generate distinct cell fates by changing the landscape in sequence and hence navigating cells to different final states.

Circuit function depends on the specific protein expression levels in the circuit. We then establish a protein expression predictor taking into account adjacent transcriptional regions' features through construction of ~120 synthetic gene circuits (operons) in *Escherichia coli*. The predictor's utility is further demonstrated in evaluating

genes' relative expression levels in construction of logic gates and tuning gene expressions and nonlinear dynamics of bistable gene networks.

These combined results illustrate applications of synthetic gene networks to understand the cell fate determination and state transition dynamics in multistable systems. A protein-expression predictor is also developed to evaluate and tune circuit dynamics.

DEDICATION

To my parents, Yihe and Ailan, who have always encouraged and supported me.

Thank you for your love and blessings.

To my wife, Ya, who has supported me in this journey.

Thank you for being there for me.

ACKNOWLEDGMENTS

First and foremost, I offer my sincerest gratitude to my advisor, Dr. Xiao Wang. His continuous support, patient guidance, and teachings have cultivated the scientist that I currently am. In his laboratory, I was afforded great opportunities and freedom to do what I want, and order what I need. I am grateful for his constant encouragement and belief in me during the bad times of the project.

I would also like to express my gratitude to my committee members: Dr. Karmella Haynes, Dr. Pamela Marshall, Dr. David Nielsen, Dr. David Brafman, and Dr. David Frakes, for their precious time and insightful comments and advices, which helped me to improve the dissertation.

I would like to thank my colleagues Riqi Su, Qi Zhang, David Menn, and Lezhi Wang for participating all my projects. Thanks to all the other Wang lab members for their advice and discussions on the projects. I offer my regards to all those who supported me in any way in preparing this dissertation.

My warmest thanks and heartfelt gratitude to my family for their constant support and encouragement. I am especially grateful to my sister, Lingling, for taking care of our parents when I am away from home during the five years.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER	
1. INTRODUCTION	1
1.1 Synthetic Gene Networks	1
1.2 Applications of Synthetic Gene Networks	3
1.2.1 Build to Understand: Multistability	3
1.2.2 Build to Understand: Pattern Formation	6
1.2.3 Synthetic Gene Networks in Single-cell Organisms: Biosensor, Biofuel, and Biopharmaceutical	8
1.2.4 Therapy-oriented Synthetic Gene Networks in Mammalian Cells	10
2. QUORUM-SENSING CROSSTALK DRIVEN SYNTHETIC CIRCUITS: FROM UNIMODALITY TO TRIMODALITY	12
2.1 Introduction	12
2.2 Results	14
2.2.1 Dissecting the Crosstalk Between LuxR/I And LasR/I Using Synthetic Circuits	14
2.2.2 Signal Crosstalk Induces Distinct Responses from Positive Feedback Circuits	17
2.2.3 Promoter Crosstalk Induces Unexpected and Complex Bimodal Responses	22

CHAPTER	Page
2.2.4 Bimodality Results From Circuit-Host Interactions	24
2.2.5 Trimodality Predicted by Expanded Model	26
2.2.6 Experimental Validation of Trimodal Responses by Lowering Growth Temperature	29
2.3 Discussion	34
2.4 Materials and Experimental Methods	36
2.4.1 Strains, Growth Conditions and Media	36
2.4.2 Plasmids Construction	37
2.4.3 Flow Cytometry	38
2.4.4 Hysteresis Experiment	38
2.4.5 Growth Curve Assay	39
2.4.6 Microfluidics, Fluorescence Microscopy, and Image Processing	40
2.5 Mathematical Modeling	41
2.5.1 Deterministic Model Construction	41
2.5.2 Stochastic Simulation Coupled with Genetic Mutation	45
2.5.3 Determinations of Parameter Values	48
3. ENGINEERING OF A SYNTHETIC QUADRASTABLE GENE NETWORK TO APPROACH WADDINGTON LANDSCAPE AND CELL FATE DETERMINATION	54
3.1 Introduction	54
3.2 Results	56
3.2.1 MINPA Circuit Construction and Multistability Analysis	56
3.2.2 Systematical Multistability Evaluation of MINPA and its Sub-Networks	58

CHAPTER	Page
3.2.3 Bifurcation and Hysteresis Verification of Multistability	65
3.2.4 Experimental Demonstration of Model-Guided Quadrastability of MINPA	69
3.3 Discussion	77
3.4 Materials and Experimental Methods	78
3.4.1 Strains, Media, and Chemicals	78
3.4.2 Plasmids Construction	79
3.4.3 Flow Cytometry	80
3.4.4 Sequential Induction and Hysteresis	81
3.4.5 Cell Growth Array Under Inductions	83
3.4.6 Microfluidics and Microscopy	83
3.5 Mathematical Modeling	85
4. DEVELOPMENT OF A PROTEIN EXPRESSION PREDICTOR TO TUNE GENE EXPRESSION AND CIRCUIT DYNAMICS	86
4.1 Introduction	86
4.2 Results	88
4.2.1 Protein Expression is Significantly Influenced by its Adjacent Genes And Position in the Operon	88
4.2.2 Quantitative Characterizations of Adjacent Gene Regulation in Synthetic Operons	92
4.2.3 Protein Expression is also Affected by Non-Coding ATR Sequences	96
4.2.4 Comprehensive Model of ATR Regulation	98

CHAPTER	Page
4.2.5 Predictor-Guided Circuit Design for Synthetic Logic Gates	101
4.2.6 Tuning Gene Expression with Synthetic 5' Adjacent Transcriptional Regions	106
4.2.7 Using Synthetic ATRs to Tune Toggle Switches	109
4.3 Discussion	111
4.4 Materials and Experimental Methods	115
4.4.1 Strains, Media and Chemicals	115
4.4.2 Plasmid Construction	115
4.4.3 Flow Cytometry Measurements	116
4.4.4 RT-qPCR	119
4.4.5 Sample Preparation and Microscopy	121
4.4.6 Hysteresis Experiment	121
4.4.7 Minimum Free Energy Calculation	122
4.5 Statistical Analysis and Mathematical Modeling	122
4.5.1 Statistical Analysis and Model Fitting	122
4.5.2 Deterministic Model Construction and Prediction for the Logic Gate ..	136
4.5.3 Parameter and Bifurcation Analysis for the Synthetic Toggle Switch Circuit	140
5. CONCLUSIONS AND OUTLOOK	141
5.1 Conclusions	141
5.2 Outlook	143
REFERENCES	146

LIST OF TABLES

Table	Page
1. Autoinducer Information	13
2. Plasmids Used in the Circuits' Construction	38
3. Parameters for the Genetic Mutation Event in the Stochastic Simulation Of Lasr- <i>PLux</i> Positive Feedback System.....	52
4. Parameters for the Three Positive Feedback Models.....	53
5. Components from the Registry of Standard Biological Parts.....	81
6. Model Evaluation for Each Gene's Expression in the AND Logic Gate	104
7. Igem Registry of Standard Biological Parts Used in this Study.	117
8. Transcriptional Factors Amplified from <i>E. Coli</i> Wild Type Strain.	119
9. Model Selection for Construct of <i>X-GFP</i> in Figure 26A.....	125
10. Model Selection for Construct of <i>X-GFP-Y</i> in Figure 26C	126
11. Model Selection for Construct of <i>GFP-X</i> in Figure 26E	127
12. A Combined Linear Model for All the Constructs in Figure 28.....	129
13. Model Selection for Noncoding ATR Constructs in Figure 28	133
14. A Comprehensive Model with Energetic Changes for the Constructs in Figure 26.	134
15. A Comprehensive Model with Energetic Changes for Noncoding ATR Constructs in Figure 28.....	135

LIST OF FIGURES

Figure	Page
1. Diagram of Assembly and Applications of Synthetic Gene Networks.....	10
2. QS Crosstalk Dissected Using Synthetic Gene Circuits.....	16
3. Characterizing the Crosstalk to the <i>PLux</i> and <i>Plas</i> Promoters Using Synthase Genes.	19
4. Signal Crosstalk Causes Shrinkage of Bistable Region.....	20
5. Promoter Crosstalk Induces Mutation and Leads to Population Heterogeneity	23
6. Hysteresis of The <i>Lasr-PLux</i> Circuits.....	25
7. Model Predictions and Experimental Validations of Mutation-Induced Trimodality ..	28
8. Temperature Influence on Cell Growth Rate and Population Distribution	30
9. Fluorescence Microscopy Validation of Mathematical Model Predictions.....	32
10. Quasi-Potential U and the Transition Dynamics Between Stable Steady States in the <i>Lasr-PLux</i> Positive Feedback System (Without Genetic Mutation).....	33
11. Parameters Determination in the Model	51
12. Conceptual and Experimental Design of MINPA and Its Sub-Networks.....	58
13. Abstract Diagrams and Molecular Implementation of the Eight MINPA Sub-Network Topologies.....	60
14. Sequential Induction Rational and Experimental Validation in Synthetic Toggle Switch Circuit	62
15. Sequential Induction of MINPA and Its Sub-Networks	64
16. Time Course Results of Sequential Induction for the MINPA (T15) Circuit.....	66
17. MINPA Has Four Individual Cell States and Bifurcation Analysis of MINPA	68
18. Hysteresis of MINPA with Induction of Arabinose and AHL (<i>AR/AL</i>).....	70

Figure	Page
19. Model-Guided Quadrastability of MINPA Through Triple Induction	72
20. Experimental Validations of Model-Predicted Quadrastability Using Flow Cytometry	73
21. Cells' States Under Induction with the First Inducer, Microfluidic Results to Demonstrate Quadrastability with IPTG and Atc Induction, and Time Course of Sequential Induction of AHL, Atc And Ara	75
22. Cell Growth Rates Under Each Inducer and Inducer Combinations.	76
23. Protein Expression is Significantly Influenced by its Adjacent Genes and Position in Synthetic Operons	89
24. RT-qPCR Results to Investigate mRNA Level for the Circuits in Figure 23B.	90
25. Gene Position in the Tri-Cistronic Circuit Impacts GFP Expression	91
26. Quantitative Characterization of Adjacent Gene Regulation in Synthetic Operons...	93
27. Sliding Window Analysis for Local GC Content and Model Fitting	95
28. Quantitative Characterization of Non-Coding ATRs Regulation on Gene Expression	97
29. Comparison of GFP Expression Between Coding ATRs and Noncoding ATRs	98
30. A General Model for ATR Regulation on Gene Expression.	99
31. Model Guided Circuit Design for Synthetic Logic Gates.....	102
32. Model Simulation and Experimental Validation of GFP Dynamics for Synthetic Logic Gates	103
33. Tuning Gene Expression with Synthetic 5' Adjacent Transcriptional Regions	107
34. Using Synthetic ATRs to Modulate Bistability of Toggle Switches	110

Figure	Page
35. Fit Diagnostics in the Combined Model Integrating the Three Scenarios in Figure 26	130
36. Outlier and Leverage Diagnostics for the Response (GFP) in the Combined Model	131

CHAPTER 1

INTRODUCTION

Synthetic gene networks have evolved from simple proof-of-concept circuits to complex therapy-oriented networks over the past fifteen years. This advancement has greatly facilitated expansion of the emerging field of synthetic biology. Synthetic gene networks have been developed for understanding biological design principles, developing biosensors for diagnosis, producing industrial and biomedical compounds, and treating human diseases.

1.1 Synthetic Gene Networks

After fifteen years of rapid development, synthetic biology has started to become an engineering discipline with the aim to create, control and program cellular behaviors for basic research, and industrial and biomedical applications^{1,2}. By taking a bottom-up approach, synthetic biologists rationally assemble various biological modules like genetic parts (promoters, coding sequence, terminators, etc...) together to create a “circuitry”, which is often called a synthetic gene circuit or network, to carry out a wide range of functions. In an integrated circuitry, genes do not work independently but in a network. For example, the lysis-lysogeny decision of bacteriophage lambda is tightly regulated by CI-Cro genetic switch³. CI protein can repress Cro to maintain a lysogenic state, while Cro inhibits CI expression to promote lytic development.

Most existing synthetic gene networks utilize one or more of the three regulation mechanisms: transcriptional, translational, and post-translational^{4,5}. Transcriptional circuits are engineered based on cell transcriptional regulatory machinery involving promoters, transcriptional factors, and RNA polymerase. Early examples of such synthesized circuits include the toggle switch which is constructed from a pair of repressor genes that inhibit the transcription of each other, and can be induced to flip between two stable steady states⁶, and the repressilator which is composed of three repressor-promoter interactions to form a cyclic negative-feedback loop and exhibits periodic oscillating behavior⁷. Since then, a myriad of transcriptional networks for diverse purposes have been constructed, including negative feedbacks⁸, genetic counters⁹, synchronized oscillators¹⁰, band-pass filters¹¹, pulse generator¹², edge detector¹³, and various genetic logic gates such as AND, OR, and XNOR^{14,15}. Similar to electronic circuits, logic functions can also operate in living cells producing certain outputs based on environmental and cellular inputs. For example, the *lac* operon in *Escherichia coli* works as an AND gate, where β -galactosidase is only produced in the presence of both lactose and cyclic adenosine monophosphate¹⁶. This presents living cells as a conducive environment for the operation of synthesized genetic circuits designed with synthetic biology tools for conducting different applications.

Moreover, many engineered riboswitches are developed to regulate gene expression at the translational level. These RNA-based genetic switches can be easily modulated to target gene networks and reprogram cellular behaviors. For example, Green and colleagues recently engineered “Toehold switches” using non-coding RNAs to construct multi-input AND logic evaluator and regulate endogenous gene expression¹⁷.

Post-translational circuits generally rely on creating novel chimeric proteins, such as protein receptors, to rewire natural signaling pathways and reshape the dynamics of signaling transduction^{4,18}. Also, it is necessary to note that some hybrid synthetic circuits are engineered by integrating RNA devices into transcriptional regulatory networks¹⁹, and some synthetic gene networks are constructed and expressed *in vitro*²⁰.

1.2 Applications of Synthetic Gene Networks

1.2.1 Build to Understand: Multistability

Whether in bacteria or in human, gene regulatory networks of thousands of genes and interactions play critical roles in fundamental biological processes such as metabolism, signaling transduction, cell differentiation and development²¹. However, it is almost impossible for traditional biologists to systematically investigate the large-scale interconnected networks in their natural context using conventional tools. Synthetic gene networks, on the other hand, provide a clear and isolated platform for uncovering the complex regulatory mechanisms, such as gene expression noise, multistability, cell decision-making, and cell-cell communications^{22,23}. These network topologies are usually abstracted from natural analogues, built with standardized genetic modules from scratch, and investigated in living cells. Mathematical modeling as a complementary tool is also employed to quantitatively understand the network dynamics and guide experimental design.

Multistability is a mechanism that cells use to achieve a discrete number of mutually exclusive states in response to environmental inputs, which enables cells to make the appropriate decisions at each condition. In prokaryotic cells, multistable

switches are common for cellular decision-making, such as the lysis/lysogeny switch of phage lambda^{3,24} and sporulation/competence in *Bacillus subtilis*^{25,26}. In multicellular organisms, multistable switches also play essential roles in the regulation of cell-cycle oscillator during cell mitosis²⁷, Epithelial-to-Mesenchymal transition and cancer metastasis^{28,29}, and the well-known cell differentiation process, which is a manifestation of cellular state determination in a multistable system^{30,31}.

Bistability, the simplest case for multistability, enables cells to switch back and forth between two stable steady states, and has been largely investigated in natural and synthetic systems. In general, bistability can arise from mutually inhibitory networks or positive-feedback loops³². A myriad of modeling and experimental studies have been performed to investigate the nonlinear dynamics and state transitions of bistable systems^{6,23,27}. As a milestone achievement in synthetic biology, the toggle switch⁶ in 2000 paved the road to develop synthetic biology with their groundbreaking work on the “toggle switch” circuit, in which two genes inhibited each other through targeting corresponding promoters. By using external inducers, the circuit could be controlled to toggle between two stable steady states (bistable) predicted by their mathematical model.

This proof-of-concept design has also been implemented in mammalian cells and yeast. For example, the Fussenegger group³³ engineered a mammalian epigenetic transgene switch in which E-KRAB and PIP-KRAB expressed on two individual plasmids could inhibit each other by binding to ETR and PIR operators in their respective promoters. The induction of antibiotics indicated that this epigenetic circuitry exhibited two stable expression states: high E-KRAB with low PIP-KRAB, and low E-KRAB with high PIP-KRAB. Furthermore, the bistable expression profiles were fully reversible even

after rounds of expression switching. Interestingly, the system also showed long-term bistability in mice, suggesting that synthetic gene networks could be used as therapeutic devices in clinic in the future.

Positive feedback is another ubiquitous topological structure found in nature that has the capacity to generate bistability. For example, the key regulatory mechanism for bacteria quorum-sensing systems is positive feedback motifs, which enable cells to make binary decisions in responding to environmental signals³⁴. In eukaryotes, positive feedback loops embedded in gene networks regulate stem cell differentiation and development. For instance, positive feedback between the transcription factor *PU.1* and the cell cycle controls lymphoid and myeloid differentiation³⁵; positive feedback between *Sox2* and *Sox6* in neural progenitor cells represses neuronal differentiation³⁶; and *Cdkn1c* interacts with *Myod* to form a positive feedback that drives muscle differentiation³⁷.

Robust bistable responses can also be achieved through coupling multiple positive feedback loops. Guided by theoretical model-based calculations, Chang *et al.*³⁸ tried to identify parameters controlling the size of a bistable range in order to build ultrasensitive systems. The authors created a composite system with two coherent positive feedback loops, where promoter *glnK* drove expression of *glnG* and the *lacZYA* operon, respectively. The expression of *glnG* was simultaneously inhibited by the LacI protein and activated by the inducer IPTG. Functionally, *glnG* could auto-activate *glnK* transcription, forming a positive feedback, while the LacY gene product galactoside permease could facilitate cellular uptake of IPTG to promote *glnG* expression by deactivating LacI repression, forming another positive feedback. Experimental results showed that the double-positive feedback circuit exhibited potent bistability over a ~480-

fold range of induction concentrations, whereas circuits with a single positive feedback showed a less than ~12-fold range.

The current experimental researches, however, mostly focus on bistable switches, involving transitions between only two states. And demonstrations, from a combination of experiments and computational modeling, for the existence and operation of such a landscape in a higher dimensional multistable system are still lacking. Our recent results found that quorum sensing (QS) crosstalk can be engineered to yield trimodal responses resulting from noise-induced state transitions and circuit-host interactions³⁹ (Figure 1).

1.2.2 Build to Understand: Pattern Formation

Cell-cell communication exists in unicellular and multicellular species, and is responsible for coordinating collective population behaviors, such as biofilm formation⁴⁰, cell differentiation⁴¹ and vertebrate embryonic development⁴². Quorum sensing is a widespread cell-cell communication mechanism in the bacteria world⁴³, such as the LuxR/LuxI system in *Vibrio fischeri* and the LasR/LasI system in *Pseudomonas aeruginosa*. So far, quorum-sensing mechanisms, coupled with engineering principles, have been employed to program population control^{44,45}, build synchronized oscillations^{10,46}, produce diverse cell phenotypes^{47,48}, control biofilm signaling⁴⁰, produce pattern formation^{49,50}, and to construct synthetic ecosystems⁵¹.

In 2005, Basu *et al.* programmed a synthetic multicellular system to form bullseye, ellipse, heart, and clover patterns based on local AHL gradients around sender cells⁵². Morphogen diffusion on the solid plate established a natural gradient that could be sensed by receiver cells and induced differential responses at distinct regions. By combining

mathematical modeling and experimental evidence, this study provided us with a better understanding of the multi-scale mechanisms underlying pattern formation in development.

In a recent study from the *You* group, a novel pattern forming mechanism was developed using a synthetic gene circuit in *E. coli*⁵³. This circuit is composed of an auto-activating motif driven by the activator T7 RNAP and negative regulation of the activator induced by AHL from a positive feedback module. Bacteria harbouring this circuit generated a self-organized ring pattern. This research sheds light on a novel morphogen timing mechanism to generate spatial patterns in developmental processes. Further studies also indicate that the synthetic circuit could generate robust scale invariance in bacterial colony growth by controlling the temporal dynamics of diffusible morphogens⁵⁴.

To date, diverse mechanisms have been proposed about how cells produce spatial patterns, either dependent on or independent of morphogen gradients. Liu *et al.*⁵⁰ recently used synthetic approaches to reveal another mechanism that could generate spatial patterns by coupling cell motility with density. The authors engineered a synthetic gene circuit having two modules: the constitutively expressed LuxR/LuxI system to monitor local cell density (density-sensing module), and a module with LuxR/LuxI regulating *cheZ* expression to control cell motility (motility-control module). At high cell densities, the LuxR-AHL complex will inhibit *cheZ* transcription, resulting in a loss of motility. However, *cheZ* reintroduction made cells regain motility at low cell densities. When placed in the middle of semi-solid agar plates, the engineered *E. coli* automatically and sequentially developed periodic stripes with alternative high and low cell densities as cells moved radially outwards. This study reveals that a recurrent spatial structure can be

generated by tuning the cell's motility and density, which may provide novel insights for developmental systems.

1.2.3 Synthetic gene networks in single-cell organisms: Biosensor, Biofuel, and Biopharmaceutical

Started with engineering of simple gene circuits in *E. coli* cells, today synthetic biology has upgraded into complex higher-order gene networks and systems with predictable functions in mammalian cells and real-world applications for human health^{1,55}. Application-oriented synthetic gene networks in single-cell organisms for solving realistic problems are also underway. Biosensor is one of the major applications. Generally, biosensor is composed of two genetic parts, a sensing element responsible for recognizing and interacting with the analyte of interest, and a detector element to transform the interaction signal and report it in a user-friendly way⁴. For example, Wang *et al.* designed a set of single-input cellular biosensors which contain an environmental-responsible promoter and corresponding transcriptional factors and a fluorescent output to detect the concentration of copper, mercury, zinc, and cadmium in an aqueous environment⁵⁶. Moreover, Kotula *et al.* engineered living bacteria with synthetic genetic circuits comprising a trigger element and a memory element to sense, record and report environmental stimuli in the mammalian gut, which lays the foundation for developing living diagnostics and therapeutics for human⁵⁷. Recently, Pardee *et al.* integrated synthetic gene networks with paper-based technology to develop low-cost and practical paper-based biosensors to monitor glucose and strain-specific Ebola virus⁵⁸ (Figure 1).

In addition, engineering of biosynthetic pathways for cost-effective and high-yield microbial production of interests (such as biofuels) is another important application of synthetic gene networks. Metabolic pathways of interests are often found in natural organisms but the natural yield is usually very low owing to host's tight regulation. So it is promising to use synthetic biology approach to improve host's production capacity either by optimizing the metabolic flux or by transplanting the whole biosynthetic pathway into industrial model microbes, such as *E. coli* and *Saccharomyces cerevisiae*^{4,59}. For example, Atsumi and colleagues transferred the butanol-producing pathway from *Clostridium acetobutylicum* into *E. coli* and genetically modified a number of *E. coli* genes to improve 1-butanol production⁶⁰. Also, *S. cerevisiae* was used as a cell factory to overproduce fatty acid ethyl esters (biodiesel) by eliminating non-essential fatty acid utilization pathways⁶¹. So far, bioproduction of many biofuels including ethanol, butanol, biodiesel, 2-propanol, and hydrogen have been successfully explored and scaled-up through engineering of biosynthetic pathways^{62,63}.

Furthermore, biosynthetic pathway engineering is applied to produce rare pharmaceutical drugs. One of the successful examples is production of the antimalarial drug artemisinin, which is a sesquiterpene endoperoxide originally produced by the plant *Artemisia annua*. In 2006, Keasling group programmed *S. cerevisiae* to produce high titres of artemisinin precursor artemisinic acid by engineering of farnesyl pyrophosphate biosynthetic pathway and introducing amorphadiene synthase and a novel cytochrome P450 monooxygenase from *A. annua*⁶⁴ (Figure 1). Recently, they further demonstrated the complete biosynthetic pathway for high-level production of artemisinic acid and developed an efficient bioprocess for conversion of artemisinic acid to artemisinin⁶⁵.

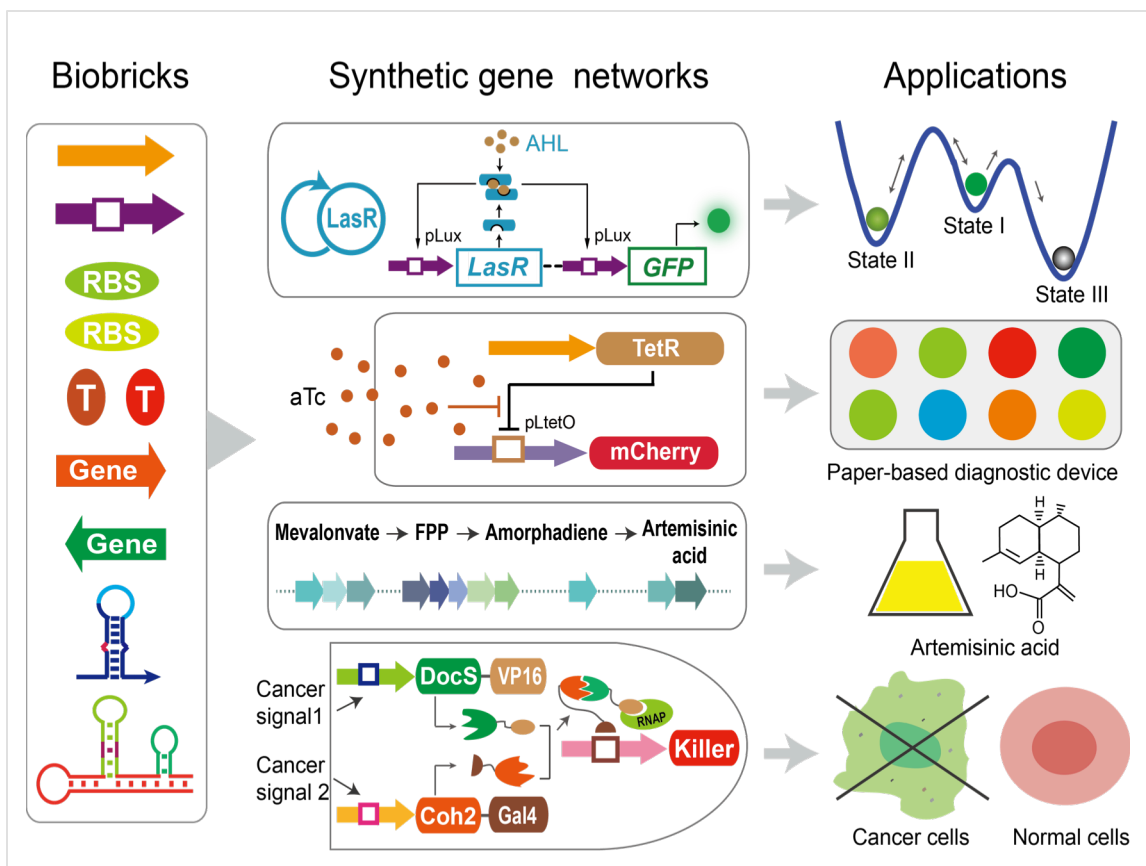


Figure 1: Diagram of assembly and applications of synthetic gene networks⁶⁶. From left to right: standardized biological blocks (biobricks) are employed to assemble synthetic gene networks for different applications. From up to bottom: Build synthetic gene network to understand biological mechanism; genetic circuits for paper-based biosensor; biosynthetic pathway for pharmaceutical drug; genetic circuit for anti-cancer therapy.

1.2.4 Therapy-oriented synthetic gene networks in mammalian cells

The synthetic gene network is moving into clinical therapies⁶⁷. Engineering synthetic gene networks as an alternative to the conventional pharmacotherapy to combat human diseases is promising. The basic idea is using synthetic gene circuits as a smart sensor-actuator device to functionally interface with the host's complex endogenous networks to monitor disease signals *in vivo* and make therapeutic responses automatically. Here, we briefly introduce applications of synthetic gene networks to treat cancer, diabetes, obesity, and immunotherapy.

One of the challenges for cancer therapy is to separate cancer cells from surrounding normal cells. So the principle is to design devices, which can specifically sense cancer signals such as overexpressed oncogenes and then initiate the killing process. For example, Nissim and colleagues constructed a dual-promoter integrator to target and kill cancer cells based on their cancerous context (Figure 1). Each of the two cancer cell-specific promoters drive a chimeric protein expression, which are then combined together to activate the thymidine kinase expression to kill the cell. So this AND logic circuit will be only activated in host cells expressing the two signals simultaneously⁶⁸. A similar strategy was also used in a recent study in which the authors engineered a multi-input RNAi-based logic circuit to specifically identify cancer cells⁶⁹. This cancer cell classifier functions through detecting the expression levels of sets of endogenous microRNAs in HeLa cells and triggering cell apoptosis when the expression levels match a predefined profile.

Last but not the least, synthetic gene circuits are also applied for immunotherapy. For example, Chen *et al.* engineered a synthetic ribozyme switches to regulate gene expression and T-cell proliferation *in vitro* and *in vivo*⁷⁰. Wendell Lim's group developed synthetic feedback modulators and pause switches using bacterial virulence proteins to precisely tune T-cell response amplitude and control human primary CD4+ T-cell activation⁷¹.

CHAPTER 2

QUORUM-SENSING CROSSTALK DRIVEN SYNTHETIC CIRCUITS: FROM UNIMODALITY TO TRIMODALITY

2.1 Introduction

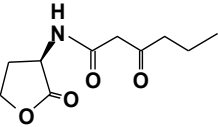
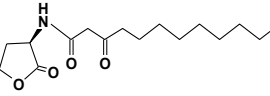
Quorum-sensing (QS) is a widespread mechanism bacteria use to regulate gene expression and coordinate population behavior based on local cell density⁷². It is achieved through the binding of QS regulators with their cognate signal molecules (autoinducers) to regulate downstream QS pathways. Autoinducers are produced inside the cell and diffuse into and out of bacterial cells. Therefore, an autoinducer's intracellular concentration correlates with local cell density⁷². There are diverse QS mechanisms allowing for bacterial communication: gram-positive bacteria generally use two-component systems mediated by peptides while gram-negative bacteria primarily use LuxR/I-type systems mediated by acylated homoserine lactones (AHL)^{73,74}. Many bacterial activities are regulated by QS, such as antibiotic production, biofilm development, bioluminescence, colonization, sporulation, symbiosis, and virulence⁷²⁻⁷⁶.

With well-defined and characterized biological properties, several QS regulators and corresponding autoinducers have also been used for synthetic gene networks. For example, LuxR/LuxI and/or LasR/LasI pairs were used to generate programmed patterns^{52,53}, trigger biofilm formation^{77,78}, develop synthetic ecosystems and program population dynamics^{40,51}, and construct synchronized oscillators^{10,46}, edge detectors¹³,

and pulse generators¹². RhlR/RhII has also been used in the study of generic mechanisms of natural selection⁷⁹ and for carrying out biological computations as chemical ‘wires’⁸⁰.

However, effects of QS crosstalk, functional interactions between QS components that are not naturally paired, remain unexplored. For example, widely used LuxR-family regulators share extensive homologies and structural similarities in their corresponding autoinducers. LuxR and LasR proteins bind their respective natural ligands, homoserine lactones (HSL) 3-oxo-C6-HSL (3OC6HSL, hereafter denoted as C6) and 3-oxo-C12-HSL (3OC12HSL, hereafter denoted as C12) to activate promoter *pLux* and *pLas*, respectively (Table 1)^{72,81,82}. However, the LuxR protein can also bind other HSLs, such as C7HSL and 3OC8HSL⁸³. LasR can also bind C12 to activate the *pLux* promoter in addition to its natural pairing *pLas* promoter⁵¹. Implications of such crosstalk on gene regulation and cell response remain largely unknown.

Table 1: Autoinducer information.

Full name	Abbreviation	Molecular structure	Coding gene	Original regulator and promoter	Organism
N-(3-Oxohexanoyl)-L-homoserine lactone	3OC6HSL		LuxI (coding the enzyme, which synthesizes 3OC6HSL)	LuxR, <i>pLux</i>	Aliivibrio fischeri
N-(3-Oxododecanoyl)-L-homoserine lactone	3OC12HSL		LasI (coding the enzyme, which synthesizes 3OC12HSL)	LasR, <i>pLas</i>	Pseudomonas aeruginosa

Here, we use rationally designed gene networks to probe crosstalk between the LuxR/I and LasR/I systems and investigate their elicited bistable behaviors from positive feedback topologies. By using synthetic biology approaches, all combinations of

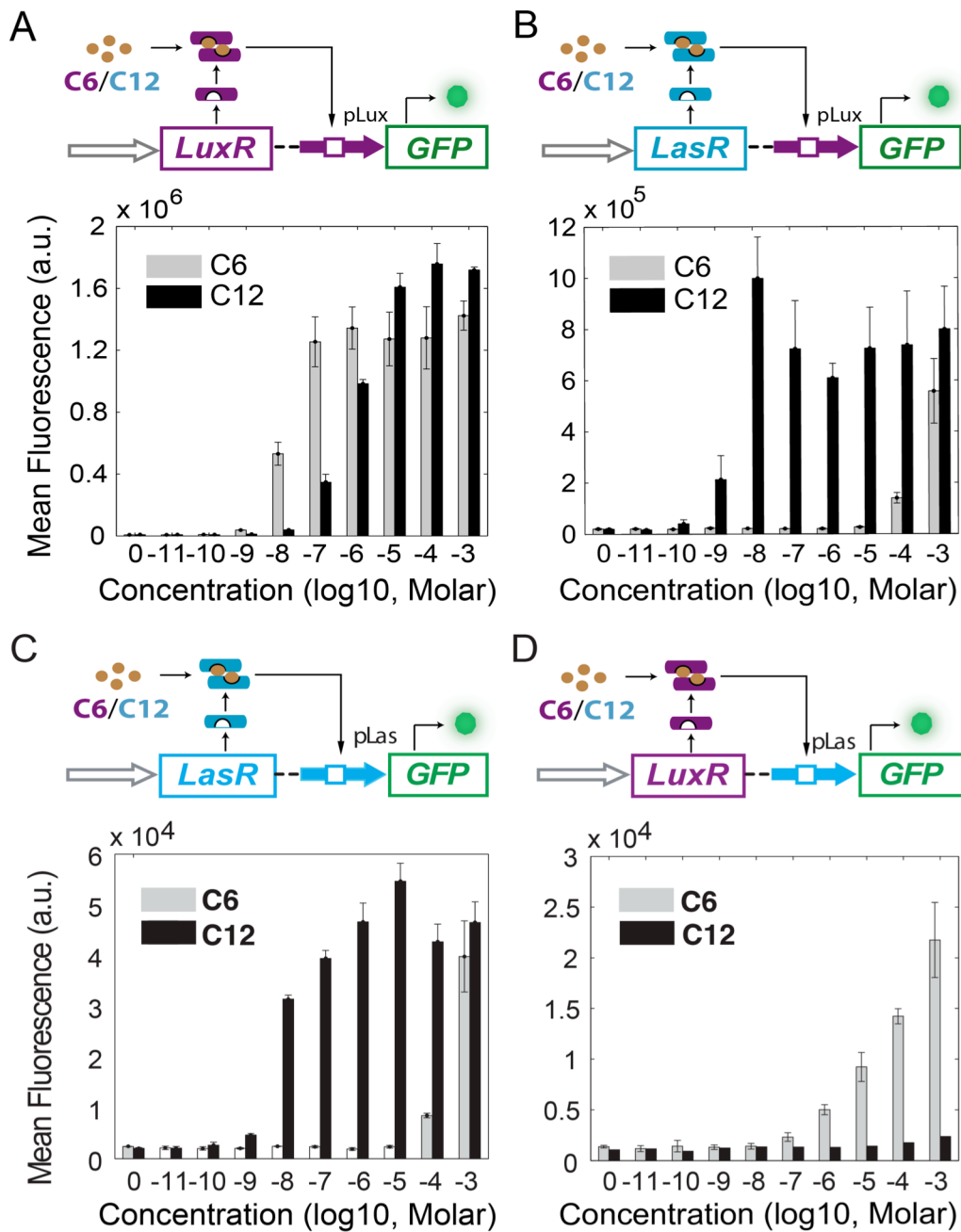
autoinducer, regulator gene, and promoter were tested to show that QS crosstalk can be dissected into signal crosstalk and promoter crosstalk. When studied in the context of a synthetic positive feedback gene network, our results indicate that QS crosstalk leads to distinct dynamic behaviors: signal crosstalk significantly decreases the circuit's induction range for bistability, but promoter crosstalk causes transposon insertions into the regulator gene and yields trimodal responses due to a combination of mutagenesis and noise induced state transitions. To fully understand this complex response, we developed a mathematical model that takes into account all of these factors to simulate and predict how varying the transposition rate can modulate this trimodality, which was verified experimentally. This reveals a novel factor of host-circuit interactions in shaping complex responses of synthetic gene networks.

2.2 Results

2.2.1 Dissecting the crosstalk between LuxR/I and LasR/I using synthetic circuits.

To characterize possible crosstalk between LuxR/I and LasR/I signaling systems, four synthetic circuits, CP (constitutive promoter)-LuxR-*pLux* (Figure 2A), CP-LasR-*pLux* (Figure 2B), CP-LasR-*pLas* (Figure 2C), and CP-LuxR-*pLas* (Figure 2D), were first built to test all autoinducer-regulator-promoter combinations' impact on gene expression activation. C6 and C12 were applied respectively to all constructs, and green fluorescent protein (GFP) expression under regulation of *pLux* or *pLas* was measured as the readout.

It can be seen in Figure 2A that in addition to its natural partner C6, LuxR can also bind with C12 molecules to activate *pLux*, which suggests that the binding with C6 or C12 results in a similar conformational change of LuxR and therefore its activating



C

Summarization of the crosstalk between LuxR/I and LasR/I signal systems.

	LuxR- 3OC6HSL	LuxR- 3OC12HSL	LasR- 3OC6HSL	LasR- 3OC12HSL	LuxR- LuxI	LuxR- LasI	LasR- LuxI	LasR- LasI
pLux	+	+	-*	+	+	+	-	+
pLas	-	-	-*	+	-	-	-	+

1: Black "+" indicates the original pairs; red "+" signifies pairs showing crosstalk.

2: "*" indicates LasR could only activate the promoter at high 3OC6HSL concentration.

Figure 2: QS crosstalk dissected using synthetic gene circuits. (A) LuxR can cross-talk with C12 to activate *pLux*. Top panel: schematic diagram of a synthetic gene circuit where a constitutive promoter (gray arrow) regulates LuxR (purple rectangle) expression. LuxR protein, when bound with C6 or C12 and dimerization, can activate *pLux* (purple arrow) to induce GFP (green rectangle) expression. The autoinducers, genes, and promoters are color coded so that naturally paired partners are in the same color. Bottom panel: dose response of the circuit when induced with C6 (gray) or C12 (black). (B) LasR can cross-talk with *pLux* when bound with C12. Top panel: schematic diagram of a circuit similar to that in (A), where a constitutive promoter regulates LasR (cyan rectangle) expression. LasR protein, when bound with C6 or C12, can activate *pLux* to induce GFP expression. Bottom panel: Dose response of this circuit when induced with C6 (gray) or C12 (black). Bar heights are averages of three independent flow cytometry measurements. (C) High concentrations of C6 can crosstalk with LasR-*pLas*. Top panel: schematic diagram of the synthetic gene circuit (CP-LasR-*pLas*). Bottom panel: dose response of the circuit when induced with C6 or C12. Compared to original pair of LasR-C12, *pLas* promoter can only be activated by LasR with extremely high C6 concentration (signal crosstalk). (D) Promoter crosstalk of C6-LuxR to *pLas* is observed under high concentrations of autoinducer. Top panel: schematic diagram of the circuit (CP-LuxR-*pLas*). Bottom panel: Dose response of this circuit when induced with C6 or C12. LuxR can bind with C6 to activate *pLas* starting from 10^{-6} M doses (promoter crosstalk), while it cannot with C12. (E) Summary of crosstalk induction of all 16 different combinations, including inductions by both chemicals and corresponding synthase genes. The four combinations shown in (A) and (B) are highlighted with a gray background.

functions remain uninterrupted. Such an activation of a natural QS regulator-promoter pair by a cross-talking autoinducer is here termed signal crosstalk. It can be seen that this signal crosstalk can fully activate the system with comparable induction dosages. However, similar tests of signal crosstalk of C6 with the Las regulator-promoter pair (Figure 2C) only show comparable induction when the autoinducer concentration is as high as 10^{-3} M. This suggests that the efficacy of signal crosstalk is QS system specific.

In addition to promiscuous autoinducer binding caused signal crosstalk, the systems studied also displayed crosstalk between regulators and promoters, here termed promoter crosstalk. It is shown in Figure 2B that, in addition to being able to activate *pLas*, LasR significantly activate *pLux* when induced with its natural cognate ligand C12,

though not with C6, which suggests that LasR's DNA binding domain can recognize both *pLas* and *pLux* when bound with its natural partner. This promoter crosstalk is robust over a wide range of autoinducer concentrations. Similar tests of promoter crosstalk of C6-LuxR to *pLas* (Figure 2D) show only weak induction. This suggests that the efficacy of promoter crosstalk is also QS system specific. It should also be noted that a third type of crosstalk, regulator crosstalk, in which naturally paired autoinducer and promoter function through a cross-talking regulator protein, only exhibited minimal levels of activation (gray bar in Figure 2B and black bar in Figure 2D).

To further verify the crosstalk under physiologically relevant dosages of autoinducers, synthase genes *LuxI* and *LasI* were introduced to replace commercial chemicals in eight different circuits (Figure 3). The results further confirm that *pLux* could be activated by LuxR with *LuxI* or *LasI*, as well as LasR with *LasI*. This is consistent with the results above with commercial chemicals, indicating the crosstalk categorization is also applicable *in vivo*. All combinatorial activations between LuxR/I and LasR/I systems are summarized in Figure 2E, with crosstalk highlighted in red. Taken together, detectable crosstalk between LuxR/I and LasR/I systems can be categorized into two types: *LasI* (C12) can crosstalk with the LuxR protein to induce *pLux* transcription (signal crosstalk), and the LasR-*LasI* (C12) complex can also crosstalk with and activate the *pLux* promoter (promoter crosstalk).

2.2.2 Signal crosstalk induces distinct responses from positive feedback circuits.

Next, synthetic positive feedback circuits were constructed to investigate the impact of QS crosstalk in the context of gene regulatory networks. It is shown that the

core of many bacteria's QS decision-making circuits is a positive feedback motif⁸⁴⁻⁸⁹. Because of its potential bistability, such a topology enables the bacteria to make appropriate binary decisions in response to changing environments^{32,90}. Synthetic positive feedback circuits serve as suitable platforms to probe the effects of signal and promoter crosstalk within the framework of gene regulatory networks.

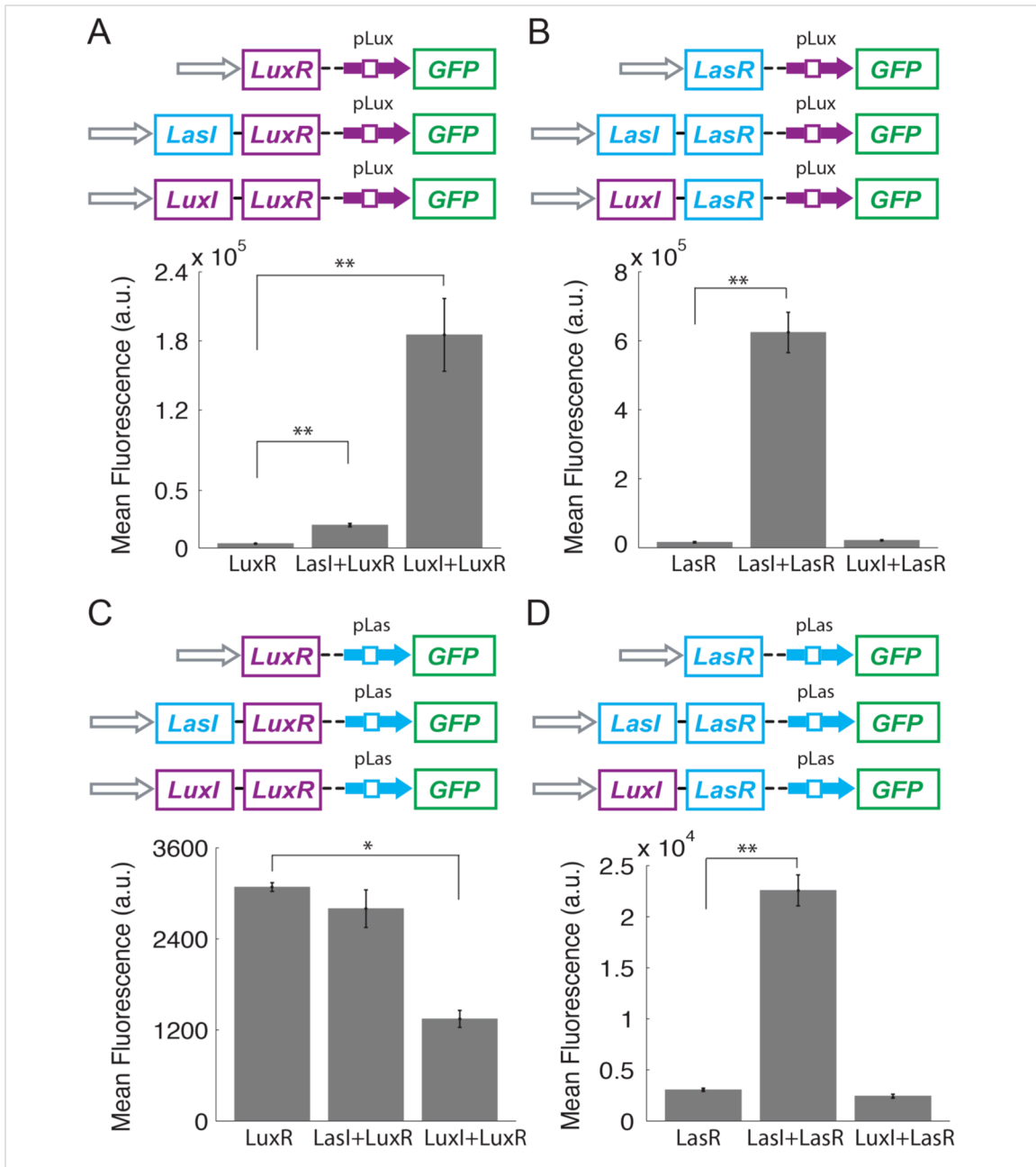


Figure 3: Characterizing the crosstalk to the *pLux* and *pLas* promoters using synthase genes. (A-B) LuxR, either with LuxI or with LasI, can activate *pLux*, while LasR with LasI can activate *pLux*. Left: schematic diagram of the synthetic gene circuits constructed to test crosstalk. LasI (cyan) and LuxI (purple) synthesize 3OC12HSL and 3OC6HSL molecules in cells, respectively. Right: GFP fluorescence in cells carrying the circuits was measured by flow cytometry at 12 hr. LasI with LuxR, and LasR with LasI can significantly activate *pLux* (signal crosstalk, and promoter crosstalk, respectively). (C-D) Characterizing the crosstalk to the *pLas* promoter using synthase genes. No significant crosstalk was observed for LuxR- or LasR-*pLas* combinations. Left: schematic diagram of the synthetic gene circuits constructed. Right: GFP fluorescence in cells carrying the circuits was measured at 12 hr. Both LasI-LuxR and LuxI-LuxR cannot activate *pLas*, and the latter shows ~ two-fold inhibition, and no signal crosstalk is observed for LasR-*pLas*. All the data were averages of three independent measurements shown as mean \pm SD (* $p < 0.05$, and ** $p < 0.01$).

The design shown in Figure 4A was first constructed to study signal crosstalk. In this circuit, expression of LuxR is regulated by the promoter *pLux*, which can be activated by LuxR when induced, forming a positive feedback loop. *pLux* driven GFP expression serves as the readout for LuxR levels. Robustness of history-dependent responses (hysteresis), a hallmark of many positive feedback topologies, is used as the main measure of signal crosstalk impacts as it captures the effectiveness of the circuit's decision-making functionality^{6,23,91}.

As a benchmark, uninduced (Initial OFF) cells with the circuit were first induced with different concentrations of LuxR's natural inducer C6 and measured using flow cytometry (Figure 4B, blue). It can be seen that GFP is only turned on with 10^{-8} M or higher C6 induction. The cells treated with 10^{-4} M C6 (Initial ON) were then collected and diluted into new medium with the same concentrations of C6 (Figure 4B, red). These cells keep high GFP expression even with low C6 inductions (below 10^{-9} M) due to the self-sustaining nature of positive feedback loops. Taken together, these results illustrate this circuit's hysteretic response with C6 inducer concentrations between 0 and 10^{-8} M.

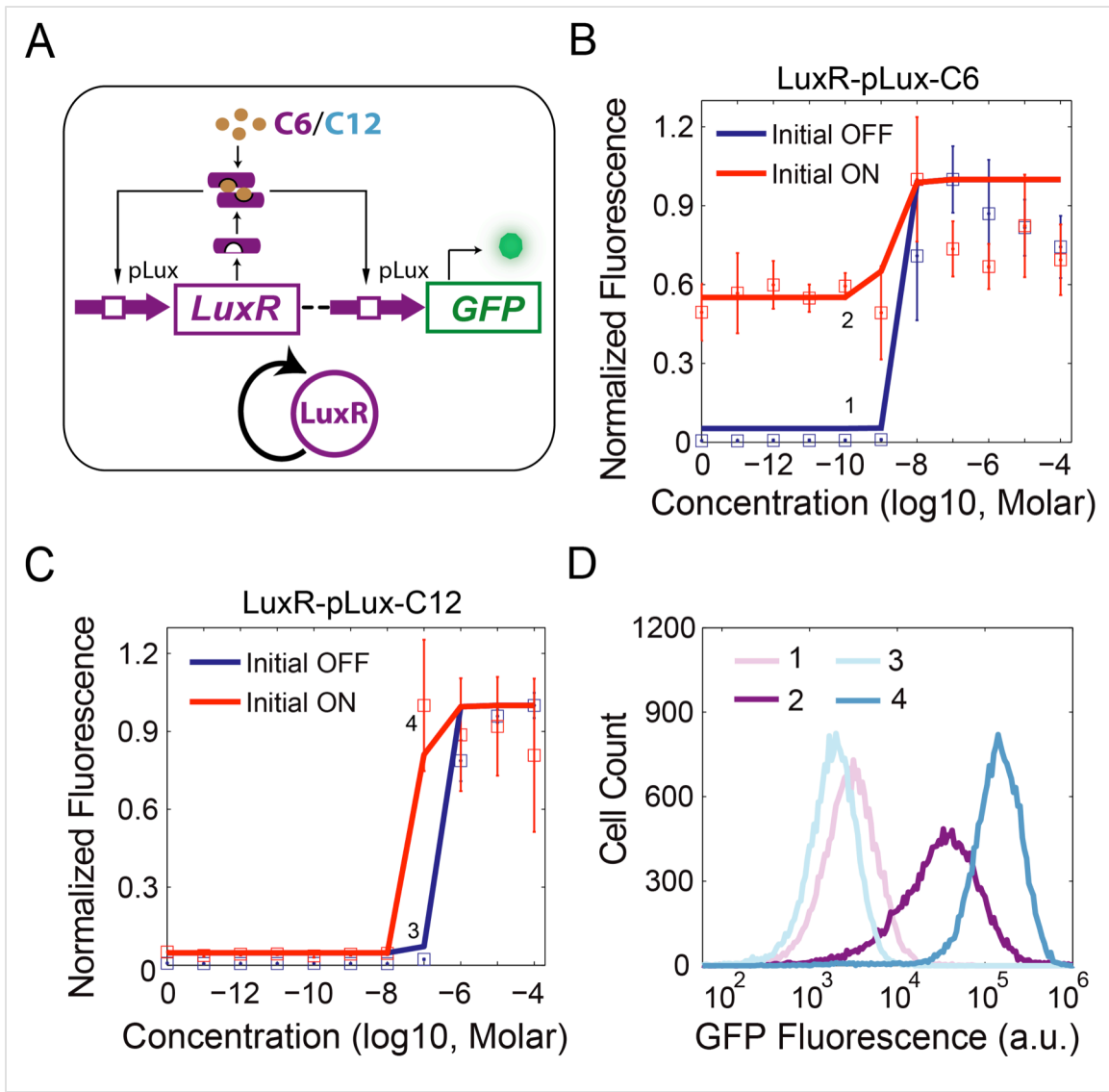


Figure 4: Signal crosstalk causes shrinkage of bistable region. (A) Schematic diagram of a synthetic gene circuit where the *pLux* promoter regulates expression of *LuxR*, which in turn can bind with C6 or C12 to further activate *pLux*, forming a positive feedback loop (shown as simplified diagram). GFP under the regulation of *pLux* serves as the readout for *LuxR* levels. (B) The average of three replicate flow cytometry measurements is plotted as a square with error bar for each dose of C6 induction, where red indicates Initial ON cells while blue denotes Initial OFF cells. Solid lines represent results calculated from model fittings. The bistable region ranges from 0 to 10⁻⁹ M C6. Labels 1 and 2 indicate representative experiments within the region to be shown as histograms in (D). (C) Similar experiments as in (B) but with C12 inductions. The bistable region ranges from 10⁻⁸ to 10⁻⁶ M C12. Labels 3 and 4 indicate representative experiments within the bistable region to be shown as histograms in (D). (D) Histograms of flow cytometry measurements labeled in (B) and (C). One representative measurement from each point is shown. No bimodal distributions are observed.

This indicates that under C6 induction the positive feedback circuit is bistable between 0 and 10^{-8} M C6 induction. However, no bimodal distribution was observed within the bistable region based on flow cytometry measurements (Figure 4D, purple and light purple), suggesting that the barrier between the two states is too high for inherent gene expression stochasticity to overcome^{6,91}.

Next, C12 was used to induce the same construct to investigate the impact of signal crosstalk on gene network regulation. Similar induction experiments were carried out and the results are shown in Figure 4C. It can be seen that this circuit also displays hysteresis, but with a much smaller bistable region between 10^{-8} and 10^{-6} M C12. Flow cytometry results within the bistable region also show no bimodal distributions (Figure 4D, cyan and light cyan).

To quantitatively understand the signal crosstalk caused shrinkage of the bistable region, an ordinary differential equation (ODE) model of LuxR-*pLux* auto-activation was developed (Details can be found in Mathematical Modeling). Two major kinetic events, LuxR transcription and translation, are described by two ODEs with all binding between chemical species incorporated into model terms. After fitting the parameters using existing literature and experimental measurements, the model can capture the experimental results (solid lines in Figure 4B and 4C) with accuracy. Inspection of model parameters reveals that the bistable region decrease caused by signal crosstalk can be largely accounted for by differential binding affinities between LuxR and C6 and C12. This suggests a new way to perturb QS decision-making through utilization of crosstalking autoinducers, which is useful for clinical therapies.

2.2.3 Promoter crosstalk induces unexpected and complex bimodal responses.

To study the impacts of promoter crosstalk, a positive feedback circuit was constructed with LasR under the regulation of *pLux* (Figure 5A). It is shown in Figure 2B that LasR can activate *pLux* when induced by C12. Therefore this circuit also forms a positive feedback loop in the presence of C12. GFP under regulation of *pLux* is again included as a readout for LasR. Experimental explorations of hysteresis were carried out and the results are shown in Figure 5B. It can be seen that initial OFF cells (blue) exhibit a non-monotonic response to C12 induction: GFP expression increases with C12 concentration, but begins to uniformly decrease when C12 induction exceeds 10^{-8} M (Figure 5B, and Figure 6A-B). Cells induced with 10^{-4} M C12 were then collected and diluted into fresh medium with the same inducer concentrations as the initial OFF cells. Flow cytometry data show that all samples exhibit unimodal minimal fluorescence signals that are even lower than the basal GFP expression of initial OFF cells (Figure 5B and 5C green, and Figure 6B).

Considering that both C12 and exogenous gene overexpression may be toxic to cells, as well as the fact that initial OFF cells can be turned on with lower induction dosages, cells induced with lower than 10^{-4} M but higher than 10^{-10} M C12 were collected as new initial ON cells to further explore possible hysteresis of this circuit. Collected cells were diluted into fresh medium with the same concentrations of C12. These new initial ON cells demonstrate the same expression pattern as the initial OFF cells when grown in inducer concentrations from 0 to 10^{-9} M, but they show much lower fluorescence values at higher concentrations. For example, the red points in Figure 5B

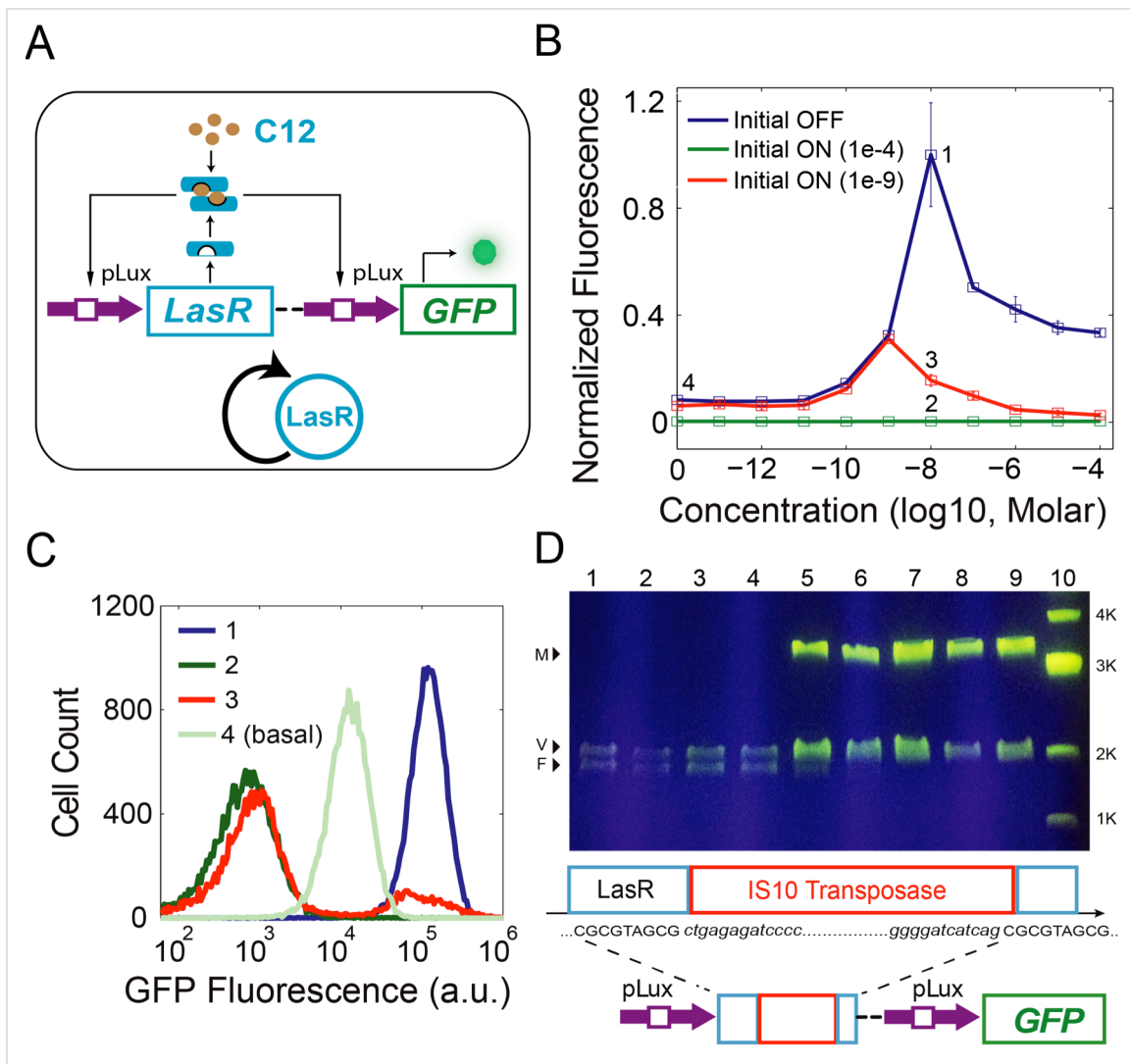


Figure 5: Promoter crosstalk induces mutation and leads to population heterogeneity. (A) Schematic diagram of a synthetic *LasR-pLux* positive feedback circuit. GFP under the regulation of *pLux* serves as the readout for LuxR levels. (B) The average of three replicate flow cytometry measurements is plotted as a square with error bars for each dose of C12 induction. Blue denotes Initial OFF cells, while green and red indicate the Initial ON cells induced with 10^{-4} M C12 and 10^{-9} M C12 before being re-diluted into concentrations of C12, respectively. Labels 1, 2, 3, and 4 indicate experiments to be shown in detail as histograms in (C). (C) Histograms of flow cytometry measurements labeled in (B). One representative measurement from each point is shown. A bimodal distribution is only observed for label 3: which is Initial ON cells (induced with 10^{-9} M C12 before redilution) at 10^{-8} M C12. (D) DNA analysis for the Initial ON samples shown as red in (B). Top: Plasmid DNA was extracted and digested with *EcoRI* and *PstI*, and agarose gel electrophoresis results indicated gene mutation happened in samples with 10^{-8} M and higher doses of C12. Lane 1 is the wild-type plasmid as the control, lanes 2 to 9 are samples in 10^{-11} to 10^{-4} M C12, and Lane 10 is the 1kb DNA marker. V: vector; F: wild-type DNA fragment (the *LasR-pLux* positive feedback circuit); M:

mutated fragment. Bottom: Schematic representation of the mutation and the features of IS10 transposase insertion: the target site (first *CGCGTAGCG*) in the LasR gene, its duplication (second *CGCGTAGCG*) due to insertion of *IS10* transposase, and the *IS10* sequence (red box and shown in italics).

illustrate the GFP average of 10^{-9} M induced initial ON cells when collected and re-diluted into a range of C12 concentrations (See Figure 6C for results with other initial induction dosages). Examination of the flow cytometry measurements of these ON cells reveals that bimodal distributions emerge within the concentration range of 10^{-8} M to 10^{-4} M C12. Interestingly, one peak of the distribution is at the high state and the other is at the minimal expression state, even lower than basal expression (Figure 5C, red). So unlike classic bimodal responses due to bistability, LasR-*pLux* positive feedback exhibits bimodality with the lower peak's expression even weaker than the OFF state. To exclude the possibility that this bimodality is triggered by inherent properties of the LasR-C12 complex, similar hysteresis experiments were carried out for the linear CP-LasR-*pLux* circuit (Figure 2B). Results show that the initial OFF and ON cells both exhibit unimodal expression without hysteresis (Figure 6D). The bimodality is, therefore, unique to the initial ON cells with LasR-*pLux* positive feedback.

2.2.4 Bimodality results from circuit-host interactions.

The remaining question is: what is the cause of the minimal expression state? To resolve this problem, new initial ON samples at concentrations of 10^{-11} M to 10^{-4} M C12 (Figure 5B) were collected. Their plasmids were extracted and digested for genotyping. The agarose gel electrophoresis results show that a new band (~3.2 kb) replaces the original fragment band (wild type, ~1.9 kb) for samples in 10^{-8} ~ 10^{-4} M C12, and that a

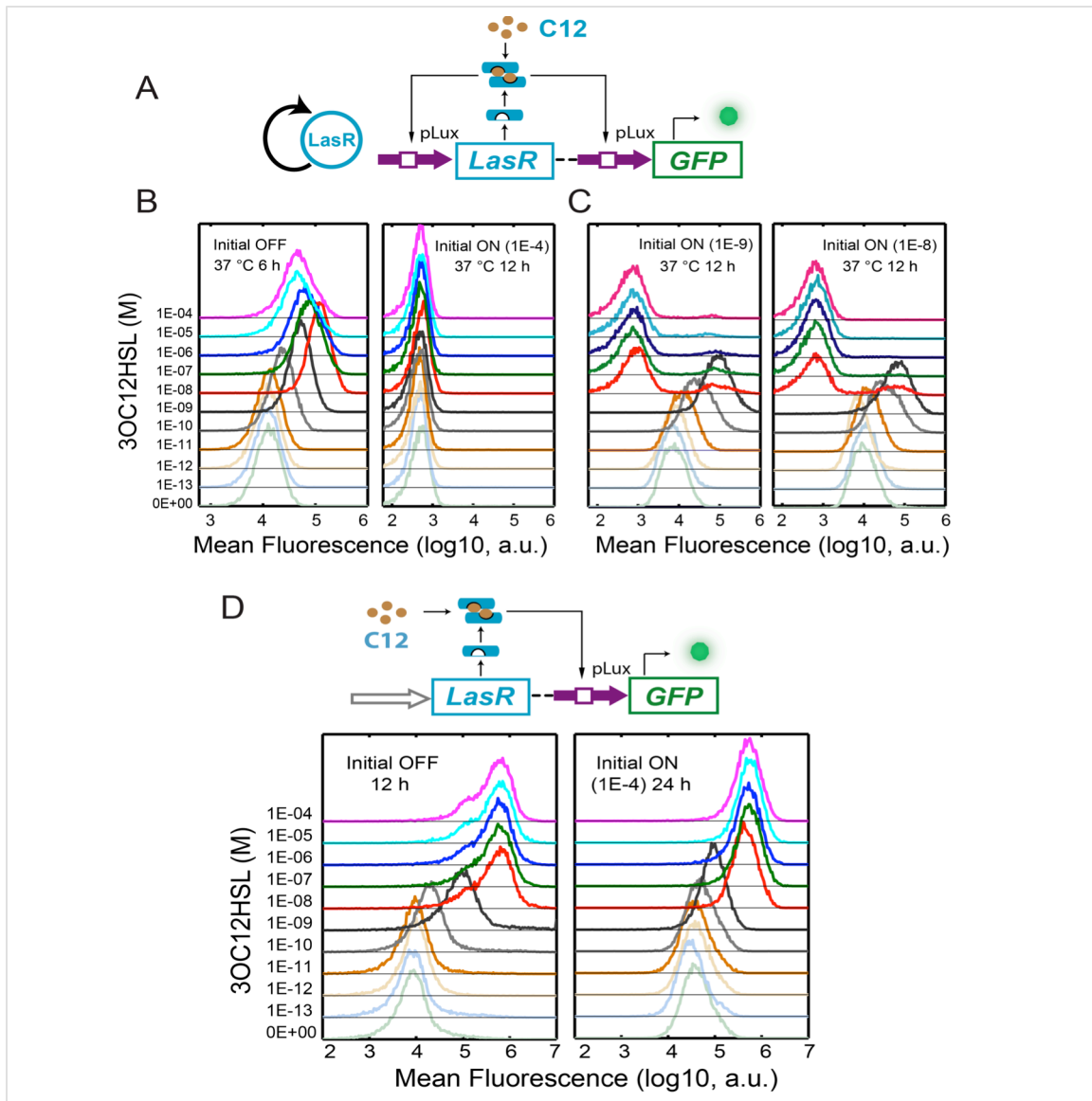


Figure 6: Hysteresis of the LasR-*pLux* circuits. (A) Schematic representation of the LasR-*pLux* positive feedback loop induced with C12. (B) Flow cytometry measurements of GFP expression for initial OFF cells (left) at 6 hr and initial ON cells (right) at 12 hr and 37°C under different concentrations of C12 induction. For initial OFF cells, GFP expression increases with C12 concentration, but begins to decrease uniformly when C12 induction exceeds 10⁻⁸ M. For initial ON cells (induced with 10⁻⁴ M C12 before redilution), all the samples exhibit unimodal minimal fluorescence signals that are even lower than the basal GFP expression of initial OFF cells. (C) Initial OFF cells were first induced with 10⁻⁹ or 10⁻⁸ M at 37°C for 6 hours to become the new Initial ON cells, which were then collected and rediluted into fresh media with different doses of C12. The two Initial ON cells show a similar GFP distribution pattern: unimodal distributions similar to the initial OFF cells for samples in the lower inducer concentrations of 0 to 10⁻⁹ M, and bimodal distributions within the higher concentration range of 10⁻⁸ to 10⁻⁴ M C12. (D) C12 induced hysteresis of the CP-LasR-*pLux* circuit. Flow cytometry measurements

of GFP expression for initial OFF cells (left) at 12 hr and initial ON cells (right, induced with 10^{-4} M C12 for 6 hours before redilution) at 24 hr and 37°C under C12 induction. Results show that the initial OFF and ON cells show a similar distribution pattern, and both exhibit unimodal expression without hysteresis.

faint original-fragment band can also be seen for samples with 10^{-8} and 10^{-7} M C12 inductions (Figure 5D). Further sequencing analyses verify that an *IS10* transposase is inserted into the LasR gene at the 682 bp site and this insertion is flanked by two 9 bp direct repeats *5'-CGCGTAGCG-3'* (Figure 5D), which is consistent with reported hotspots for *IS10* insertion⁹².

The insertion abolishes LasR's ability to activate downstream GFP expression, which in turn causes the cells' fluorescence signal to be even weaker than basal expression when LasR is intact. Cells with this type of mutation form the low GFP peak in the bimodal distributions in Figure 5C. On the other hand, cells that do not mutate are able to maintain a high GFP expression due to positive feedback, forming the GFP ON peak of the bimodal distributions. Taken together, the combination of gene network activated GFP expression and mutation caused GFP inhibition drive the emergence of a bimodal distribution.

2.2.5 Trimodality predicted by expanded model.

In light of the verified mutation in the LasR-*pLux* positive feedback system, the mathematical model was expanded to take into account crosstalk triggered genetic changes to better describe the circuit. To enable comparison with flow cytometry results, the ODEs were transformed into corresponding biochemical reactions and simulated stochastically⁹³. In addition, each cell was assigned a probability of mutation throughout

the simulation (Figure 7C inset), which is dependent on the cell's current LasR/GFP level and the transposition rate. Once mutated, the cells had only minimal GFP expression strength and remained mutated until the end of the simulation. Finally, growth rate differences between wild type and mutated cells were computed from experiments (Figure 8A) and taken into consideration in the simulation. Results of stochastic simulations of this expanded model are shown in Figure 7A, exhibiting the bimodal distribution observed experimentally (red curves in Figure 7A, simulation; and 7B, experiment).

To further investigate the impact of this mutation on the circuit's functions, simulations were carried out with perturbed parameters to mimic various scenarios. First, the transposition rate was artificially set to zero, and the simulations show that the system can also exhibit a bimodal distribution (Figure 7A, blue), with the OFF peak exhibiting basal GFP expression. Bimodality has been reported to arise from stochastic state switching of a bistable system without any genetic changes^{6,91,94}. The same mechanism leads to simulated bimodality of this LasR-*pLux* circuit when there is no mutation.

While it is almost impossible to eliminate mutation, it is possible to decrease the transposition rate experimentally. To explore the impacts of mutation in a more realistic scenario, simulations were carried out with positive but smaller transposition rates. Interestingly, the system demonstrates a trimodal distribution (Figure 7A, green). In this distribution, there are three groups of cells: ON, OFF, and Mutated. Those cells initialized at the ON state freely transition to and from the OFF state, due to the system's bistability. Meanwhile, all cells have the chance to mutate and stay mutated (Figure 7C).

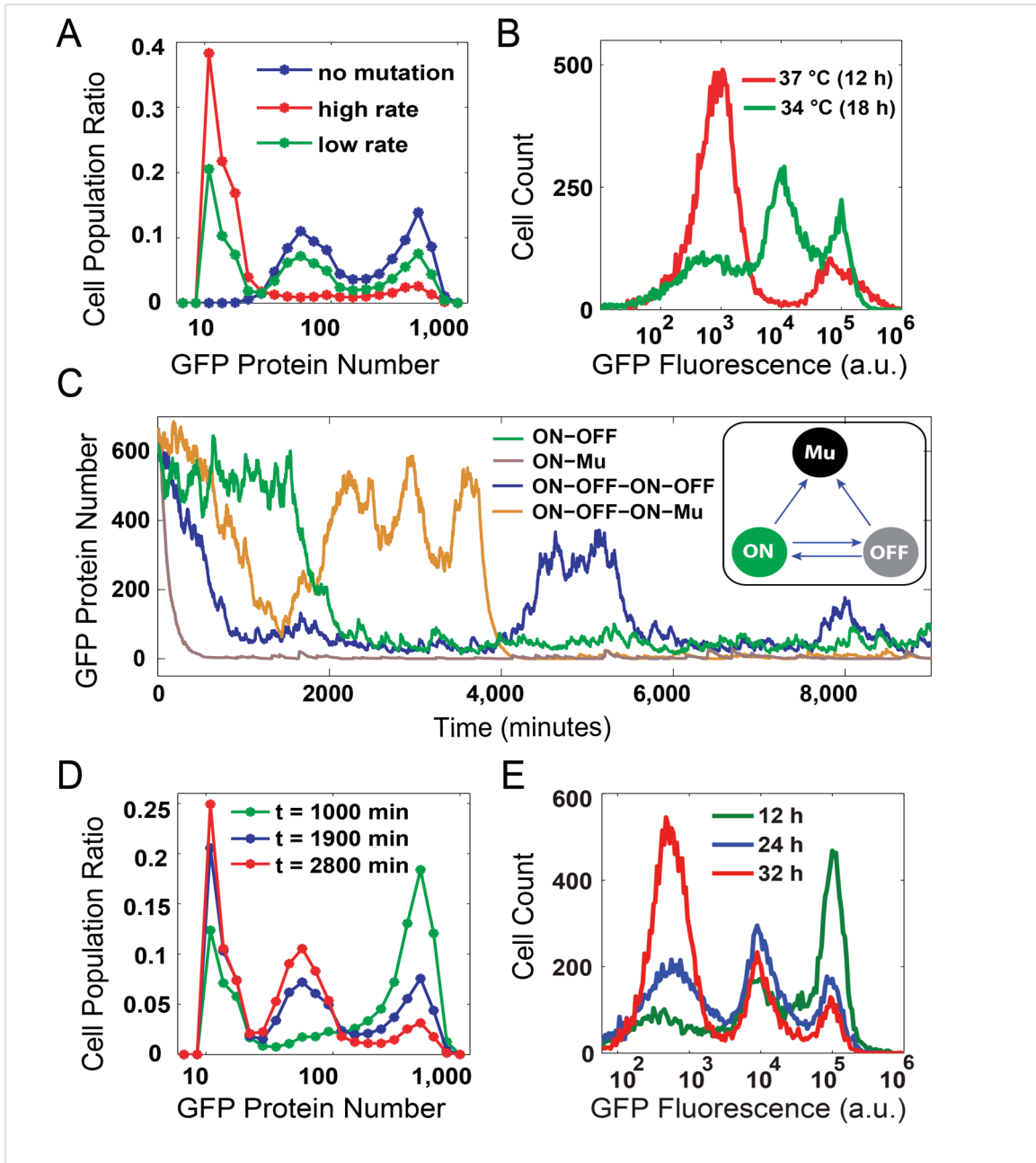


Figure 7: Model predictions and experimental validations of mutation-induced trimodality. (A) Model predictions of GFP expression at several transposition rates: high (red, $k_3=3.6e-6$), low (green, $k_3=4e-7$), and none (blue, $k_3=0$). Histograms were constructed from 8000 single cell stochastic simulations at 1000 ($k_3=3.6e-6$) and 1900 ($k_3=0$ and $k_3=4e-7$) minutes. (B) Experimental validation of the model predictions in (A). Red and green curves correspond to the high and low transposition rates from (A), and they exhibit similar bi- and trimodal responses, respectively. No blue curve is included because mutation could not be eliminated entirely experimentally. (C) Representative stochastic simulations of single cell fluorescence starting from the ON state. All possible

transitions are shown. Inset diagram illustrates all possible state transitions in the simulation. (D) Model predictions of GFP expression with low transposition rate showing temporal evolution of the population from primarily ON cells at an early time (green), to trimodal distributions at intermediate time (blue), eventually falling into a primarily Mutated state at late time (red). (E) Flow cytometry measurements taken at 12 hours (green), 24 hours (blue), and 36 hours (green). Populations show similar dynamics to those predicted by the model in (D), starting with a large ON peak, transitioning to a trimodal distribution, then into primarily Mutated or OFF cells.

Given enough time and the right measurement window, all three groups of cell would be visible. Within this window, the portion of ON and OFF cells will gradually decrease and the number of mutated cells will increase because the mutation is irreversible. The effect of decreased transposition rate is essentially slowing down the ON to Mutation transition rate and giving enough time for ON to OFF transitions and hence the emergence of the OFF peak. Time courses of the simulations demonstrate gradual emergence and evolution of these three populations of cells (Figure 7D).

2.2.6 Experimental validation of trimodal responses by lowering growth temperature.

Previous reports indicated that transposition frequency can be perturbed by growth temperatures⁹⁵⁻⁹⁷. To tune the transposition rate, experiments were carried out with cells cultured at a lower 34 °C temperature, which was shown to obviously slow down crosstalk triggered mutation of this circuit (Figure 8B). Consistent with model predictions, initial ON cells induced with 10^{-8} M C12 exhibited a trimodal response when the growth temperature was tuned from 37 °C to 34 °C (Figure 7B, green). Moreover, temporal evolution of the proportion of each subpopulation was also consistent with model predictions: the portion of ON cells gradually decreased, the Mutation portion

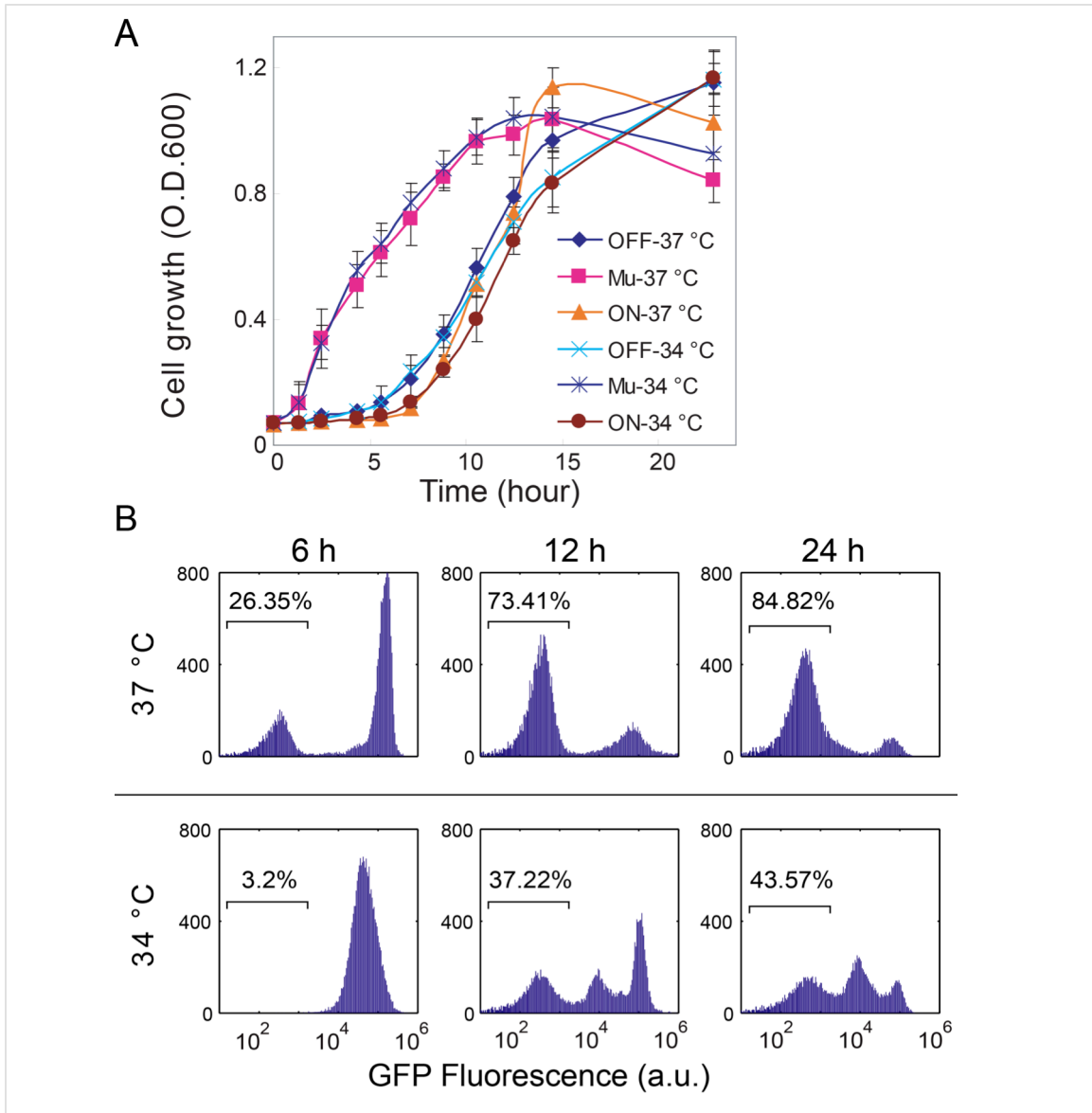


Figure 8: Temperature influence on cell growth rate and population distribution. (A) Growth curves for initial ON, OFF and Mutated cells in 10^{-8} M C12 at 37 °C and 34 °C. The initial ON and OFF cells' growth curves were similar, with a long lag phase in 10^{-8} M C12, while the Mutated cells directly entered exponential growth phase. All populations reached stationary phase after about 15 hours. The three cell types show similar growth curves at 37 °C and 34 °C, indicating that growth temperature does not significantly influence their growth rate. (B) Temperature changes the transposition rate. Top: temporal evolution of the initial ON cells grown in 10^{-8} M C12 at 37 °C. Bottom: time course of the same initial ON cells grown in 10^{-8} M C12 but at 34 °C. Flow cytometry was used to measure the GFP fluorescence at 6 hr, 12 hr, and 24 hr. For each measurement, the percentage of Mutated state cells was calculated. Data shows that higher temperature increases the transposition rate and *IS10* transposase insertion, which promotes the transition from the ON state to the Mutated state.

increased, and the OFF portion increased first and then decreased as time went on (Figure 7E). Growth rates of cells at Mutated, ON, or OFF states were also measured and show no difference when cultured at these two different temperatures (Figure 8A). The emergence of the OFF peak, therefore, is fully accounted for by the decrease of transposition rate, which slows down the direct transitions from ON to Mutation and therefore gives the cells time to layover at the OFF state. This is also evidenced by the smaller portion of Mutated cells when grown at 34 °C compared with 37 °C (Figure 8B).

Furthermore, a microfluidic platform coupled with time-lapse imaging was also employed to verify model predictions⁹⁸. Cells were pretreated with 10^{-9} M C12 until steady state as the initial ON cells before being loaded into the device and induced with 10^{-8} M C12 at 34 °C to mimic experimental protocols used in Figure 7E. Initially, there was only one ON cell loaded into the trap (Figure 9A). At the 8th hour, it can be seen that two populations began to emerge: some cells became OFF and some stayed ON. Mutations started to occur shortly after the 8th hour, and the OFF and Mutation cells accounted for around 90 percent of the population after 16 hours. Eventually mutation state cells took up the majority of the population. There also existed several OFF cells which became ON again, owing to stochastic gene expression noise, but they eventually exhibit a similar evolving process: ON to OFF and Mutation (Figure 9B), which is consistent with the stochastic model simulations shown in Figure 7C.

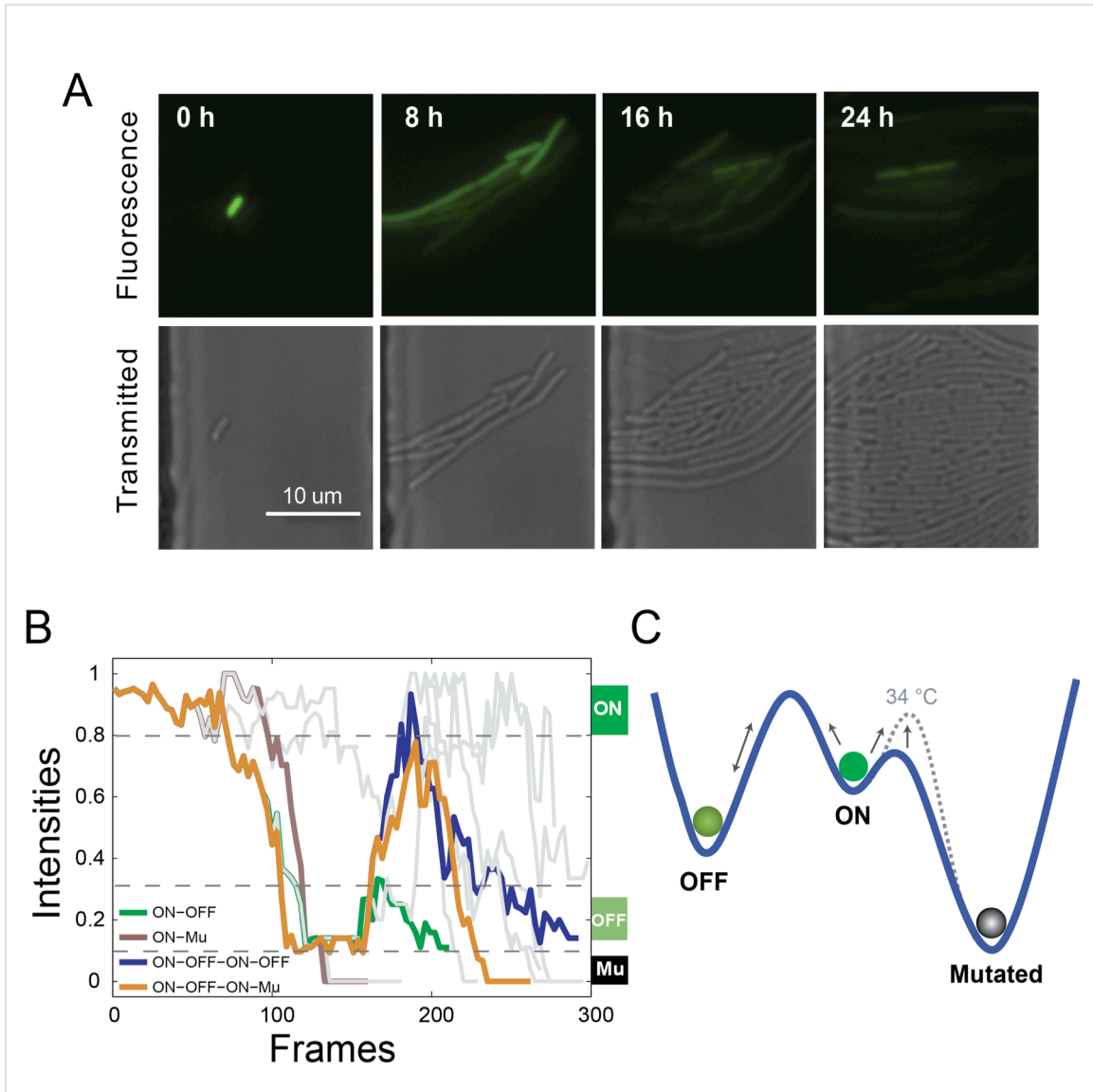


Figure 9: Fluorescence microscopy validation of mathematical model predictions. (A) GFP fluorescence (top) and phase contrast (bottom) images of cells growing in the microfluidic chamber at 0, 8, 16, and 24 hours. Magnification: 40x. (B) Normalized fluorescence expression of representative cells from (A), showing similar behavior to that predicted by the model from Figure 4C. Four cells are colored corresponding to the scenarios in Figure 7C, and the other 11 cells are grey. Each trajectory follows one cell, with the trajectory branching as the cells divide. One frame equals five minutes. (C) Diagram of the mechanism for trimodality. Each “valley” represents one state. Blue curve represents the landscape at 37 $^{\circ}$ C, and the dotted grey curve is the landscape at 34 $^{\circ}$ C. At 37 $^{\circ}$ C, ON state cells can more easily transit to Mutated state because of the low barrier; while at 34 $^{\circ}$ C, the barrier between ON and Mutated states increases, resulting in more ON cells transit to OFF state and promote the emergence of trimodality.

Altogether, the flow cytometry and microfluidic data confirmed the model predicted trimodality, which arises from bistability of positive feedback circuit and host-circuit interactions. In the context of positive feedback circuit, the ON and OFF state can transit freely to each other, but it is easier for ON state cells to transit to the OFF state because of the asymmetric energy barrier (Figure 10). However, the ON cells can also go to the Mutated state, which has an extra advantage of growth (Figure 8A). So compared to OFF state, the majority of ON cells would choose to go to Mutated state at 37 °C, which leads to the bimodal distribution (Figure 5C). When the growth temperature tuned down to 34 °C, the transposition frequency obviously decreased, meaning that the barrier between ON and Mutated state increases. Hence, more ON cells would transit to OFF state, which promotes the emergence of trimodality (Figure 9C).

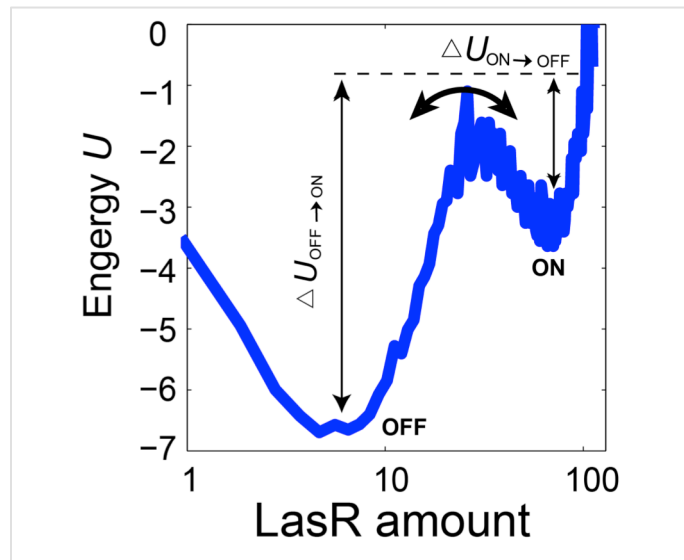


Figure 10: Quasi-potential U and the transition dynamics between stable steady states in the LasR-*pLux* positive feedback system (without genetic mutation). The lower ‘valley’ (with lower potential U) is the stable OFF state and the higher is the stable ON state. According to the stochastic simulation, the energy barrier $\Delta U_{\text{OFF} \rightarrow \text{ON}}$ is much greater than $\Delta U_{\text{ON} \rightarrow \text{OFF}}$, which suggests it is easier for ON state cells to transition to the OFF state. The energy function is calculated according to the probability density distribution of steady state LasR concentrations in each cell.

2.3 Discussion

QS is a ubiquitous mechanism in nature, and its regulator-autoinducer pairs, such as LuxR/LuxI and LasR/LasI, have been used in synthetic biology for a wide range of applications^{12,13,40,51–53,77–80,99,100}. However, evolutionary pressures from limited resources in a competitive environment promote promiscuous bacterial communication, which takes the form of either different genera of bacteria producing the same types of autoinducers or non-specific regulator-autoinducer binding^{51,73,77,101–103}. Therefore, QS regulator-autoinducer pairs are not orthogonal, and there is crosstalk between them. Dissecting the crosstalk is critical for unraveling the underlying principles of bacterial decision-making and survival strategies for both natural and synthetic systems.

In this work, we used synthetic biology approaches to dissect QS crosstalk between LuxR/I and LasR/I. By applying engineering principles to construct modular gene networks, we were able to characterize and categorize QS crosstalk into signal crosstalk, where LuxR can bind with non-originally paired C12 autoinducer to activate *pLux*, and promoter crosstalk, where LasR bind with C12 autoinducer to activate non-originally paired *pLux*. However, regulator crosstalk, in which the naturally paired autoinducer and promoter function through a cross-talking regulator protein, was not detected in this work.

When signal crosstalk is constructed and tested in the context of positive feedback, our results showed a significant shrinkage of the bistable region. Because of this topology's bistable capability and wide presence in most bacterial QS decision-making circuits, such a decrease in bistability robustness due to QS crosstalk suggests a new

strategy for developing anti-infection therapeutics. Namely, we might exploit “artificial” crosstalk to disrupt intercellular communication specificity and collapse the group’s coordination, which could be an efficient and economic approach in medical treatments, especially for QS-dependent bacterial infection.

On the other hand, promoter crosstalk caused complex trimodal responses when embedded within a positive feedback circuit. This can only be explained when network bistability, gene expression stochasticity, and genetic mutations are all taken into consideration. These results highlight the potential for engineering gene networks to express complex behaviors due to host-circuit interactions. We computationally predicted and experimentally verified that the C12-LasR-*pLux* positive feedback circuit could drive the formation of three subpopulations from an isogenic initial culture: one population expressing high GFP expression, the second showing basal GFP expression, and the third population with no GFP expression. The high and low GFP states are the result of positive feedback enabled bistability and gene expression stochasticity-induced random state transitions: commonly reported as a hallmark of many bistable systems^{6,91,94}. This population heterogeneity is not caused by genetic factors.

The third non-GFP population is the result of genetic mutation from *IS10* insertion. The mutation only happened in the C12-LasR-*pLux* positive feedback circuit but not in CP-LasR-*pLux*-C12 (Figure 6) or the C12-LuxR-*pLux* positive feedback circuit (Figure 4). It is, therefore, possible that the special sequence arrangements of the positive feedback circuit (for example, the symmetric *pLux* promoters flanking the LasR gene) on the plasmid coupled with the stress of exogenous protein overexpression led to transposon activation and gene network destruction. This stands in contrast to previously

reported host-circuit interactions, which are primarily related to resource limitation and resulting growth defects¹⁰⁴. Here we were able to illustrate that both the components used and the topology of the network constructed could contribute to resource independent host-circuit interactions. This concept of combining nonlinear dynamics and host-circuit interactions to enrich population diversity expands our understanding of mechanisms contributing to cell-cell variability, and suggests new directions in engineering gene networks to utilize hybrid factors.

Taken together, our studies not only showcase living cells' amazing complexity and the difficulty in the refinement of engineered biological systems, but also reveal an overlooked mechanism by which multimodality arises from the combination of engineered gene circuit^{46,105-108} and circuit-host interactions¹⁰⁴.

2.4 Materials and Experimental Methods

2.4.1 Strains, growth conditions and media.

All cloning experiments were performed in *E. coli* DH10B (Invitrogen, USA), and measurements of positive feedback response were conducted in DH10B and MG1655. Cells were grown at 37 °C (unless specified) in liquid and solid Luria-Bertani (LB) broth medium with 100 µg/mL ampicillin. Chemical 3OC6HSL and 3OC12HSL (Sigma-Aldrich, USA) were dissolved in ddH₂O and DMSO, respectively. Cultures were shaken in 5 mL or 15 mL tubes at 220 rotations per minute (r.p.m), and inducers were added at OD₆₀₀~0.1.

2.4.2 Plasmids construction.

Plasmids were constructed according to standard molecular cloning protocols and the genetic circuits were assembled using standardized BioBricks methods based on primary modules (Table 2) from the iGEM Registry (www.parts.igem.org). The receiver CP-LuxR-*pLux* was constructed from six BioBrick standard biological parts: BBa_K176009 (Constitutive promoter, CP), BBa_B0034 (Ribosome binding site, RBS), BBa_C0062 (luxR gene), BBa_B0015 (transcriptional terminator), BBa_R0062 (lux promoter), and BBa_E0240 (GFP generator, RBS-GFP-T). As an example, to produce the RBS-LuxR part, LuxR plasmid was digested by *XbaI* and *PstI* to produce a fragment while the RBS plasmid was digested by *SpeI* and *PstI* as the vector. The fragment and vector were purified by gel electrophoresis (1% TAE agarose gel) and extracted using a PureLink gel extraction kit (Invitrogen). Then, the fragment and vector were ligated together using T4 DNA ligase, the ligation products were transformed into *E. coli* DH10B and clones were screened by plating on 100 µg/mL ampicillin LB agar plates. Finally their plasmids were extracted and verified by double restriction digest (*EcoRI* and *PstI*) and DNA sequencing (Biodesign sequencing lab in ASU). After confirming that the newly assembled RBS-LuxR was correct, subsequent rounds to produce the RBS-LuxR-Terminator were performed similarly until completing the entire receiver CP-LuxR-*pLux* construction. All the other receivers and positive feedback circuits were assembled similarly. Restriction enzymes and T4 DNA ligase were from New England Biolabs. All the constructs were verified by sequencing step by step. To keep all the constructs' expression consistent in the cell, we transferred all the fragments into the pSB1A3 vector before testing them.

Table 2: Plasmids used in the circuits' construction

Biobrick number	Abbreviation	Description
BBa_R0062	<i>pLux</i>	Promoter activated by LuxR in concert with 3OC6HSL
BBa_R0079	<i>pLas</i>	Promoter activated by LasR in concert with 3OC12HSL
BBa_K176009	CP	Constitutive promoter family member J23107 actual sequence (pCon 0.36)
BBa_B0034	RBS	Ribosome binding site
BBa_B0015	T	Transcriptional terminator (double)
BBa_C0062	LuxR	LuxR repressor/activator
BBa_C0079	LasR	LasR activator
BBa_C0161	LuxI	Autoinducer synthetase for AI from <i>Aliivibrio fischeri</i>
BBa_C0178	LasI	Autoinducer synthetase for PAI from <i>Pseudomonas aeruginosa</i>
BBa_E0240	GFP	GFP generator
pSB1A3	pSB1A3	High copy BioBrick assembly plasmid

2.4.3 Flow cytometry.

All the samples were analyzed at the time points indicated on an Accuri C6 flow cytometer (Becton Dickinson, USA) with 488 nm excitation and 530±15 nm emission detection (GFP). The data were collected in a linear scale and noncellular low-scatter noise was removed by thresholding. All measurements of gene expression were obtained from at least three independent experiments. For each culture, 20,000 events were collected at a medium flow rate. Data files were analyzed using MATLAB (MathWorks).

2.4.4 Hysteresis experiment.

For OFF→ON experiments, initially uninduced overnight culture was diluted at 1:100 ratio into fresh media, grown at 37 °C and 220 r.p.m for about 1.5 hr ($OD_{600} \sim 0.1$), then distributed evenly into new tubes and induced with various amounts of 3OC6HSL or 3OC12HSL. Flow cytometry analyses were performed at 6, 12, and 21 hours to monitor the fluorescence levels, which generally became stable after 6 hours induction according

to our experience. For ON→OFF experiments, initially uninduced cells were induced with 10^{-4} M (or 10^{-9} M) autoinducer and tested by flow cytometry to ensure they were fully induced. Cells were then collected with low-speed centrifugation, washed twice, resuspended with fresh medium (same amount as original culture), and at last inoculated into fresh medium with varying inducer concentrations at a 1:80 ratio. For the LasR-*pLux* positive feedback system, we only diluted once and grew them for 6, 12, 18, 24, or 32 hours, but for the other hysteresis experiments, the ON cells were collected and diluted twice into new medium with the same concentrations of 3OC6HSL or 3OC12HSL at 12 hr and 24 hr.

2.4.5 Growth curve assay.

First, different initial states cells were collected: initial OFF cells were cells grown overnight without inducers, initial ON cells were initial OFF cells induced with 10^{-9} M C12 for 12 hours, and the Mutated cells were cells induced with 10^{-4} M C12 for 12 hours, diluted into fresh media with 10^{-4} M C12, and grown at 37 °C for another 12 hr. Before the growth experiment, all the cells' fluorescence was tested by flow cytometry to verify their states. Growth rate was measured by using absorbance at 600 nm with a plate reader (BioTek, USA). Cells from each state were then diluted into fresh LB media (1000 μ L, O.D. \sim 0.06) with 10^{-8} M C12 and grown at 37 or 34 °C. For each sample, OD was measured by using 200 μ L cultures in a 96-well plate and tested over 24 hours. The experiments were independently replicated three times.

2.4.6 Microfluidics, fluorescence microscopy, and image processing.

The use of microfluidic devices coupled with fluorescence measurement allowed us to measure gene network dynamics in single cells. Media flow direction and speed was controlled through hydrostatic pressure. A detailed description of the chip can be found in reference²¹. Once the cell was loaded into the trap, the flow was reversed and its rate was slowed to $\sim 120 \mu\text{m}/\text{min}$ to ensure that the cells would not be washed away and would receive enough nutrients. Furthermore, care was taken to avoid introducing bubbles to any part of the chip as they considerably disrupt flow. The chip temperature was maintained at $34 \text{ }^\circ\text{C}$ with an external microscope stage (Tokai Hit, Japan). Inducer concentrations were controlled by adjusting the heights of the inducer-containing media syringes relative to one another.

Images were taken using a Nikon Eclipse Ti inverted microscope (Nikon, Japan) equipped with an LED-based Lumencor SOLA SE Light Engine with the appropriate filter sets. The excitation wavelength for GFP was 472 nm, and fluorescence emission was detected with a Semrock 520/35 nm band pass filter. Phase and fluorescent images were taken under a magnification of 100X, and perfect focus was maintained automatically using Nikon Elements software.

Initially OFF cells (K-12 MG1655) with the positive feedback loop plasmid were grown until log phase ($\text{OD}_{600} \sim 0.3$), collected and resuspended with fresh LB media with 0.075% Tween-80 (Sigma-Aldrich, USA), then loaded into the traps (one or two cells for each trap, for best result). Initially ON cells (K-12 MG1655) induced with 10^{-9} M 3OC12HSL (6 hours) were collected, washed, then resuspended with 10^{-9} M inducer and loaded into the trap. $100 \mu\text{g}/\text{mL}$ ampicillin was added into media 1 and 2, but only media

2 was augmented with the corresponding inducer. The microfluidic device was used to control the chemical concentration by switching between medias 1 and 2. For initial ON cells, media 2 was provided to the cells for the duration of the experiment. To prevent photobleaching and phototoxicity to the cells in the trap, exposure time was limited to 100 ms for GFP.

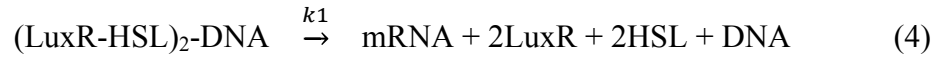
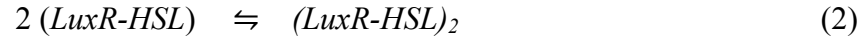
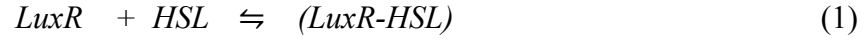
Images were taken every 5 minutes for about 28 hours in total. The pixels in all images are normalized to 0 - 1 range before analysis. One image was chosen for quantification every 15 minutes (i.e. three images). For each cell, the intensity was calculated by averaging three selected points (left, middle, and right) in the cell and then subtracting the background. Since all the cells are offspring of the first initial ON cell, each branch in Figure 9B stands for one progeny. The cells that were washed away or had less than three generations were not analyzed.

2.5 Mathematical Modeling

2.5.1 Deterministic Model Construction

In the positive feedback loop circuit, LuxR production is controlled by the *pLux* promoter with only one LuxR-HSL binding site, which is bound and activated by the complex of LuxR and the autoinducer (3OC6HSL or 3OC12HSL, hereafter denoted as C6 and C12, respectively). GFP expression, as a reporter of the system, is regulated by the same *pLux* promoter and therefore follows the dynamics of LuxR. Therefore, we can directly analyze the LuxR dynamics in the model, comparing the output to the cells' fluorescence, without any loss of explanatory power. Since the LuxR-*pLux* and LasR-*pLux* positive feedback systems are characterized similarly and described by the same

mathematical equations, we explain only the technical details for the LuxR-*pLux* positive feedback loop. Our model is based on the following biochemical reactions:



where *LuxR* is the monomer form of LuxR protein; *HSL* is the autoinducer 3OC6HSL; (*LuxR-HSL*) is the complex of LuxR bound with *HSL*; (*LuxR-HSL*)₂ is the dimer of (*LuxR-HSL*); (*LuxR-HSL*)₂-DNA represents (*LuxR-HSL*)₂ binding to the *pLux* promoter; *mRNA* is the messenger RNA of the LuxR gene; *k*₁ and *k*₂ are the transcription and translation rates, respectively; *d*₁ and *d*₂ are the degradation rates of mRNA and LuxR, respectively.

After C6 concentration reaches a certain threshold, LuxR binds to HSL molecules and forms the active LuxR monomers in the form of (*LuxR-HSL*) (Reaction 1). To quantitatively capture the relationship between the autoinducer concentration and the active LuxR monomers, a Hill function is employed to represent the fraction of LuxR monomers bound by HSL (*f*):

$$f = [\text{HSL}]^{ni} / ([\text{HSL}]^{ni} + K_i^{ni}) \quad [\text{Eq1}]$$

where ni is the binding cooperativity (Hill coefficient) between LuxR and HSL, and K_i represents the dissociation constant between LuxR and HSL (the HSL concentration producing half conversion of LuxR monomers into *LuxR-HSL* complexes). It should be noted that different autoinducers will have different K_i values. Here we assume that the activator LuxR is abundant, and the fraction of active LuxR is independent from LuxR abundance in the cell.

LuxR needs to form a dimer to bind the promoter and activate transcription. We describe the relationship between the dimer and the monomer as the following expression:

$$[\text{LuxR}_2] = [\text{LuxR}]^2 / K_d \quad [\text{Eq2}]$$

where K_d is the dissociation constant for LuxR dimerization. According to reaction (2), two (*LuxR-HSL*) molecules bind together to form a dimer and activate transcription. Additionally, it is necessary to point out that even without autoinducer *LuxR₂* can still bind the *pLux* promoter and initiate leaky transcription of downstream genes. Taken together, the concentration of the functional LuxR dimer that will bind to *pLux* and activate its transcription is:

$$C = (c_0 + f^2) * [\text{LuxR}]^2 / K_d \quad [\text{Eq3}]$$

where C represents the concentration of functional LuxR dimer (*(LuxR-HSL)₂* and *LuxR₂*); c_0 is the fraction of *LuxR₂* that can recognize and bind *pLux* in the absence of autoinducers; K_d is the dissociation constant for dimerization.

(LuxR-HSL)₂ then recognizes and binds to the *pLux* promoter to form the *(LuxR-HSL)₂-DNA* complex together with RNA polymerase and other transcription factors to initiate transcription and produce mRNA (Reactions 3 and 4). So the expression of

mRNA can be modeled as:

$$S_m = c_l + k_l C / (C + K_n) \quad [\text{Eq4}]$$

where S_m represents the production of mRNA; c_l represents the basal mRNA expression without LuxR protein; k_l is the transcription rate; K_n is the dissociation constant between C and $pLux$ promoter.

After transcription, mRNA is translated into LuxR protein (Reaction 5). Here we simplify the whole translation process and capture the production of LuxR protein in the form of:

$$S_p = k_2 * [\text{mRNA}] \quad [\text{Eq5}]$$

where S_p represents the synthesis of LuxR and k_2 is the translation rate.

Next, we take the constitutive degradation of mRNA in the cell into account (Reaction 6) with the equation:

$$D_m = d_1 * [\text{mRNA}] \quad [\text{Eq6}]$$

where d_1 is the degradation rate of mRNA.

Similarly, the degradation of LuxR protein (Reaction 7) is:

$$D_p = d_2 * [\text{LuxR}] \quad [\text{Eq7}]$$

where d_2 is the degradation rate of LuxR.

Finally, we combine the synthesis and degradation (Eq4, 5, 6, and 7) to find the rates of change of the concentrations of mRNA and LuxR:

$$\begin{cases} d[M]/dt = S_m - D_m \\ d[R]/dt = S_p - D_p \end{cases} \quad [\text{Eq8}]$$

where M and R represents mRNA of LuxR and LuxR monomers, respectively.

Combining all the parameters, the two ODE equations can be rewritten as follows:

$$\left\{ \begin{array}{l} \frac{d[M]}{dt} = c_1 + \frac{k_1 C}{C + K_n} - d_1[M]; \\ \frac{d[R]}{dt} = k_2[M] - d_2[R]. \end{array} \right. \quad \text{Where} \quad \begin{array}{l} C = \frac{(c_0 + f^2) \cdot [R]^2}{K_d} \\ f = \frac{[HSL]^{n_i}}{[HSL]^{n_i} + K_i^{n_i}} \end{array} \quad [\text{Eq9}]$$

These two ordinary differential equations were used to model the three positive feedback loops: LuxR-*pLux-C6*, LuxR-*pLux-C12*, and LasR-*pLux-C12*. Owing to the signal and promoter crosstalk, the dissociation constants K_i , K_d , and K_n may be different, as may also be the case with the Hill coefficients and leaky expression without autoinducer. Setting of parameter values is introduced below.

2.5.2 Stochastic Simulation Coupled with Genetic Mutation

The Gillespie algorithm was employed to perform stochastic simulations of the positive feedback loops⁹³. According to our deterministic model (Eq9), two equations capture the time evolution of the biochemical reactions. In this model, there are four independent events in total – mRNA production, mRNA decay, LuxR production, and LuxR decay – which are translated directly to the stochastic model. Simulation data was collected for 8000 cells, and each simulation was run for 40000 steps.

The energy-like function $U(x)$, which denotes the probability and direction of transitions between attractors in a noisy environment, can also be used to interpret state transitions¹⁰⁹. After finishing all simulations, we first calculated the amount of LasR present in each cell (assuming the cells had reached steady state), then divided by the total number of cells. This yielded a probability density distribution of steady state LasR concentrations, which was used to calculate the energy function $U(LasR)$ by the following approach¹⁰⁹:

$$U(\text{LasR}, t) \sim -\ln(P(\text{LasR}), t) \quad [\text{Eq10}]$$

where $P(\text{LasR}, t)$ is the steady-state probability for each LasR concentration at a given time t . In practice, the $P(\text{LasR}, t)$ was derived from the following equation:

$$P(\text{LasR}) = \text{hist}(\text{LasR})/\text{Cellnum} \quad [\text{Eq11}]$$

where $\text{hist}(\text{LasR})$ is a histogram of the amount of LasR in each cell and Cellnum is the total number of simulated cells. The energy-like function U gave us a more vivid and direct understanding of the quasi-potential landscape and the transition dynamics between stable steady states in this positive feedback system. The transition rates between ON and OFF states are decided by the energy barrier ΔU (Figure 10). Unlike the typical bimodality emerged from bistable systems, C12-LasR-*pLux* positive feedback loop displayed an asymmetric bimodal distribution at a population level, which only happened from ON state to OFF state. The model suggests that this asymmetry comes from the different energy barrier of switching between ON and OFF states (Figure 10).

To take the genetic mutation in the LasR-*pLux* positive feedback circuit into account, we added another event in addition to mRNA and LasR production and degradation. Since the genetic mutation only happened in initial ON cells, and because it is easier for cells in high C12 concentration to mutate, we inferred that more LasR in the cell resulted in a higher mutation probability. Moreover, the mutation occurred in the LasR open reading frame, so theoretically the mutation probability is positive as long as the LasR gene is present. Here, we used a Hill function to describe the probability of mutation:

$$Pm = [\text{LasR}]^n / (K^n + [\text{LasR}]^n) \quad [\text{Eq12}]$$

where Pm represents the probability of mutation; n is the Hill coefficient indicating the

cooperativity of mutation causing factors related to LasR concentration; and K represents the dissociation constant in the complicated biochemical reactions. In the Gillespie simulation, the mutation event, independent of the other four events, was described mathematically as:

$$Mu = k_3 * Pm * [LasR] \quad [Eq13]$$

where k_3 is the transposition rate; $[LasR]$ is the amount of LasR in the cell at a given time, and Pm is the probability of mutation as described above. Generally, once the mutation has happened, the LasR gene is broken into two parts and the functional mRNA of LasR cannot be produced any more. Mutated cells theoretically retain the ability to switch state. However, the probability of this occurring is small. In practice, for each cell, when the mutation event had occurred, the transcription rate (k_l) and leaky expression from $pLux$ (c_l) were reduced to very low values, the cell would remain mutated, and the simulation was ended. By tuning the transposition rate, we fit the parameters according to experimental data, which we then used to make predictions.

Next, since the ON, OFF, and Mutation cells have different growth curves under the same experimental conditions, growth rate differences between the three populations were added into the model. From the growth curves, it can be seen that the initial ON and OFF cells' growth curves were similar, with a long lag phase in $1e-8$ M C12, while the Mutation cells directly entered exponential growth. All three populations went to stationary phase after about 15 hours (Figure 8A). Instead of using a population balance model, we employed a simple and efficient method to combine the stochastic model with population dynamics. The cells with greater growth rate would acquire an extra

advantage in their final quantity: each of the three original populations was multiplied by its relative growth rate and then its ratio in the three populations was adjusted.

To simplify the case, we chose three time points (2.5 hr, 7.1 hr and 12.5 hr) and compared their O.D. values (by $OD^{\text{Mutation}}/OD^{\text{ON}}$, $OD^{\text{OFF}}/OD^{\text{ON}}$, and $OD^{\text{ON}}/OD^{\text{ON}}$: ON cells grew slowest) and then made an average to get an averaged relative growth rate, which then was taken into the simulation results. So the final amount of Mutation cells (F_{mu}), OFF cells (F_{off}) and ON cells (F_{on}) are:

$$F_{mu} = S_{mu} * (OD^{\text{Mutation}}/OD^{\text{ON}});$$

$$F_{off} = S_{off} * (OD^{\text{OFF}}/OD^{\text{ON}});$$

$$F_{on} = S_{on} * 1;$$

where S_{mu} , S_{off} , and S_{on} are the primary number of cells which finished the simulation in the Mutation, OFF, and ON states, respectively. Therefore, the proportions of Mutation cells (P_{mu}), OFF cells (P_{off}), and ON cells (P_{on}) are:

$$P_{mu} = F_{mu}/(F_{mu} + F_{off} + F_{on});$$

$$P_{off} = F_{off}/(F_{mu} + F_{off} + F_{on});$$

$$P_{on} = F_{on}/(F_{mu} + F_{off} + F_{on});$$

In this way, the population with a greater growth rate acquired an advantage in its quantity under identical conditions.

2.5.3 Determinations of parameter values

In the *E. coli* cells, even though the transformed plasmid is high-copy, there is also a maximum expression value. According to the BioNumber database¹¹⁰, each protein generally has no more than 1000 copies. Therefore, we chose 1000 molecules per cell to

be the maximum expression value of LuxR and LasR. All other parameters were adjusted under this assumption.

Specifically, the transcription rate (k_1), translation rate (k_2), and degradation rates of mRNA and LuxR (d_1 and d_2 , respectively) were estimated from previous reports and the BioNumbers database (Table 3). Since *pLux* was the only promoter used in the positive feedback circuits, the leaky expression without LuxR or LasR (c_1) did not change between simulations, and it was estimated to be 0.08 min^{-1} . In addition, according to experimental results, basal GFP expression in the absence of autoinducers (c_0) in the LasR-*pLux* positive feedback circuit is about three times larger than in its LuxR-*pLux* counterpart (Figure 11A-B). Therefore c_0 was set to 0.03 and 0.007 for LasR-*pLux* and LuxR-*pLux*, respectively. The Hill coefficients (n_i) and dissociation constants (K_i) between LuxR/LasR and the C6/C12 were fitted from the dose response curves (Figure 11C-F) by the same fitting method used in our previous work^{23,105}. Considering experimental variations, parameters were adjusted within 10% relative error. The generic parameters K_d and K_n are constant and fit to make the model consistent with experimental results (Figure 4B-C). With these fitted parameters, our model captured the experimental hysteresis results and provided insights to understand the difference between the three positive feedback loop variants induced by QS crosstalk. For example, K_d in LuxR-*pLux*-C12 positive feedback was smaller than in LuxR-*pLux*-C6, while K_n was larger for LuxR-*pLux*-C12. This suggests that C12 might bind more easily to LuxR (relative to C6), but the original LuxR-C6 pair has higher affinity for the *pLux* promoter. Additionally, K_n in the LasR-*pLux* positive feedback loop is much bigger for LasR-C12 than for either

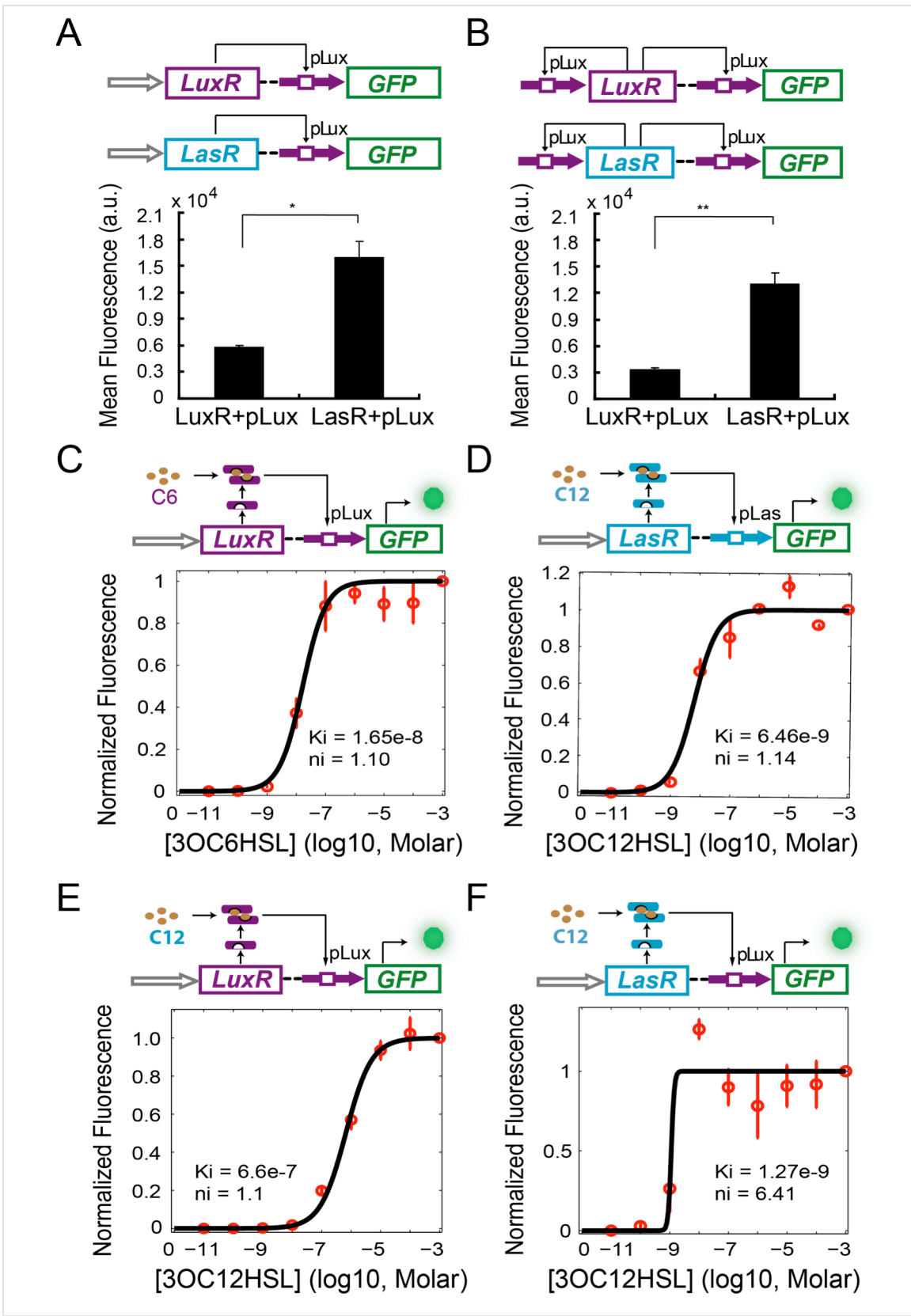


Figure 11: Parameters determination in the model. (A) Comparison of the basal GFP expression from the *pLux* promoter between the two linear CP-LuxR-*pLux* and CP-LasR-*pLux* circuits. (B) Comparison of the basal GFP expression from the *pLux* promoter between the two LuxR-*pLux* and LasR-*pLux* positive feedback circuits. All the data shows that the leakage from the *pLux* promoter in LasR-*pLux* circuits is greater than in LuxR-*pLux* circuits. All the data were averages of three independent measurements shown as mean \pm SD (* $p < 0.05$, and ** $p < 0.01$). Parameters determination from experimental tests: (C) the CP-LuxR-*pLux* circuit induced with C6; (D) the CP-LasR-*pLas* circuit induced with C12; (E) the CP-LuxR-*pLux* circuit induced with C12; (F) the CP-LasR-*pLux* circuit induced with C12. (C) and (D) are the original pairs used to test the functionality of all modules, while (E) and (F) were used to characterize the signal and promoter crosstalk. All of the red data points represent the mean of three independent measurements shown as mean \pm SD. The solid black curves, corresponding Hill coefficients (n_i), and dissociation constants (K_i) between LuxR/LasR and C6/C12 were fitted from the dose response curves by the same fitting method used in our previous work.

LuxR-C6 or LuxR-C12, which indicates that the LasR-C12 dimer has less affinity for *pLux*, and therefore it is more difficult for the system to reach saturation. The parameter combination for the LasR-*pLux* positive feedback loop was used in the stochastic simulation and for predicting trimodality.

To fit the probability of the LasR gene's mutation against experimental results at 37°C, we first approximated the Hill coefficient (n) and the dissociation constant (K) based on the difference between fluorescence values at the ON and OFF states. Different n and K combinations were generated, and it was discovered that $n = 5$ and $K = 400$ best fit the experimental data (Figure 7A-B). In addition, previous reports indicated that transposition rates of IS elements in *E. coli* usually range from $1e-3$ to $1e-7 \text{ min}^{-1}$ ^{96,111}. So the transposition rate in our model was estimated ($k_3 = 3.6e-6 \text{ min}^{-1}$) according to the final experimental data (Figure 7B). To predict the trimodal response, k_3 was adjusted but all the other parameters were held constant. With $k_3 = 4.0e-7 \text{ min}^{-1}$, the simulation exhibited trimodality, validated by the experimental results at 34 °C (Figure 7D-E).

All the parameter values are listed in Table 3 and 4.

Table 3: Parameters for the genetic mutation event in the stochastic simulation of LasR-*pLux* positive feedback system.

Parameters	Description	Value	Source
n	Cooperativity of <i>IS10</i> transposase binding to the plasmid of LasR	5	Estimated and experiment indicated
K	Dissociation constant between transposase and the plasmid DNA	400	Estimated and experiment indicated
k3 (37 °C)	Transposition rate at 37 °C (min ⁻¹)	3.6e-6	Approximated according to experimental results and Ref. ^{96,111}
k3 (34 °C)	Transposition rate at 34 °C (min ⁻¹)	4.0e-7	Approximated according to experimental results
c1	Transcription rate after gene mutation (min ⁻¹)	0.01	Estimated and experiment indicated
k1	Leakage without LasR protein after gene mutation (min ⁻¹)	0.005	Estimated and experiment indicated

Table 4: Parameters for the three positive feedback models.

Parameter	Description	LuxR- <i>pLux</i> -3OC6HSL	LuxR- <i>pLux</i> -3OC12HSL	LasR- <i>pLux</i> -3OC12HSL	Source
k1	Transcription rate (min^{-1})	1.8	1.8	1.8	Ref. ¹¹⁰
k2	Translation rate (min^{-1})	1.6	1.6	1.6	Ref. ¹¹⁰
d1	LuxR/LasR degradation rate (min^{-1})	0.01	0.01	0.01	Ref. ^{52,112}
d2	mRNA degradation rate (min^{-1})	0.33	0.33	0.33	Ref. ^{110,113}
c1	Leakage without LuxR or LasR protein (min^{-1})	0.08	0.08	0.08	Approximated
c0	Leakage without AHL	0.007	0.007	0.03	Approximated
Kd	Dissociation constant of LuxR-HSL dimerization	600	180	720	Approximated
Kn	Dissociation constant of $[\text{LuxR-HSL}]_2$ binding DNA	2.6	14.7	177	Approximated
Ki	HSL concentration producing half occupation of <i>pLux</i> promoter	1.6e-8	6.6e-7	6.9e-9	Measured by experiments
ni	Hill coefficient	1.3	1.1	6.4	Measured by experiments

CHAPTER 3

ENGINEERING OF A SYNTHETIC QUADRASTABLE GENE NETWORK TO APPROACH WADDINGTON LANDSCAPE AND CELL FATE DETERMINATION

3.1 Introduction

Multistability is a mechanism that cells use to achieve a discrete number of mutually exclusive states in response to environmental inputs, such as the lysis/lysogeny switch of phage lambda^{3,24} and sporulation/competence in *Bacillus subtilis*^{25,26}. In multicellular organisms, multistable switches are also common in the cellular decision-making including the regulation of cell-cycle oscillator during cell mitosis²⁷, Epithelial-to-Mesenchymal transition and cancer metastasis^{28,29}, and the well-known cell differentiation process, which is a manifestation of cellular state determination in a multistable system^{30,31}. However, loss of multistability can drive cells to acquire metastatic characteristics and stabilize highly proliferative, pathogenic cellular states in cancer¹¹⁴.

C. H. Waddington in 1957 hypothesized the “epigenetic landscape” to explain canalization and fate determination mechanism during cell differentiation¹¹⁵. In this hypothesis, differentiation is depicted as a marble rolling down a landscape with multiple bifurcating valleys and eventually settles at one of the local minima, corresponding to terminally differentiated cells. More recent theoretical studies further proposed the local minima to be modeled as steady states or attractors of dynamical systems, which can be

mathematically described using differential equations^{116,117}. As such, cell differentiation can be interpreted as a state transition process on a multistable dynamic system.

A myriad of theoretical analysis have investigated the functioning of such systems and quantified the Waddington landscape and developmental paths through computation of the probability landscape for the underlying gene regulatory networks¹¹⁷⁻¹¹⁹. Recent studies also revealed that the potential landscape and the corresponding curl flux are crucial for determining the robustness and global dynamics of non-equilibrium biological networks^{120,121}. Furthermore, the multiple stable steady states have been predicted beyond the bistable switches with or without epigenetic effects, which is reflected in slow timescales^{120,122}. Experimental researches, however, mostly focus on bistable switches, involving transitions between only two states. And demonstrations, from a combination of experiments and computational modeling, for the existence and operation of such a landscape in a higher dimensional multistable system are still lacking. Moreover, it remains unknown how gene regulatory networks (GRNs), gene expression noise, and signal induction together shape the attractor landscape and determine a cell's developmental trajectory to its final fates¹²³⁻¹²⁵.

Complex contextual connections of GRNs have impeded experimentally establishing the shape and function of the cell fate landscape. Rationally designed and tunable synthetic multistable gene networks in *E. coli*, however, could form well-characterized attractor landscapes to enable close experimental investigations of general principles of GRN regulated cellular state transitions. Since the functioning of these principles only requires the most fundamental aspects of gene expression regulation, they would also be applicable for cell differentiation regulations in mammalian cells. Here, we

combine mathematical theory, numerical simulations, and synthetic biology to probe all possible sub-networks of mutually inhibitory network with positive autoregulations (MINPA, Figure 12A), which has been hypothesized to have multistability potentials^{31,119}. Moreover, MINPA and its sub-networks are recurring motifs enriched in GRNs regulating hematopoietic development (Gata1-Pu.1,¹²⁶), trophoctoderm differentiation (Oct3/4-Cdx2,¹²⁶), endoderm formation (Gata6-Nanog,^{117,128}), and bone, cartilage, and fat differentiation (RUNX2-SOX9-PPAR- γ ,^{129,130}).

3.2 Results

3.2.1 MINPA circuit construction and multistability analysis

Engineered circuit of MINPA (Figure 12B) is designed to use two hybrid promoters, *Para/lac* and *PLux/tet*, which are characterized experimentally to show small leakage and high nonlinearity (Figure 12D-E). For MINPA topology, hybrid promoter *Para/lac* drives *AraC* and *TetR* expression, representing the node X in Figure 12A, whereas *PLux/tet* controls *LuxR* and *LacI* transcription, representing the node Y. AraC and LuxR activate *Para/lac* and *PLux/tet* in the presence of Arabinose and AHL (3OC6HSL) respectively, forming positive autoregulations. IPTG inhibits the repressive effect of LacI on TetR expression, while aTc counteracts TetR repression on LacI. Hence, the two nodes form the topology presented in the conceptual design shown in Figure 12A. Green fluorescent protein (GFP) and mCherry serve as the corresponding readouts of *PLux/tet* and *Para/lac* activities in living cells (Figure 12B).

Topologies of MINPA and all its subnetworks can be divided into four layers, from one- to four-dimensional networks based on the number of regulatory edges (Figure

12C) and further categorized into nine groups (R, A, R-A, RA, R², A², RA², R²A, R²A²) based on the configurations of activation and inhibition.

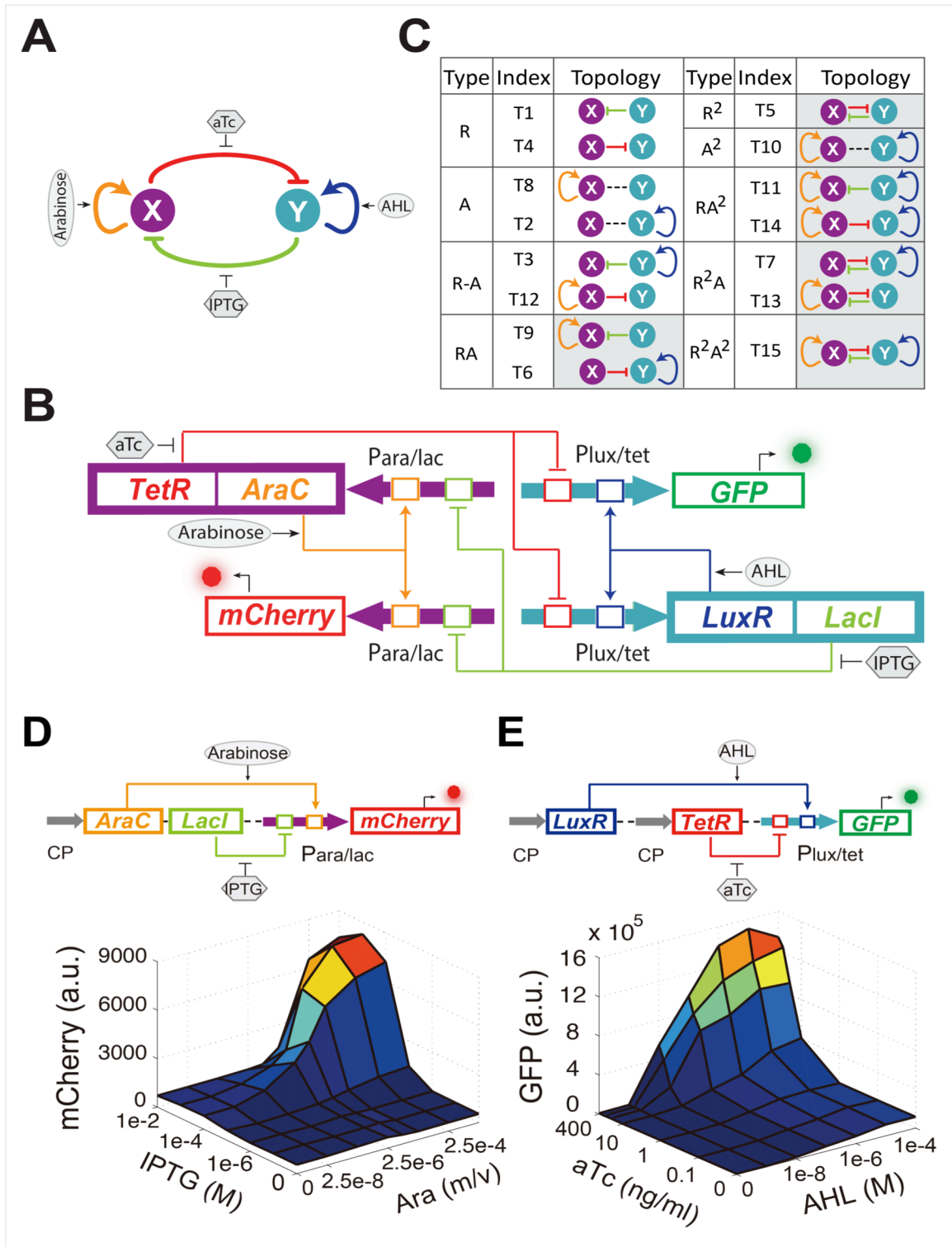


Figure 12: Conceptual and experimental design of MINPA and its sub-networks. (A) Abstract diagram of MINPA topology, where X and Y mutually inhibit (T-bars) each other and auto-activate (arrowheads) itself. Four inducers to regulate the four color-coded regulatory edges are also listed. (B) Molecular implementation of the MINPA network. *Para/lac* (purple arrow) is activated by AraC (yellow) and repressed by LacI (light green), while *PLux/tet* (cyan arrow) is activated by LuxR (blue) and repressed by TetR (red). Arabinose and AHL (oval) can induce AraC and LuxR activation, respectively. IPTG and aTc (hexagon) can respectively relieve LacI and TetR inhibition. GFP and mCherry serve as the readout of *Para/lac* and *PLux/tet*, respectively. Therefore, TetR and AraC collectively form the node X in (A), color-coded as purple rectangle. Similarly, LuxR and LacI collectively form the node Y in (A), color-coded as cyan rectangle. Genes, promoters and regulations are color-coded corresponding to the topology in (A). (C) List of MINPA and its 14 sub-networks. T represents “topology”. R represents “repression”, and A represents “autoactivation”. Superscript is used to describe the number of such types of edges. Topologies with shaded background were later constructed and analyzed experimentally. (D-E) Top: Biological devices for testing promoter *Para/lac* (B) and *PLux/tet* (C), respectively. Fluorescence was measured by flow cytometry at 12 hr and 24 hr (not shown) after adding the inducers. All the data points were averaged from three repeated experiments. Grey arrows represent constitutive promoters (BBa_K176009). Bottom: Dynamic responses for *Para/lac* (D) and *PLux/tet* (E) through induction with Arabinose (Ara) and IPTG, and AHL and aTc, respectively. Presented data was the mean value of three replicates. mCherry and GFP serves as the readout of the two promoters.

3.2.2 Systematical multistability evaluation of MINPA and its sub-networks.

In order to experimentally evaluate dynamic properties of these networks, we constructed nine circuits including tunable positive feedbacks (T6 and T9), mutual inhibition (T5), dual-positive feedbacks (T10), and their combinations (T7, T11, T13, T14 and T15, Figure 13). One-dimensional networks (T1, T4, T2 and T8) and trivial two-dimensional networks (T3 and T12) are excluded for their low multistability probability. All motifs were constructed using the same set of components (Figure 12).

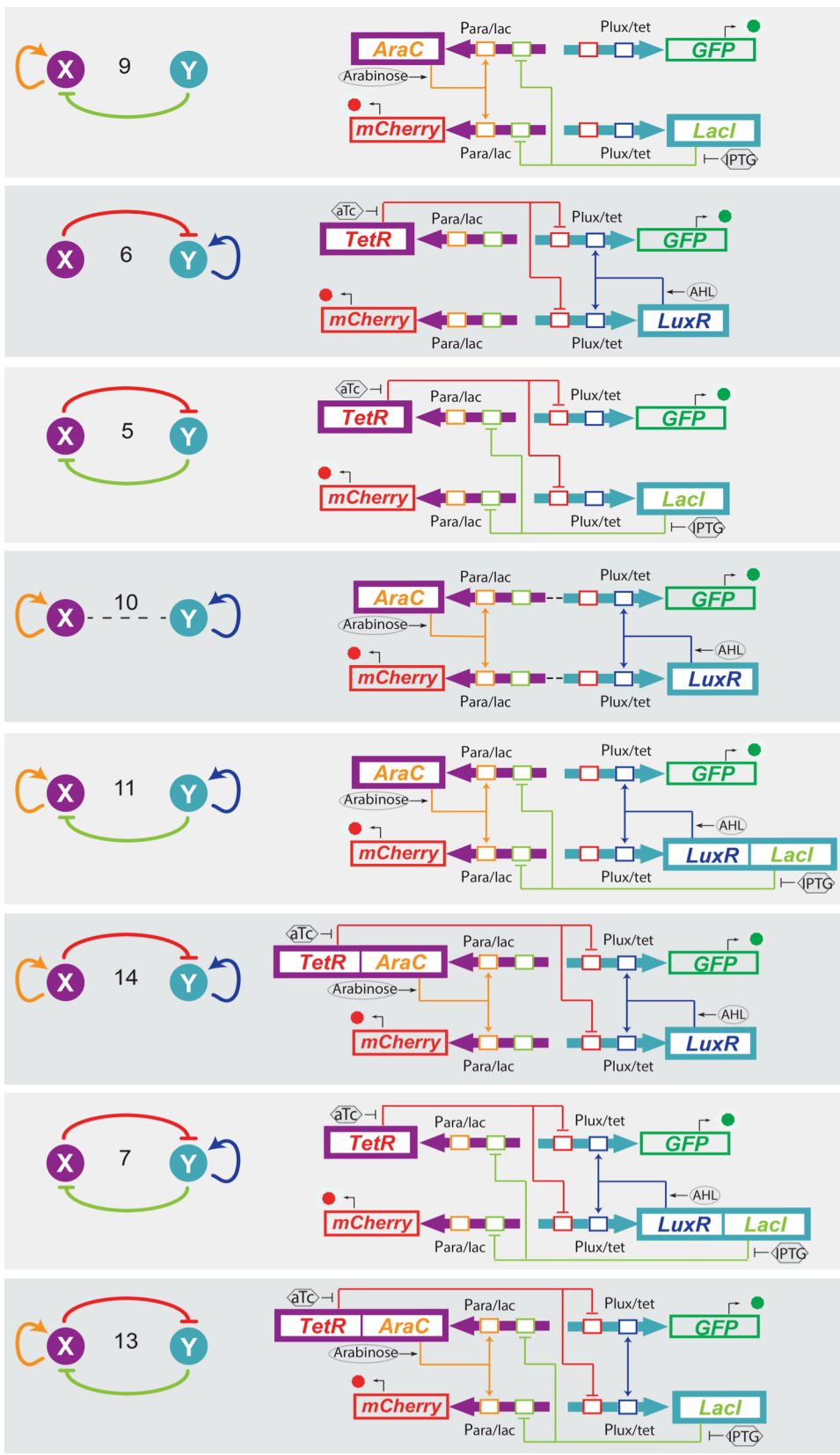


Figure 13: Abstract diagrams and molecular implementation of the eight MINPA sub-network topologies. The subnetworks include tunable positive feedbacks (T6 and T9), mutual inhibition (T5), dual-positive feedbacks (T10), and their combinations (T7, T11, T13, and T14). Genes, promoters, and regulations are color-coded corresponding to the topology on the left side.

Probing a circuit's multistability typically requires thorough hysteresis experiments covering wide ranges of doses for all inducers^{6,91,131}, which becomes infeasible for nine complex networks with four inducers. To improve the efficiency of probing multistability and tunability, we designed a "sequential induction" method to accelerate exploration of unknown high dimensional bifurcation spaces, instead of conventional "back and forth" hysteresis on one parameter dimension. The main concept relies on the fact that multistable gene networks could exhibit discontinuous jump from one state to another in response to changing parameter (inducer) combinations. Taking the classic "toggle switch" as an example, the circuit can be tuned by two external inducers and its two-parameter bifurcation diagram has a stretched *S* shape (Figure 14A). Initialized at an arbitrary state A, the cells could reach State C in the bistable region directly when induced with both inducers simultaneously. If the cells are first induced by Inducer I to go to state B, they will also reach State C after Inducer II is added. However, if the same dose of Inducer II is applied first, cells will cross the bifurcation plane to state D on the low-Response surface and then reach state E with addition of Inducer I (Figure 14A). State C and E are two different steady states with the same induction dosages, illustrating hysteresis and verifying multistability.

To test our theoretical analysis, a synthetic toggle switch circuit was constructed (Figure 14B-C). Following experimental design principles, we designed a protocol to

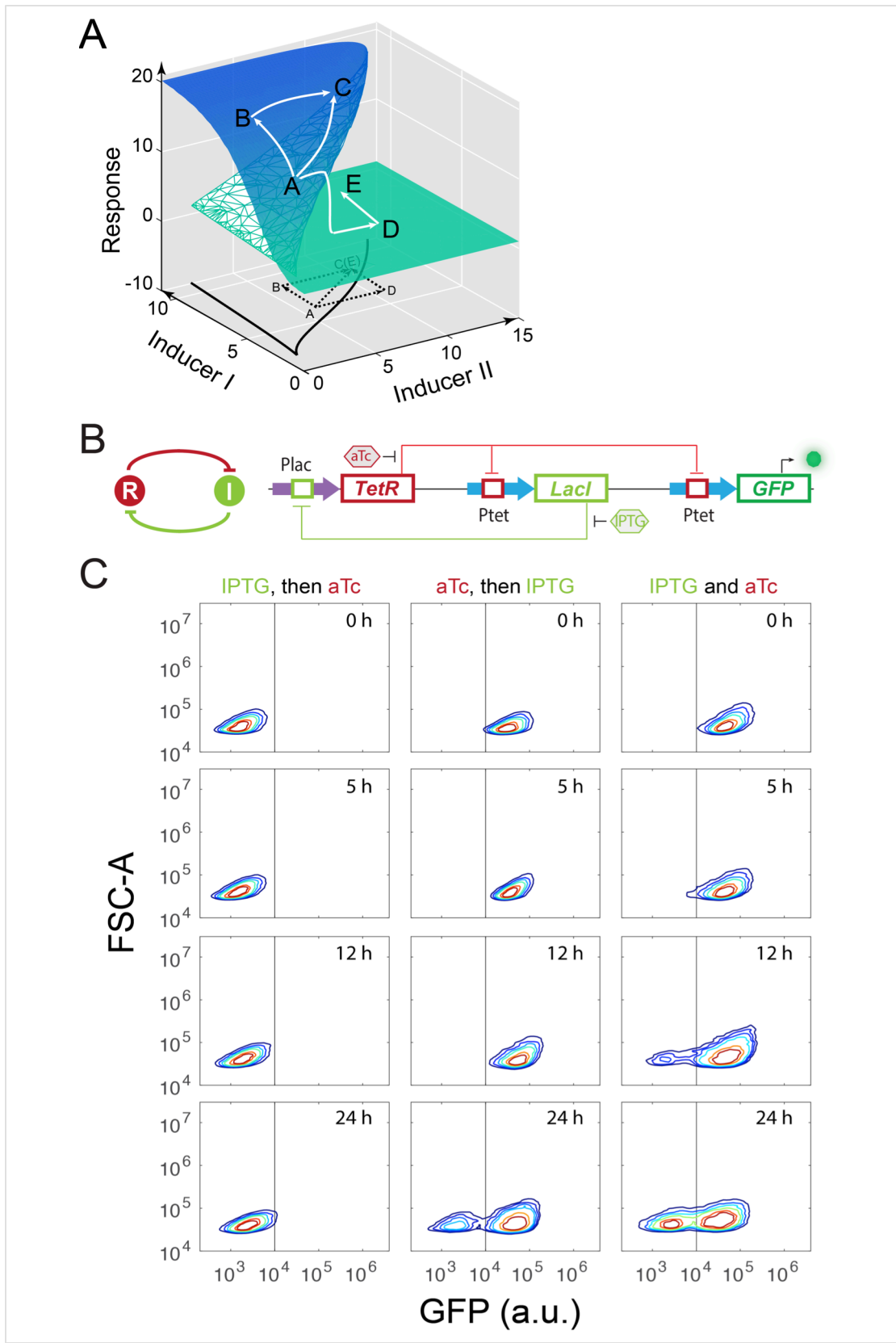


Figure 14: Sequential induction rationale and experimental validation in synthetic toggle switch circuit. (A) Schematic illustration of rationale for sequential induction. This two-parameter bifurcation diagram of a bistable toggle-switch depicts all steady state values of response (Z-axis) with combinations of inducer I and II (X and Y axes). Arrows illustrate order and direction of inductions and consequent steady state value changes. Solid lines on the X-Y plane are the boundaries of bistability. Dashed lines on the X-Y plane are projections of solid white arrowheads. (B) Abstract diagram and molecular implementation of the toggle switch circuit. TetR (R) and LacI (I) mutually inhibit each other through binding to *Ptet* and *Plac* promoter, respectively. IPTG and aTc (hexagon) can respectively relieve LacI and TetR inhibition. GFP serves as the readout of *Ptet*. (C) Time course results of the sequential induction. The y-axis represents forward scatter (FSC-A), and the x-axis indicates GFP fluorescence. IPTG and aTc were sequentially (left and middle columns) or simultaneously (right column) applied to induce the toggle circuit. The first inducer was added to the media for 5 hr, and then the second inducer was added. Fluorescence was measured by flow cytometry at 0, 5, 12, and 24 hr after the second inducer was added into the cultures. The concentration of IPTG and aTc is 8×10^{-5} M, and 100 ng/ml, respectively. Experiments were repeated for at least three times, and representative results were shown.

show the sequential induction effects. We first employed IPTG to induce the circuit for 5 hr, and then aTc was added. Time course results showed that cells stayed at low-GFP state till 24 hours (Figure 14C). However, cells induced with aTc first, and then IPTG mainly stayed at high-GFP state, another stable steady state under this condition. Simultaneous aTc and IPTG induction produced similar cell distributions. These results show that sequential induction can be used as a strategy to quickly explore a multistable potential landscape for complex non-equilibrium systems.

Without knowing the exact bifurcation range beforehand, such ordered sequential inductions could help quickly explore the irregular bifurcation space to reveal multistability for systems with complicated bifurcations, which is typically caused by interfering parameters. Similar sequential induction techniques have been shown to enable access of otherwise hard-to-reach cell death states in breast cancer cells¹³⁶. This strategy has also been widely employed in directed differentiation of stem cells to

specific lineages^{133–135} and reprogramming somatic cells to induced pluripotent stem cells¹³⁶. Although specific inducer concentrations are required to observe the effects of this strategy in synthetic circuits, sequential induction with pre-selected inducer combinations can help perform a coarse-grained exploration from different directions in the parameter space. Furthermore, stochastic gene expression of the circuits also contributes to cellular population distribution thus leads to pronounced sequential induction effects, given experimentally feasible amount of time, when the system is entering its multistable region from different directions. Therefore, distinct final states, or even different population distributions, under sequential induction strongly suggests the existence of nonlinear dynamics, including multistability.

Using the sequential induction approach, we tested the nine circuits using flow cytometry. Cells were first induced by inducer I, inducer II was then added into the media for another 24 hours. Depending on the network configuration, four different dual-inducer combinations were used. For example, Arabinose and IPTG were applied sequentially and simultaneously to T9, T13, T11 and T15, respectively (Figure 15A). It can be seen only T15 exhibits significant expression difference between three induction patterns, while the others show little change (Figure 15A and Figure 16A). It should be noted that T15 also exhibits tri-modality of fluorescence expression, suggesting multistability given the presence of gene expression noise, which is partially consistent with our computational predictions. Similarly, AHL and aTc were applied to T6, T7, T14, and T15, respectively (Figure 15B and Figure 16B). Results show that only T15 exhibits significant fluorescence pattern change with different inductions, whereas T6 and T7

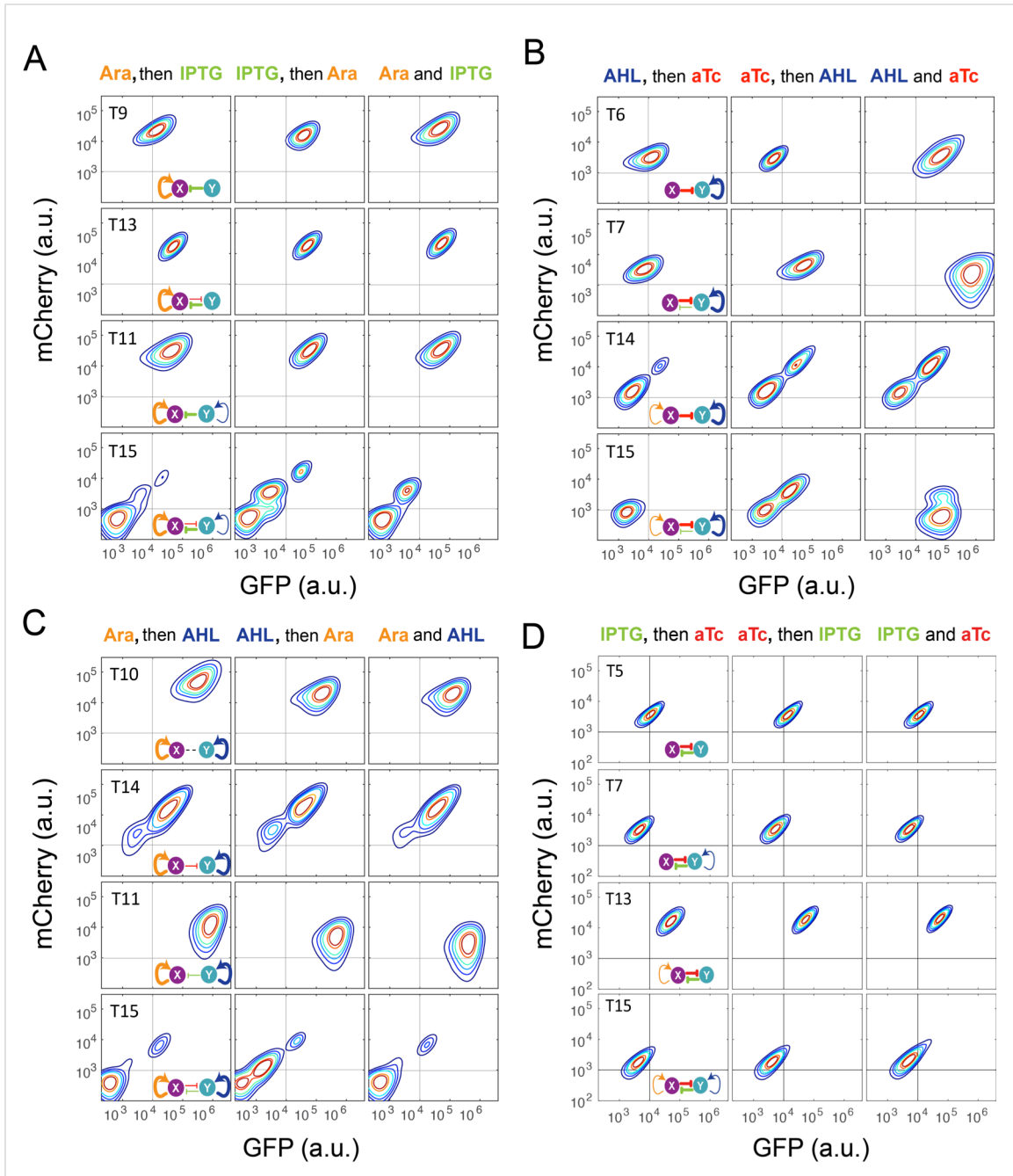


Figure 15: Sequential induction of MINPA and its sub-networks. (A) Arabinose (Ara) and IPTG were sequentially (left and middle columns) or simultaneously (right column) applied to induce T9, T13, T11, and T15. T: topology. The concentration of Arabinose and IPTG is $2.5 \cdot 10^{-5}$ m/v, and $5 \cdot 10^{-5}$ M, respectively. To indicate the effects of inducers, we used the same color for applied inducers and its regulated connections, which were also shown in bold lines. The other non-regulated connections are represented by thin lines. (B) AHL and aTc were sequentially (left and middle) or simultaneously (right) applied to induce T6, T7, T14, and T15. The concentration of AHL and aTc is $1 \cdot 10^{-4}$ M,

and 200 ng/ml, respectively. (C) Ara and AHL were sequentially (left and middle) or simultaneously (right) applied to induce T10, T14, T11, and T15. The concentration of Arabinose and AHL is 2.5×10^{-5} m/v, and 1×10^{-8} M, respectively. (D) IPTG and aTc were sequentially or simultaneously applied to induce T5, T7, T13, and T15. The concentration of IPTG and aTc is 1×10^{-4} M and 200 ng/ml, respectively. Samples were treated with the first inducer till OD_{600} is about 0.15 and then the second inducer was added. Cells were grown for another 24 hours before measured by flow cytometry. The experiments were performed in triplicate and repeated two times, and representative results are presented. The inducers are color-coded as visual assistance to indicate which edge of inset diagram it regulates.

exhibit minor uniform shifts of expression. T14, although exhibiting bimodality, only shows a ratio change of two populations between three inductions and no sign of bifurcation. Sequential induction by Arabinose and AHL combinations has little effect on T10, T14 and T11, but T15 displays three notable populations for AHL-then-Arabinose induction (Figure 15C and Figure 16C). IPTG and aTc were also tested on T5, T7, T13 and T15, but no notable dynamics were observed (Figure 15D and Figure 16D). Taken together, T15, the full MINPA topology, shows the most variety and complexity in population heterogeneity under sequential inductions, suggesting this circuit has the highest potential to generate complex multistability within our induction range and hence enable us to approach the Waddington landscape.

3.2.3 Bifurcation and hysteresis verification of multistability

Next, operating principles and full tunability of T15 (MINPA) were further examined by using four inducers (Arabinose, AHL, aTc, and IPTG) to fine tune the strength of regulations and perturb the system (Figure 17A). Uninduced cells showed low GFP and low mCherry expression (low-low state, LL). In the presence of AHL and aTc, high GFP and low mCherry (GFP state) is observed; low GFP and high mCherry

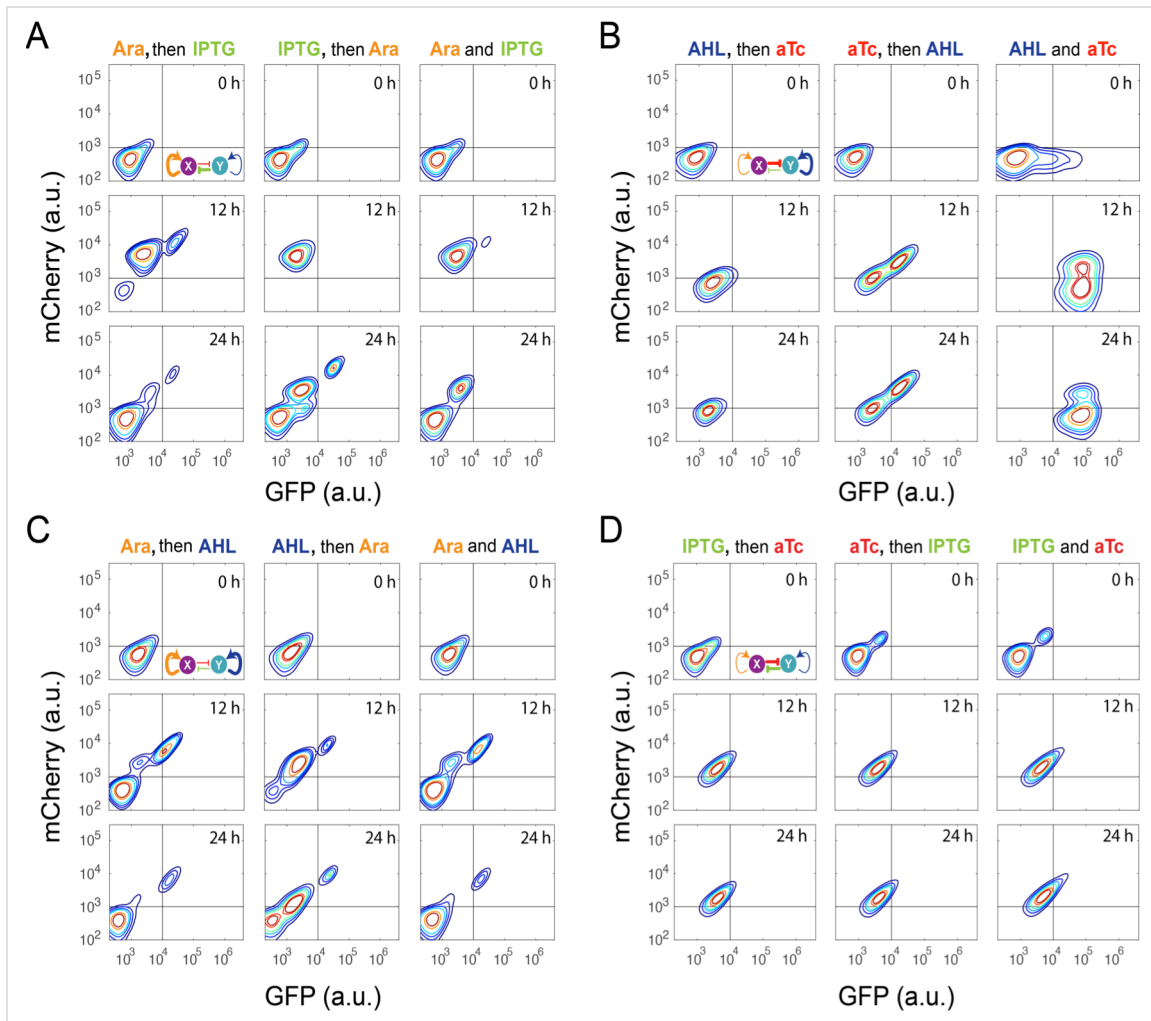


Figure 16: Time course results of sequential induction for the MINPA (T15) circuit. (A) Arabinose (Ara) and IPTG were sequentially (left and middle columns) or simultaneously (right column) applied to induce T15. The first inducer was applied for 5 hr, and then the second inducer was added into the culture. Fluorescence was measured by flow cytometry at 0, 12, and 24 hr after the second inducer was added into the culture. The concentration of Ara and IPTG is $2.5 \cdot 10^{-5}$ m/v, and $5 \cdot 10^{-5}$ M, respectively. (B) AHL and aTc were sequentially (left and middle) or simultaneously (right) applied to induce T15. The first inducer was applied for 6.5 hr, and then the second inducer was added into the culture. The concentration of AHL and aTc is $1 \cdot 10^{-4}$ M, and 200 ng/ml, respectively. (C) Ara and AHL were sequentially (left and middle) or simultaneously (right) applied to induce T15. The first inducer was applied for 5 hr, and then the second inducer was added into the culture. The concentration of Arabinose and AHL is $2.5 \cdot 10^{-5}$ m/v, and $1 \cdot 10^{-8}$ M, respectively. (D) IPTG and aTc were sequentially (left and middle) or simultaneously (right) applied to induce T15. The first inducer was applied for 6.5 hr, and then the second inducer was added into the culture. The concentration of IPTG and aTc is $1 \cdot 10^{-4}$ M, and 200 ng/ml, respectively. The inducers are color-coded as visual assistance to indicate which edge of inset diagram it regulates.

(mCherry state) emerged with induction of Arabinose; and high GFP and high mCherry (high-high state, HH) was achieved when induced with Arabinose and AHL. These results verify that our engineered MINPA circuit is functioning as designed and fully controllable with four distinct states reachable through appropriate inductions, respectively.

To help design experiments to further investigate the circuit's quadrastability, a detailed mathematical model was developed to describe the system. Using parameters derived from hybrid promoter testing experiments, bifurcation analysis was carried out to systematically quantify MINPA's dynamic behavior (Figure 17B). Figure 17B is the three-dimensional bifurcation diagram, where levels of GFP and mCherry represent the states of node X and Y, and " AR/AL " is a lumped parameter composed of a fixed ratio of the concentrations of Arabinose and AHL. Overall, it can be seen that the system, initialized without induction, is predicted to be quadrastable (shown as four colored spheres, representing LL (grey), GFP (green), mCherry (rose), and HH (golden) state, respectively) but with the low-low state to have dominant attractiveness (shown as the big gray sphere) when AR/AL is low (C1). However, when AR/AL level is within an intermediate range, relative stabilities between different states become comparable. When AR/AL level increased from C1 to C2, the circuit's quadrastability becomes well pronounced, illustrated as four similar-sized colored spheres on the same gray plane, which represents the low-low, GFP, mCherry, and high-high state, respectively (Figure 17B). As AR/AL continues to increase from C2 to C3, while the other three SSS remain stable, the stability of the GFP branch disappears. Further increase of AR/AL results in only one stable state-the high-high state, shown as the orange sphere with biggest size.

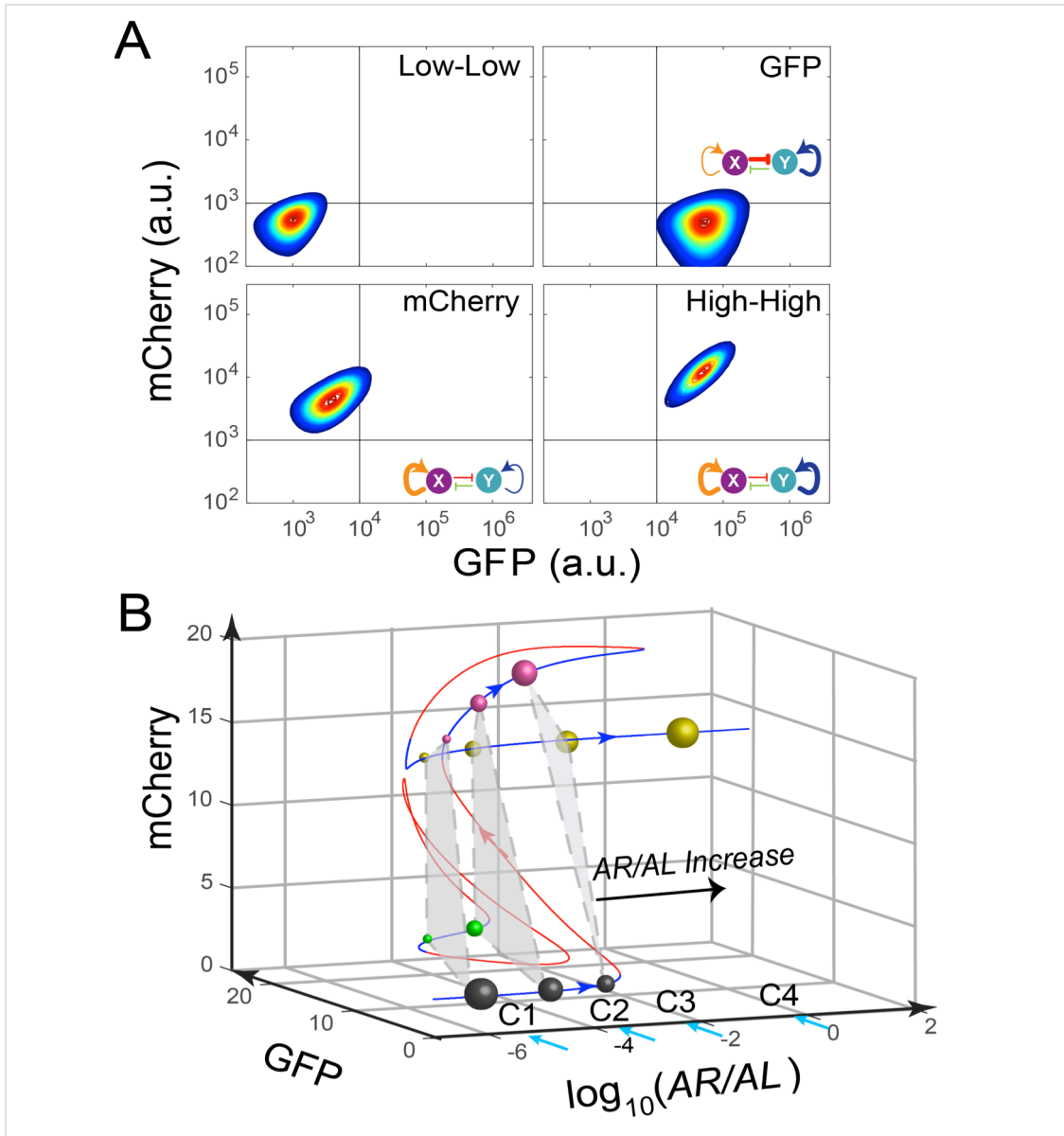


Figure 17: MINPA has four individual cell states and bifurcation analysis of MINPA. (A) Engineered MINPA is tunable to reach four individual states: low-low, GFP, mCherry, and high-high, under no induction, $1 \cdot 10^{-4}$ M AHL and 100 ng/ml aTc, $2.5 \cdot 10^{-5}$ m/v Arabinose, $1 \cdot 10^{-4}$ M AHL and $2.5 \cdot 10^{-3}$ m/v Arabinose, respectively. To indicate the effects of inducers, we used the same color for applied inducer and its regulated connection (bolder lines) in the MINPA topology. (B) 3-D bifurcation diagram of MINPA. AR/AL is a lumped parameter composed of increasing concentrations of Arabinose and AHL, but the ratio of Arabinose and AHL is fixed, i.e., $[Arabinose]/[AHL]$ is a constant. GFP and mCherry represent the states of node X and Y. Blue lines represent stable steady states, while red ones are unstable steady states. Grey, green, rose, and golden spheres represent low-low, GFP, mCherry, and high-high state, respectively. And the size of spheres correlates with the attractiveness of each state. C1, C2, C3, and C4 are four increasing concentrations of Arabinose and AHL used for experimental probing.

To establish MINPA's quadrastability and tristability as predicted, hysteresis, a hallmark of multistability^{23,39,91}, of the network was tested. Initialized at the low-low state, cells were induced by increasing doses of *AR/AL* corresponding to C1 to C4 and measured by flow cytometry (Figure 18A). As predicted, C1_{LL} (cells with initial Low-Low state grown at C1 condition) experiment demonstrates uniform low-low fluorescence profile, due to the low-low state's dominant attractiveness, and C4_{LL} shows a uniform high-high profile. Interestingly, C3_{LL} indeed illustrates tri-modality, which is the result of predicted tristability. C2_{LL} experiment, on the other hand, exhibits enough heterogeneity to signal high-high, low-low, and mCherry state, but does not illustrate significant trace of GFP state. Given that GFP state is achieved through combinational induction of AHL and aTc (Figure 17A), we hypothesize that the GFP state here is not easily accessible with AHL induction only. Next, cells initialized at high-high states were collected and diluted into fresh media with the same concentrations of *AR/AL* (Figure 18B). As predicted, these cells keep high-high expression profile even with inductions as low as C1, another demonstration that the system is already multistable at C1. Taken together, the two sets of experiments demonstrated clear hysteresis and verified existence of three of the four predicted SSS.

3.2.4 Experimental demonstration of model-guided quadrastability of MINPA

To further investigate what determines the accessibility of certain SSS in this quadrastable system and how cells navigate this attractor landscape, we take into account gene expression stochasticity²³ to sketch out MINPA's quasi-potential attractor landscape

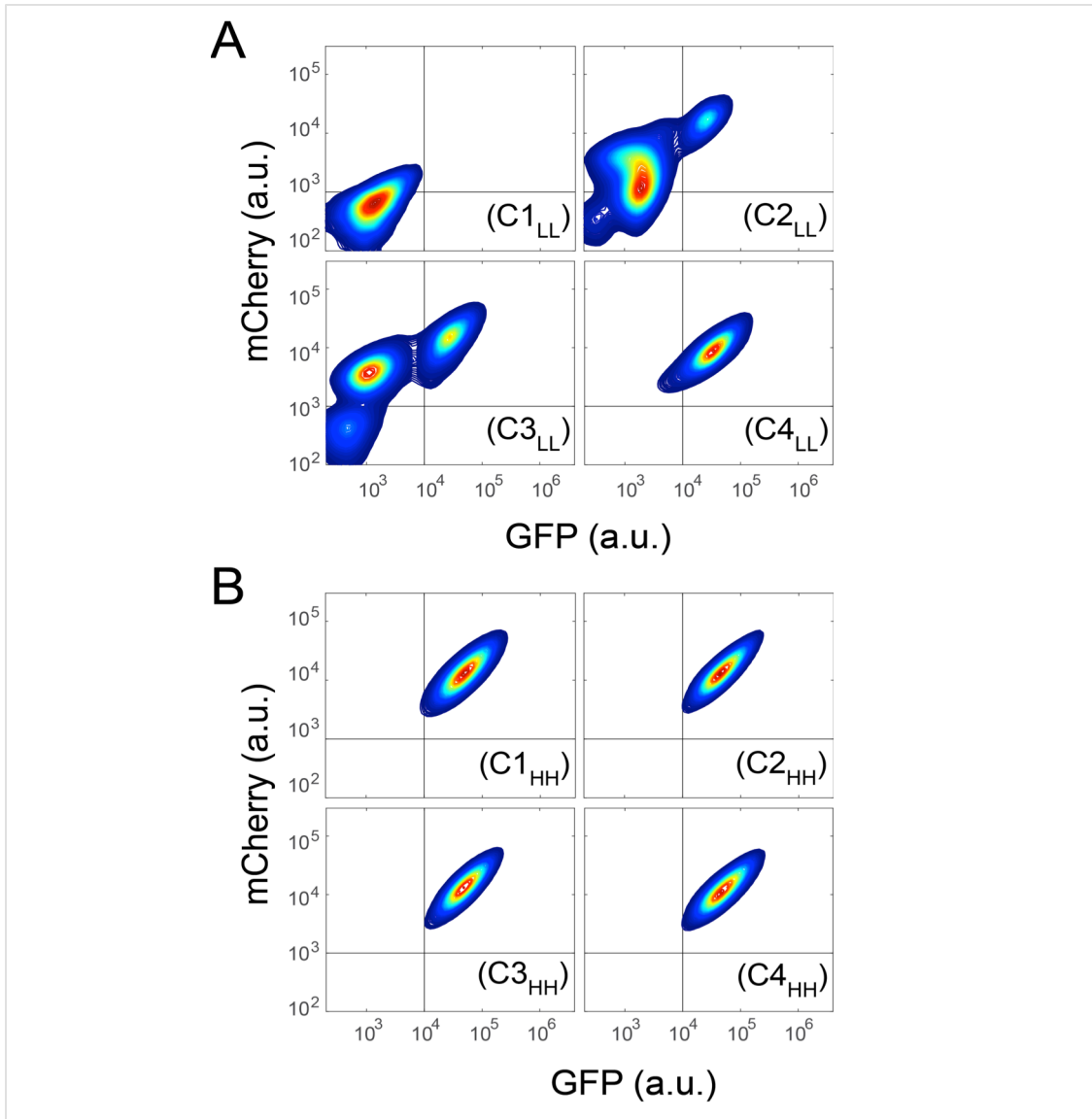


Figure 18: Hysteresis of MINPA with induction of Arabinose and AHL (*AR/AL*). Cells with initial low-low state were induced with a series of concentrations of Arabinose C1_{LL}-C4_{LL}: cells with low-low initial state (A) are induced with *AR/AL* at C1 to C4; C1_{HH}-C4_{HH}: cells with high-high initial state (B) are induced with *AR/AL* for 24 hr at C1 to C4. C1: no inducers; C2: $2.5 \cdot 10^{-6}$ m/v Arabinose and $1 \cdot 10^{-7}$ M AHL; C3: $2.5 \cdot 10^{-5}$ m/v Arabinose and $1 \cdot 10^{-6}$ M AHL; C4: $2.5 \cdot 10^{-3}$ m/v Arabinose and $1 \cdot 10^{-4}$ M AHL. Arabinose and AHL were added at the same time to induce the cells. 100,000 cells were recorded for each sample by flow cytometry.

(Figure 19), which is calculated as the negative logarithmic function of stationary distribution density in the phase space of GFP and mCherry. Using the weighted

ensemble random walk algorithm^{137,138}, the stationary density distribution can be efficiently calculated from the initial uniform distribution. It can be seen that when there is no inducer, MINPA is already quadrastable with four local minima, which is consistent with bifurcation analysis for C1 condition. Furthermore, the much stronger stability of the low-low state (deepest well, Top landscape) and high state-transition barrier explain homogeneous low-low population (C1 experiment in Figure 18A) when cells were initialized with no inductions.

Since Arabinose and AHL combination is not sufficient to enable the cells to reach all four SSS, we chose to add aTc to the mix to further facilitate cell transitions among these four SSS. Using our expanded model, we simulated simultaneous and sequential inductions and computed corresponding quasi-potential landscape (Figure 19), showing cells harboring the same MINPA network exhibiting distinct landscapes under different inductions. AHL and aTc promote a more stable GFP state (Left center), while Arabinose induction modulates the landscape to be biased toward mCherry state (Right center). When the three inducers were applied simultaneously, the landscape changes and the four states show comparable stabilities (Bottom), suggesting a higher possibility of quadramodal cell population experimentally. Experimental validation is shown as flow cytometry measurements of cells treated with Arabinose, AHL, and aTc simultaneously for 24 hours (Figure 20A and Figure 21A). Such a hybrid induction greatly facilitates the cells' transition from low-low state to the other three states so that a quadramodal distribution emerges. Single-cell time lapse microscopy results also showed that the initial low-low state cells could differentiate into GFP, mCherry and high-high state cells (Figure 21B-D). This also finally verifies predicted quadrastability of MINPA. It is

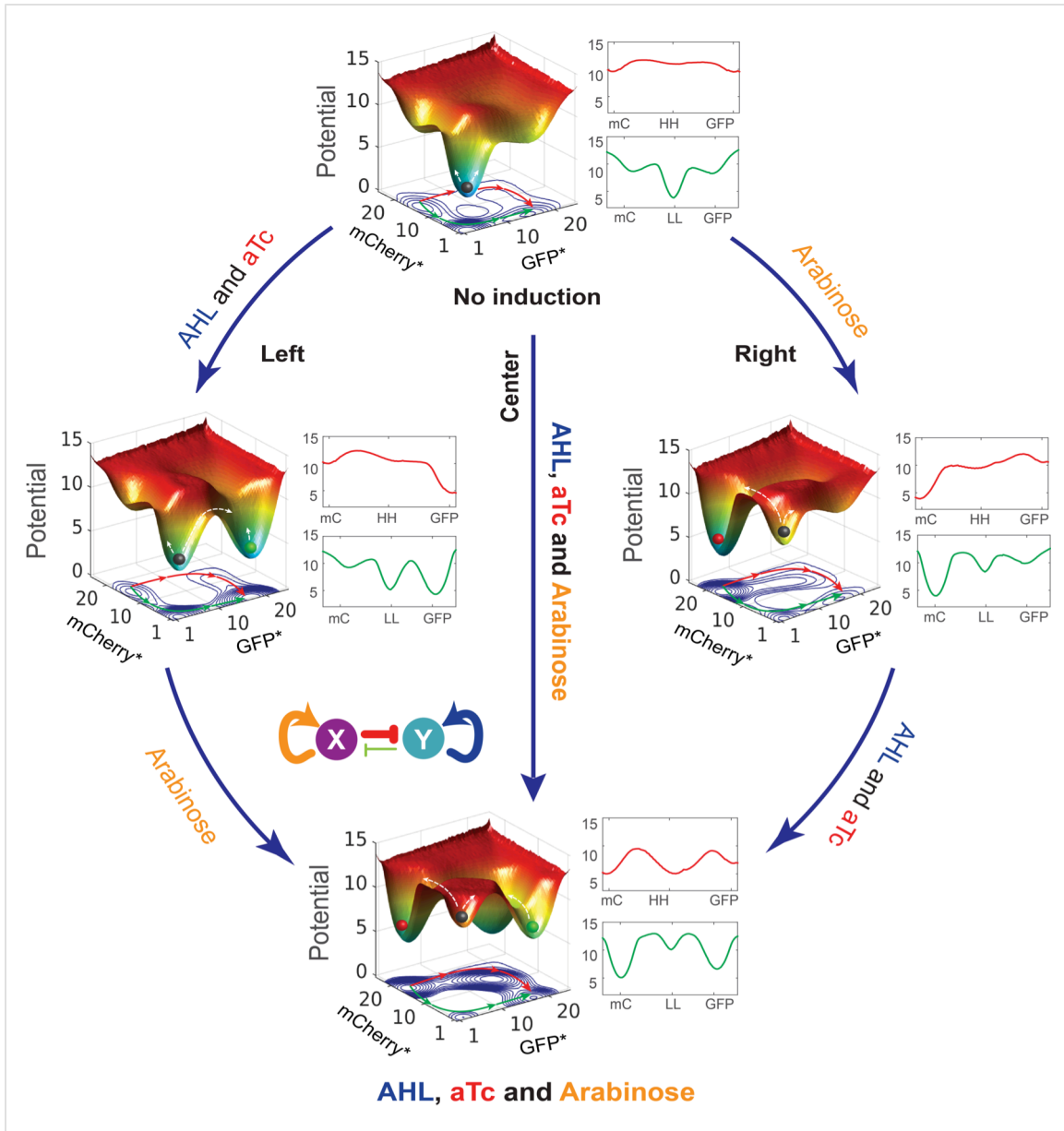


Figure 19: Model-guided quadrastability of MINPA through triple induction. Dynamic evolution of computed energy landscapes of MINPA under sequential/simultaneous inductions of Arabinose, and/or AHL and aTc. Center route: simultaneous induction with three inducers; Left route: sequential induction with AHL and aTc first, and then Arabinose. Right route: sequential induction with Arabinose, and then AHL and aTc. Deeper wells represent higher stability of corresponding states. For each three-dimensional landscape, corresponding two-dimensional state-potential plots were also shown. Red line sketches the potentials from mCherry state to high-high to GFP state while green one represents the potentials from mCherry state to low-low to GFP states. mC: mCherry; HH: high-high; LL: low-low. GFP* and mCherry* is the computed GFP and mCherry abundance from the model. To indicate the effects of inducers, we used the same color for applied inducers and its regulated connections, which were also shown in bolder lines.

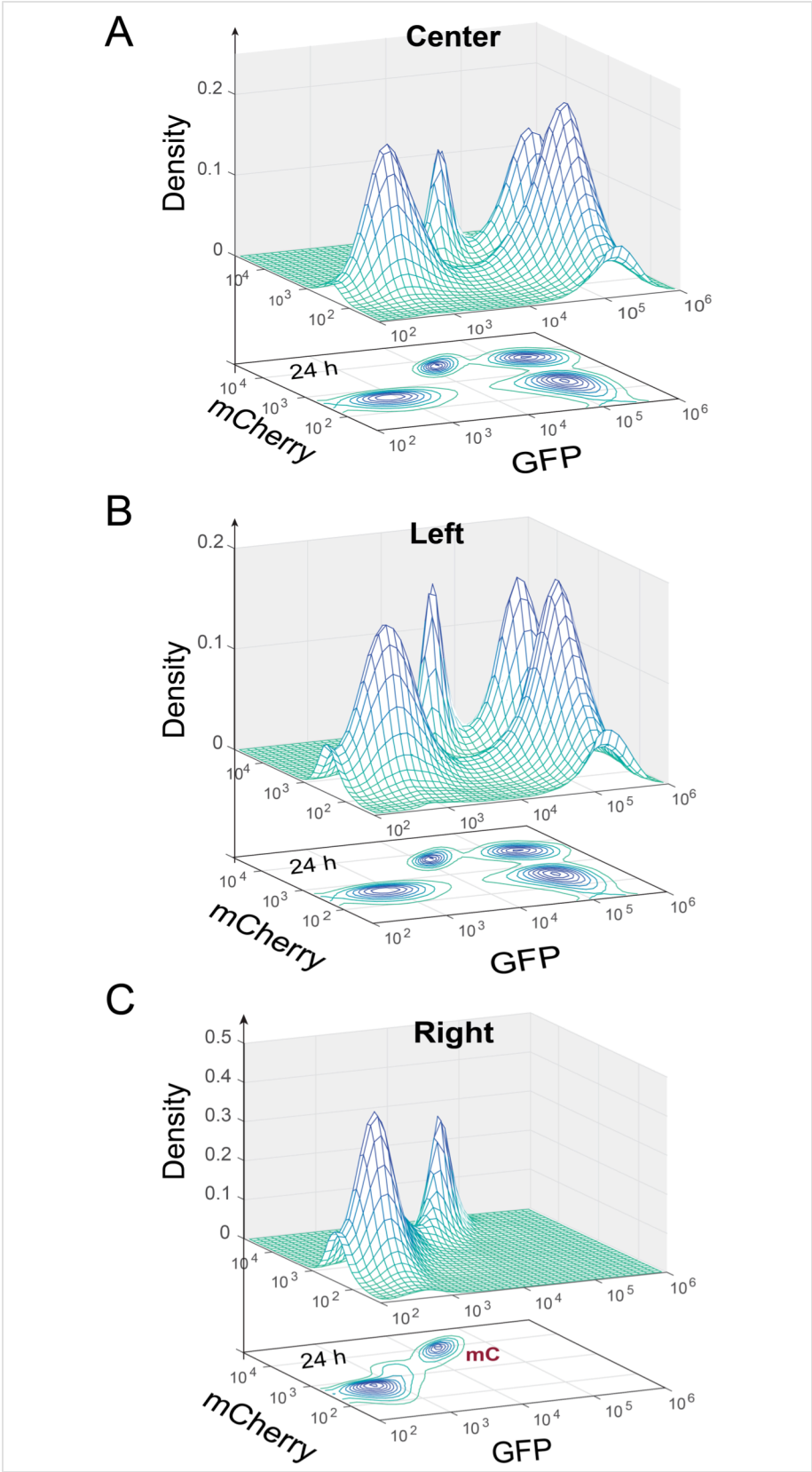


Figure 20: Experimental validations of model-predicted quadrastability using flow cytometry. Quadrastable steady states were observed when Arabinose, AHL and aTc were simultaneously added into the media (A, corresponding to the Center route in Figure 19). Four populations were also observed when AHL and aTc were first added to growth media for 6.5 hr and then Arabinose was added, and cells were grown for another 24 hours before measurement (B, corresponding to the Left route in Figure 19). Bimodality (low-low and mCherry states) was generated when Arabinose was first applied and then AHL and aTc were added (C, corresponding to the Right route in Figure 19). Concentrations for Arabinose, AHL and aTc are $2.5 \cdot 10^{-5}$ m/v, $1 \cdot 10^{-4}$ M, and 400 ng/ml, respectively. Representative results from three replicates are showed and 100,000 cells were recorded for each sample by flow cytometry.

possible that the multimodal distribution might be resulted from the different growth rates of each population under inductions. We then measured chemical inducers' effect on cell growth. As shown in Figure 22, aTc addition (aTc, AHL + aTc, and Ara + AHL + aTc) influenced cells growth and increased the lag phase for about 2.5 hours, compared to AHL or Arabinose individually. But at about 13 hours, the growth rates are almost the same. Since the timescale in the experiments is much longer (≥ 24 hr) than 13 hours, we reason the effect of inducers on the cellular growth rates would not change interpretations of experimental observations.

There are two other strategies to reach this condition: sequential inductions with AHL-and-aTc and then Arabinose (Figure 19, Left route) or Arabinose and then AHL-and-aTc (Right route). Even though the initial and final landscapes are the same, the dynamics for each route are quite different, which could lead to distinct outcomes. By comparing state barrier heights (Figure 19), we hypothesize that cells walking through the left route would start transitioning from low-low state to GFP state upon induction of AHL and aTc. Following Arabinose induction would then make the mCherry state

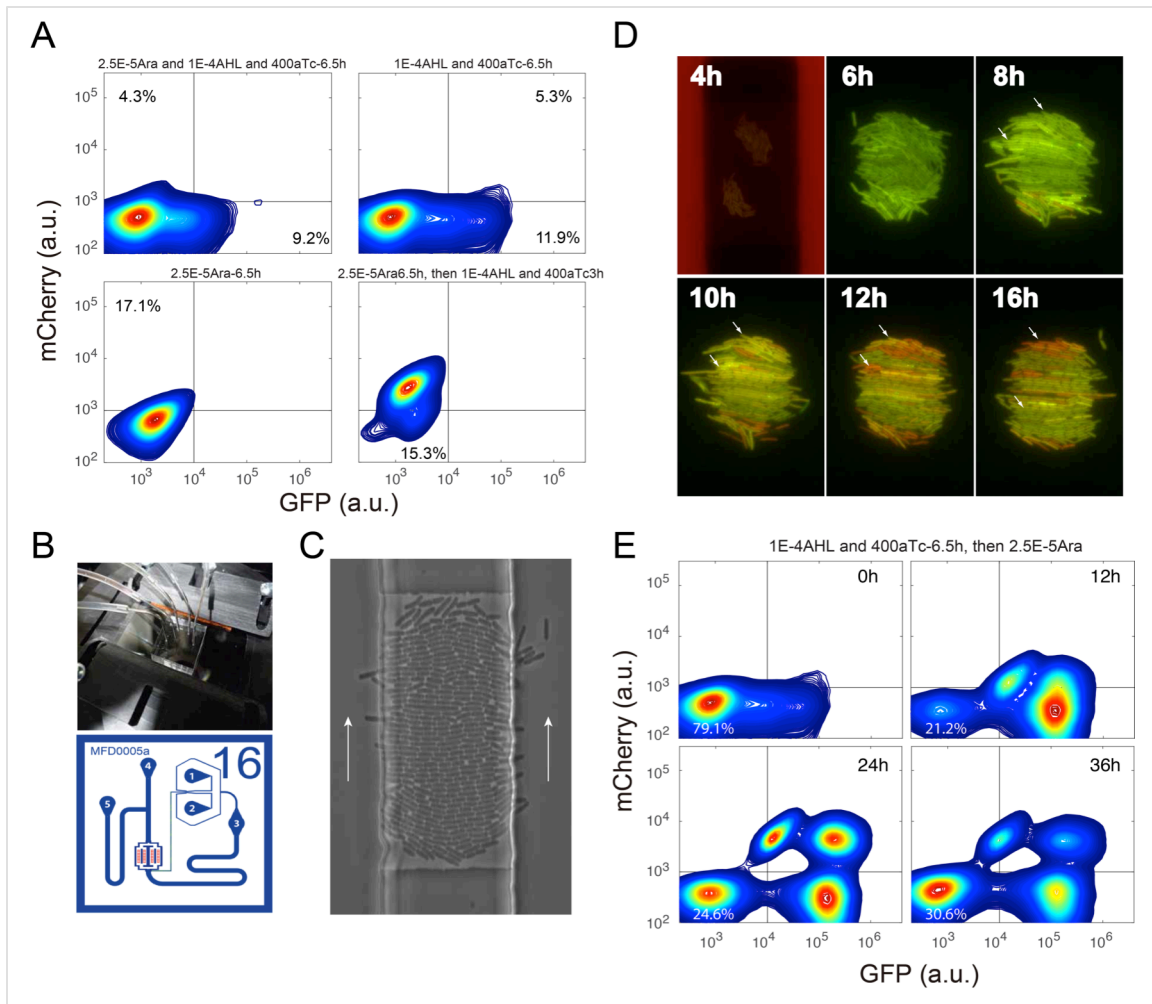


Figure 21: Cells' states under induction with the first inducer, microfluidic results to demonstrate quadrastability with IPTG and aTc induction, and time course of sequential induction of AHL, aTc and Ara. (A) Up left: Flow cytometry result for cells simultaneously induced with 2.5×10^{-5} m/v Arabinose, 1×10^{-4} M AHL and 200 ng/ml aTc for 6.5 hr. Up right: Cells were first induced with 1×10^{-4} M and 400 ng/ml aTc for 6.5 hr, and then measured by flow cytometry. About 12% cells were moving from low-low state to GFP state at 6.5 hr. Bottom left: Cells were first induced with 2.5×10^{-5} m/v Arabinose and no obvious state transition was observed at 6.5 hr. However, at 9.5 hr, most cells (84.6%) were transitioned to mCherry state (Bottom right). (B) Microfluidic setup and device design (adopted from Dr. Hasty lab). (C) Images showing *E. coli* growing in the device. White arrows indicate the flow direction. (D) Time course of the cells growing and fluorescence state change with 2×10^{-4} M IPTG and 200 ng/ml aTc induction in the trap. The red flow is medium without inducer for 6 hr, and then cells switch to medium with inducers for 18 hr. Small white arrows show single cells with state change from GFP to mCherry. Magnification: 40x. (E) Time-course sequential induction with AHL, aTc first and then Arabinose (corresponding to the Left route in Figure 4A). The indicated time point is the time after Arabinose added into the culture. 10,000 events were recorded. The low-low state cells changed from 21.2% (12 hr) to 24.6% (24 hr) to 30.6% (36 hr).

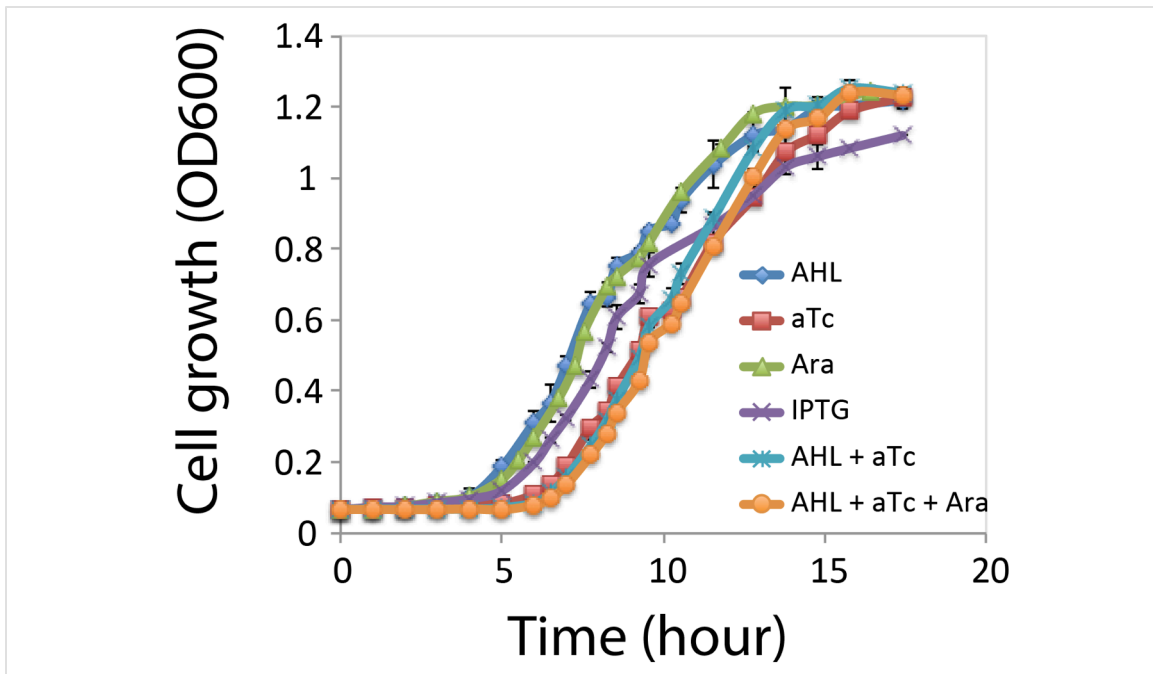


Figure 22: Cell growth rates under each inducer and inducer combinations. Growth curves for the cells under the four individual inducers: Ara; AHL; aTc; IPTG, and inducer combinations: AHL and aTc; AHL, aTc, and Ara. Ara: 2.5×10^{-5} m/v; AHL: 1×10^{-5} M, aTc: 200 ng/ml, IPTG: 1×10^{-4} M. Cells under induction with aTc has a longer lag phase (~ 2.5 hr), and all the samples reached stationary phase after ~ 13 hr. Data indicate mean \pm SD of three independent replicates.

accessible. So some cells with GFP state would transition to high-high state while some low-low state cells transition to mCherry state, resulting in cells in all four states.

Experimental testing indeed shows four stable populations (Figure 20B). At 6.5 hr of AHL and aTc induction, about 12% cells were moving to GFP state while the rest of them still stay “undecided” at low-low state (Figure 21A and 21E). This is consistent with the simulated landscape as these two states are more stable and accessible to each other (Figure 19, Left). Arabinose induction promoted some cells to transition into mCherry state while some cells continued moving into GFP state, of which some further transitioned to high-high state.

Interestingly, the right route is predicted to generate different results. When first induced with Arabinose, the mCherry valley is so deep that it would be difficult for cells to jump out to high-high state, and low-low state cells also hardly transition to GFP state due to its low attractiveness, and thus most cells would stay at mCherry and low-low state even with AHL and aTc inductions (Figure 19, Right). Experimental testing of the right route indeed only produces two populations with low-low and mCherry state (Figure 20C). With 5 hr of Arabinose induction, most cells still stay at low-low state because of slow transition to the mCherry state (Figure 21A), but 84.6% cells transitioned to mCherry state with 15.3% cells at low-low state at 9.5 hr (Figure 21A). This is consistent with our model predictions. The high barrier between the mCherry state and high-high state blocks the transition from mCherry state to high-high state, while the low attractiveness and relatively high barrier of the GFP state also decreases the probability of cells transitioning from low-low to GFP state. Hence, when AHL and aTc are applied, cells are predominantly in mCherry state with a small portion in low-low state with low probability of transitioning out, resulting in a bimodal distribution.

3.3 Discussion

Multistability and the resulting landscape has long been proposed as an underlying mechanism that cells use to maintain pluripotency and guide differentiation^{30,31,119,139,140}. Theoretical frameworks have also been established to quantify the Waddington landscape and biological paths for cell development^{117,122,141}. Experimental validation of this hypothesis and a full understanding of this mechanism will help reveal differentiation dynamics and routes for all cell types, which remains an outstanding

problem in biology. In this study, we engineered the quadrastable MINPA circuit and show that it can guide cell fate choices, represented by fluorescence expression, through shaping the potential landscape. MINPA represents one of the most complicated two-node network topologies and includes four genes to implement a web of regulations. Biological complexity correlates with the number of regulatory connections¹⁴², not the number of genes. Hence, dense connectivity and complex dynamics of MINPA may provide a framework to understand similarly densely connected gene regulatory networks.

Combining mathematical modeling and experimental investigation, this study serves as a proof-of-principle demonstration of the Waddington landscape. Furthermore, we used this circuit to demonstrate how different sequential inductions can change the landscape in a specific order and navigate cells to different final states. Such illustrations suggest mechanistic explanations of the need for fixed induction sequences for targeted differentiation to desired cell lineage. Overall, this study helps reveal fundamental mechanisms of cell-fate determination and provide a theoretical foundation for systematic understanding of the cell differentiation process, which will lead to development of new strategies to program cell fate.

3.4 Materials and Experimental Methods

3.4.1 Strains, Media, and Chemicals

All the molecular cloning experiments were performed in *E. coli* DH10B (Invitrogen, USA), and measurements of MINPA and sub-networks were conducted in *E. coli* K-12 MG1655 Δ *lacI* Δ *araCBAD* strain as previously described (from Dr. Collins Lab¹⁰⁷). The sequential induction for the toggle circuit was conducted in *E. coli*

MG1655 Δ *lacI* strain as previously described¹⁰⁷. Cells were grown at 37 °C in liquid and/or solid Luria-Bertani broth medium with 100 µg/mL ampicillin or kanamycin. Chemicals AHL (3oxo-C6-HSL, Sigma-Aldrich), Arabinose (Sigma-Aldrich, USA), isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich), and anhydrotetracycline (aTc, Sigma-Aldrich) were dissolved in ddH₂O and diluted into indicated working concentrations. Chemical aTc solution was stocked in brown vials, and experiments involving aTc were performed in cabinet without light, and cell cultures were grown in darken incubator at 37 °C. Cultures were shaken in 5 mL and/or 15 mL tubes at 220 rotations per minute (r.p.m).

3.4.2 Plasmids construction

All the plasmids (MINPA and its nine sub-networks) in this study were constructed using standard molecular cloning protocols and assembled by standardized BioBricks methods based on primary modules (Table 5) from the iGEM Registry (www.parts.igem.org). Hybrid promoter *Para/lac* was from Dr. Collins lab and amplified using forward primer: *CGGAATTCGCTTCTAGAGAATTGTGAGCGGATAAC*; and reverse primer: *CGCTGCAGGCACTAGTTTGTGTGAAATTGTTATCCG*. PCR product was purified using GenElute™ PCR Clean-Up Kit (Sigma-Aldrich), and then cut by restriction enzymes *EcoRI* and *PstI*. The purified product was inserted into pSB1K3 backbone, and finally verified by DNA sequencing. The MINPA circuit was constructed from promoter *Para/lac* and 9 other Biobrick standard biological parts: BBa_B0034 (ribosome binding site, RBS), BBa_C0080 (*AraC* gene), BBa_C0040 (*tetR* gene), BBa_K176000 (*PLux/tet* hybrid promoter), BBa_C0062 (*luxR* gene), BBa_C0012 (*lacI*

gene), BBa_B0015 (transcriptional terminator), BBa_E0240 (GFP generator), and BBa_J06702 (mCherry generator). The fragment and vector were separated by gel electrophoresis (1% TAE agarose) and purified using GenElute Gel Extraction Kit (Sigma-Aldrich). Then, fragment and vector were ligated together using T4 DNA ligase, and the ligation products were transformed into *E. coli* DH10B and clones were screened by plating on 100 µg/ml ampicillin LB agar plates. Finally, their plasmids were extracted and verified by double digestion (*EcoRI* and *PstI*). The detailed procedures of assembling by plating on 100 µg/ml ampicillin LB agar plates. Finally, their plasmids were extracted and verified by double digestion (*EcoRI* and *PstI*). The detailed procedures of assembling DNA constructs were described in our previous study ³⁹. Restriction enzymes (*EcoRI*, *XbaI*, *SpeI*, and *PstI*) and T4 DNA ligase were purchased from New England Biolabs. All the constructs were inserted into high copy number plasmid pSB1A3 and pSB1K3. All the constructs were verified by DNA sequencing (Biodesign sequencing lab in ASU) step by step.

3.4.3 Flow cytometry

All the samples were analyzed at the indicated time points on an Accuri C6 flow cytometer (Becton Dickinson, USA) with excitation/emission filters (488/530 nm for GFP, and 610 LP for mCherry). The data were collected in a linear scale and non-cellular low-scatter noise was removed by thresholding. All measurements of gene expression were obtained from at least three independent experiments. For each culture, 100,000 events were collected at a slow flow rate. Data files were analyzed using MATLAB (MathWorks).

Table 5: Components from the Registry of standard biological parts.

Biobrick number	Abbreviation in the paper	Description
BBa_C0080	AraC	AraC arabinose operon regulatory protein from <i>E. coli</i>
BBa_C0040	TetR	Tetracycline repressor from transposon Tn10
BBa_C0062	LuxR	LuxR activator from <i>Aliivibrio fischeri</i>
BBa_C0012	LacI	LacI repressor from <i>E. coli</i>
BBa_E0240	GFP	GFP generator
BBa_J06702	mCherry	RFP generator
BBa_K176002	<i>PLux_{tet}</i>	Hybrid promoter with LuxR/HSL- and TetR-binding sites
BBa_B0034	RBS	Ribosome binding site
BBa_B0015	Terminator	Transcriptional terminator (double)
BBa_K176009	CP	Constitutive promoter
pSB1K3	pSB1K3	High copy BioBrick assembly plasmid with kanamycin resistance
pSB1A3	pSB1A3	High copy BioBrick assembly plasmid with ampicillin resistance

3.4.4 Sequential induction and hysteresis

For sequential induction, initially uninduced overnight cell culture was diluted into fresh media without or with inducer I, grown at 37 °C and 220 r.p.m till OD₆₀₀ is 0.15~0.25 (the time usually takes 5~6.5 hr, depends on the inducers and concentrations). For samples induced individually by Ara, or AHL, or IPTG, it is ~5 hr; for samples induced with aTc, it takes ~6.5 hr. According to our experience, gene (GFP) is starting to

be partially expressed while steady states are not yet stable. Then inducer II was added into the culture, and grown for another 24 hr. Flow cytometry was performed at 0, 12, and 24 hr after the second inducer was added into the culture. For each set of sequential induction, the 1st scenario: add inducer I first, then add inducer II; the 2nd scenario: add inducer II first, then add inducer I; the 3rd scenario: add inducers I and II at the same time. As a control, cells without any inducer were also prepared and measured. Inducer I and II were the two of four commercial chemicals: AHL, Arabinose, IPTG, and aTc. All the experiments were repeated for at least three times and only representative results were showed.

For hysteresis experiments, initially uninduced cells were diluted into fresh media and distributed into new 5 ml tubes. Various amounts of Arabinose and AHL (3oxo-C6-HSL) were added into the media, and cells were then grown at 37 °C shaker. The initially high-high state cells induced with 2.5×10^{-3} m/v Arabinose and 1×10^{-4} M AHL were collected with low-speed centrifugation, washed twice, resuspended with fresh medium, and at last inoculated into fresh medium at a 1:100 ratio with the same series of inducer (Arabinose and AHL) concentrations. C1, C2, C3, and C4 (Figure 17B and Figure 18) are four increasing concentrations of Arabinose and AHL used for experimental probing, but the ratio of Arabinose and AHL is fixed. Specifically, cells were induced with the Arabinose and AHL at the same time (the 3rd scenario), at concentrations from C1 to C4. C1: no inducers; C2: 2.5×10^{-6} m/v Arabinose and 1×10^{-7} M AHL; C3: 2.5×10^{-5} m/v Arabinose and 1×10^{-6} M AHL; C4: 2.5×10^{-3} m/v Arabinose and 1×10^{-4} M AHL. Flow cytometry analyses were performed at 12 hr and 24 hr to monitor the fluorescence levels. Experiments were repeated two times with three replicates.

3.4.5 Cell growth array under inductions

Cells with the MINPA plasmid were cultured overnight at 37 °C and diluted into fresh media with corresponding inducers at 1:100 ratio (O.D. \sim 0.066). The four individual inducers are Ara (2.5×10^{-5} m/v), AHL (1×10^{-5} M), aTc (200 ng/ml), IPTG (1×10^{-4} M), and inducer combinations: AHL and aTc, AHL, aTc, and Ara. Cellular growth rates were measured by using 200 μ L cultures in a 96-well plate with absorbance at 600 nm on a plate reader (BioTek, USA). Three replicates were tested for each condition.

Compared to AHL or Arabinose individually, aTc addition (aTc, AHL + aTc, and Ara + AHL + aTc) influenced cells growth and increased the lag phase for about 2.5 hours. But at about 13 hours, the growth rates are almost the same. Since the timescale in our experiments is much longer (\geq 24 hr) than 13 hours, we reason the effect of inducers on the cellular growth rates would not change interpretations of experimental observations.

3.4.6 Microfluidics and microscopy

Microfluidics coupled with time-lapse imaging was employed to visualize the state transitions at the single-cell level. Cells with MINPA circuit were grown overnight, which was then re-diluted into 5 mL fresh LB medium with Kanamycin the next day. When OD₆₀₀ of the cells reached about 0.2, cells were spun down with low speed and resuspended in 5 ml of fresh medium and loaded into the device. Detailed description of chip design and device setup could be found from Hasty Lab ⁹⁸. Two media were

prepared: one with inducers and the other without. Cells in the trap were first supplied by the medium without inducer for 6 hours, and then switched to medium with inducers for another 18 hours, which was controlled by adjusting the heights of the medium syringes relative to one another. Images were taken by using Nikon Eclipse Ti inverted microscope (Nikon, Japan) equipped with an LED-based Lumencor SOLA SE. Phase and fluorescence images were taken every 5 minutes for 24 hours in total under the magnification 40x. Perfect focus was maintained automatically using Nikon Elements software.

Specifically, two media were prepared: one with inducers (Arabinose, AHL and aTc) and the other without. After cells were loaded into the trap (1 to 5 cells for a trap), the device was heated up to 37 degree and cells were supplied with LB media without inducers for 6 hours. Sulforhodamine was added as a dye to monitor nutrient transport. Then, the supplied media was switched to the media added with Arabinose, AHL and aTc for another 18 hours. Media switching was controlled by adjusting the heights of the medium syringes relative to one another.

However, cells treated with the three inducers demonstrate symptoms of significant stress and cell death, presumably due to photo toxicity compounded with flow-induced sheer stress and other mechanical stresses in the microenvironment^{143–145}. Lower concentrations of three inducer combinations were also tested but yield no significant improvement of cells viability.

Since the logic of emergence of quadrastability is enhancing the two positive feedbacks of MINPA through adding inducers Arabinose, AHL and aTc, quadrastability could also be achieved through weakening the mutual inhibition using IPTG and aTc.

Depending on the basal expression of two hybrid promoters, IPTG and aTc can promote GFP and mCherry expression to a limited extent, which in turn attenuates fluorescent proteins toxicity. Hence, we tried to use IPTG and aTc to induce quadrastability instead of the three inducers tried in flow cytometry. Experimental result showed that the initial low-low state cells could differentiate into GFP, mCherry and high-high state cells with 2×10^{-4} M IPTG and 200 ng/ml aTc induction (Figure 21D). It is interesting that the trajectory for many cells were from GFP to high-high to mCherry state. Altogether, this result further verified MINPA has the potential to generate quadrastability in living cells.

3.5 Mathematical Modeling

Ordinary differential equations are developed based on the MINPA structure. Details can be found in our paper¹³⁷.

CHAPTER 4

DEVELOPMENT OF A PROTEIN EXPRESSION PREDICTOR TO TUNE GENE EXPRESSION AND CIRCUIT DYNAMICS

4.1 Introduction

Gene circuit engineering as one of the foundation technologies has helped start the burgeoning development of synthetic biology. Based on a large collection of well-characterized biological components including promoters, ribosome binding sites, transcriptional factors, terminators, and RNA elements, complex gene circuits with designed functions can be wired using established biological principles. Toggle switch and repressilator are two of the earliest examples of engineered gene circuits^{6,7}. Now synthetic biologists are paying increasing attention to develop innovative gene circuits for drug development^{146,147}, gene therapy¹⁴⁸⁻¹⁵⁰, cancer therapy¹⁵¹, immunotherapy^{152,153}, pathogen detection^{58,154}, *in vivo* delivery¹⁵⁵, and other biotechnological applications including nitrogen fixation^{156,157} and environmental bioremediation^{158,159}.

Currently circuit assembly has two main strategies: one is monocistronic construct, in which one promoter drives one gene expression and ensures each gene being expressed independently; the other is polycistronic construct, in which one promoter transcribes multiple structural genes (operon) into a single messenger RNA but is translated into individual products (Figure 23A). Operon, a cluster of genes with functional associations, is a major genomic organization method in prokaryotic cells and also widely found in eukaryotes and viruses¹⁶⁰. This operon organization strategy ensures a coordinated gene

expression and regulation, and enables bacteria cells to rapidly respond to environmental changes. In synthetic biology, this organization (synthetic operon) facilitates rapid constructions of genetic cascades and decreases the number of wired biological components, and therefore is widely used in circuit engineering^{46,107,161–165}.

However, it remains unknown whether/how structural gene expression is impacted by immediately adjacent genes in a polycistronic operon. Two previous reports have indicated that gene position and transcriptional distance can affect gene expression in an operon^{166,167}. But little research has systematically studied the effects of adjacent gene organizations in synthetic operons on circuit gene expression, dynamics, and functionality. This factor is more prominent for synthetic operons containing a cluster of genes and complex multi-layered genetic circuits. Deciphering the effects of adjacent transcriptional region's (ATRs) regulation on gene expression would advance our understanding of determinants of gene expression in synthetic circuits, and accelerate circuit design and assembly. Such adjacent gene regulation effect has been generally neglected during engineering of synthetic gene networks, leading to unexpected circuit performance or failure. Hence, development of a predictive method to evaluate each gene's expression level in a circuit would be of great importance to circumvent the need for trial and error in circuit design and assembly.

To quantify ATRs effects on gene expression, we herein systematically analyzed the effect of adjacent genes and non-coding regions on green fluorescence protein (GFP) expression level through construction of ~120 synthetic gene circuits (operons) in *Escherichia coli*. Data-driven analysis yields a new gene expression-prediction method that strongly correlates with ATRs features including GC content, size, and stability of

mRNA folding near ribosomal binding site (RBS). We demonstrated this method's utility in evaluating genes' relative expression levels by incorporating it in the design and construction of logic gates with lower basal expression and higher sensitivity and nonlinearity. Furthermore, we designed synthetic 5'ATRs to tune protein expression levels over a 300-fold range. Finally, by combining ATR regulation and mathematical modeling, we illustrated the application of synthetic ATRs in quantitatively tuning nonlinear dynamics of bistable gene networks.

4.2 Results

4.2.1 Protein expression is significantly influenced by its adjacent genes and position in the operon

To examine whether protein expression is affected by its neighbors in a polycistronic setting, we first constructed a two-gene operon (gene *X* and *GFP*), which is driven by a constitutive promoter (Figure 23B). Flow cytometry results showed that for different *X*, GFP expression is highly diverse. Specifically, circuits with *AraC* and *RhIR* as *X* showed a comparable level of GFP fluorescence to the control (without *X* gene), while the others (*LuxI*, *TetR*, and *dnMyD88*) showed high expression variations, ranging from 6-fold to over 120-fold decrease compared to control. It is interesting to note that membrane protein *dnMyD88* shows the most significant influence on its neighbor GFP expression (Figure 23B). RT-qPCR results showed that there is little correlation between GFP fluorescence and mRNA concentration, for P1:P2 (GFP N terminal) or P3:P4 (GFP C terminal) primer pairs, (Figure 23C and Figure 24). So the variation of mRNA concentrations for each construct is insufficient to explain the fluorescence differences,

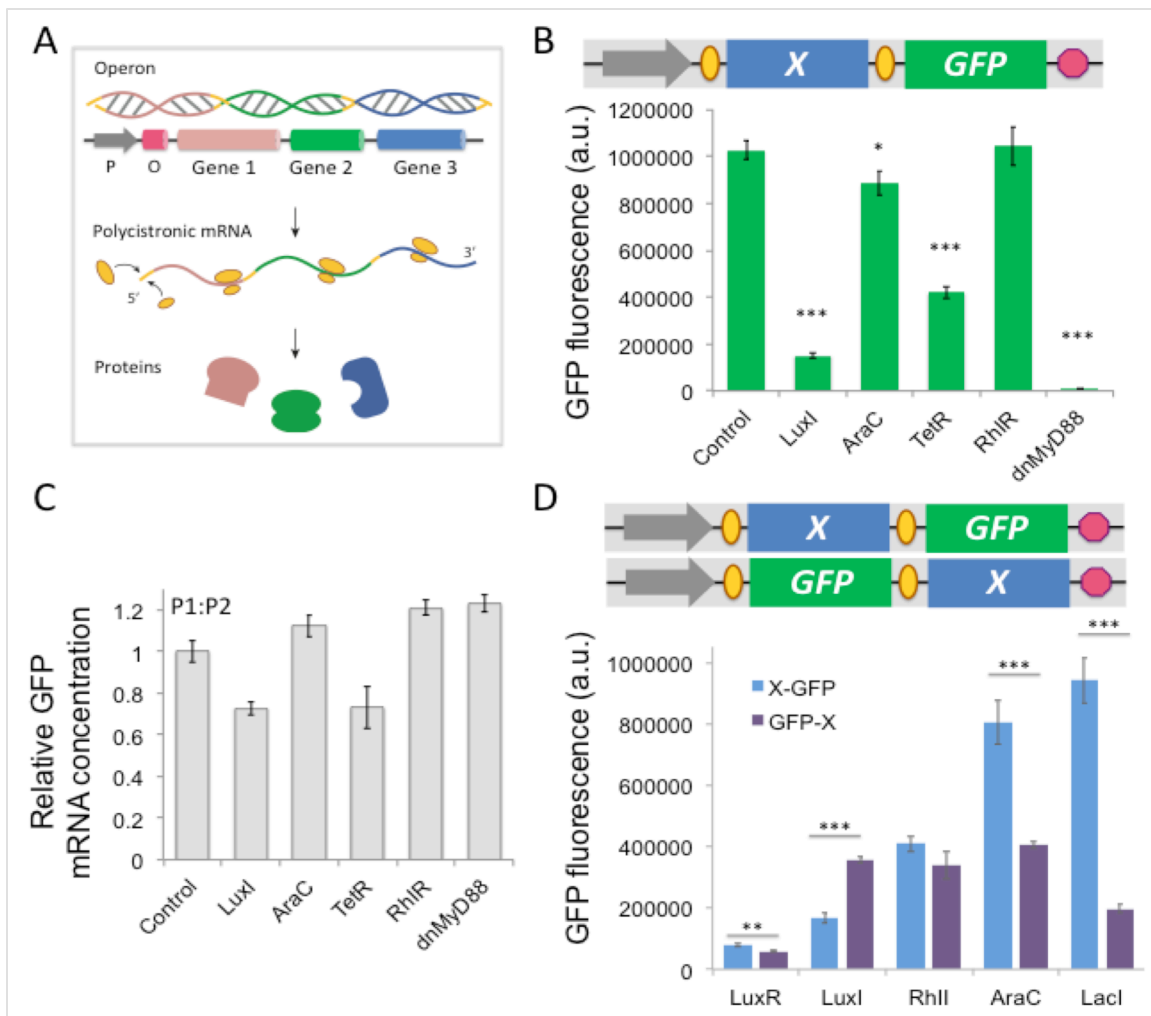


Figure 23: Protein expression is significantly influenced by its adjacent genes and position in synthetic operons. (A) Illustration of the operon structure and genes expression. The three structural genes are transcribed as a polycistronic mRNA, but translated into individual proteins. P: promoter; O: operator. Yellow oval: ribosome; (B) Top: Schematic representation of synthetic bi-cistronic gene circuits with gene *X* and *GFP*. Grey arrow: constitutive promoter; Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. Bottom: Flow cytometry results show GFP expression is influenced by its 5' ATRs. *X* represents a gene name (i.e. *LuxI*, *AraC*, *TetR*, *RhlR*, and *dnMyD88*). “Control” is without *X* gene in the circuit. Grey arrow represents constitutive promoter; yellow oval represents ribosome-binding site; red octagon represents transcriptional terminator. Rectangles with filled colors represent different genes. *X* represents different gene names used in the circuit. Data represent the mean \pm s.e. of eight replicates. (C) Relative GFP mRNA concentrations (normalized to 16S rRNA control) for the circuits in (B) determined by RT-qPCR. Primer pair P1:P2 was designed to amplify *GFP* gene from the sample cDNA. (D) Top: Schematic representation of synthetic bi-cistronic gene circuits with gene *X* and *GFP*, but with switched positions in the circuit. Gene position in the operon impacts GFP expression. Data represent the mean \pm s.e. of eight replicates. * $P < 0.05$, ** $P < 0.001$ and *** $P < 0.0001$ by student’s t-test.

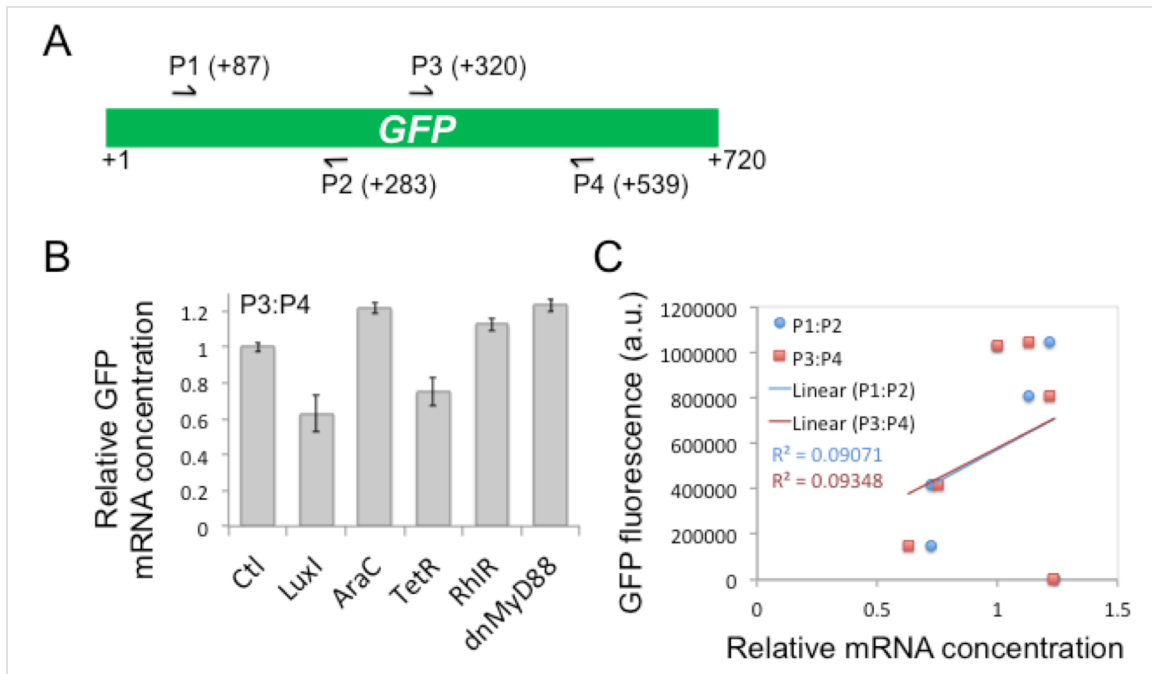


Figure 24: RT-qPCR result indicates there is little correlation between GFP protein fluorescence intensity and mRNA level for the circuits in Figure 23B. (A) Two pairs of primers (P1:P2, and P3:P4) designed to amplify *GFP* gene from the sample cDNA. The binding sites of the four primers are also indicated. (B) RT-qPCR result using primer pair P3:P4 to amplify *GFP* gene. The GFP mRNA concentrations were normalized to the 16S rRNA control. Error bar represents standard deviation of three biological replicates. (c) Correlation between the GFP fluorescence intensities and the relative GFP mRNA concentrations. Little correlation was found using primer pair P1:P2 or P3:P4.

which is in agreement with previous studies that protein and mRNA copy numbers in *E. coli* cells for any given gene are uncorrelated^{170,172}. These results indicate that a gene's expression in an operon-based gene circuit is substantially regulated by its neighboring genes, mostly at the translational level.

Next, we further investigated the influence of a gene's position on its expression. As shown in Figure 23D, higher GFP expression is observed when GFP is arranged distal to the promoter for the bi-cistronic constructs that *X* gene is *RhlI*, *AraC*, or *LacI*, while there are three cases showing similar level GFP fluorescence (*LuxR*) or higher (*LuxI*)

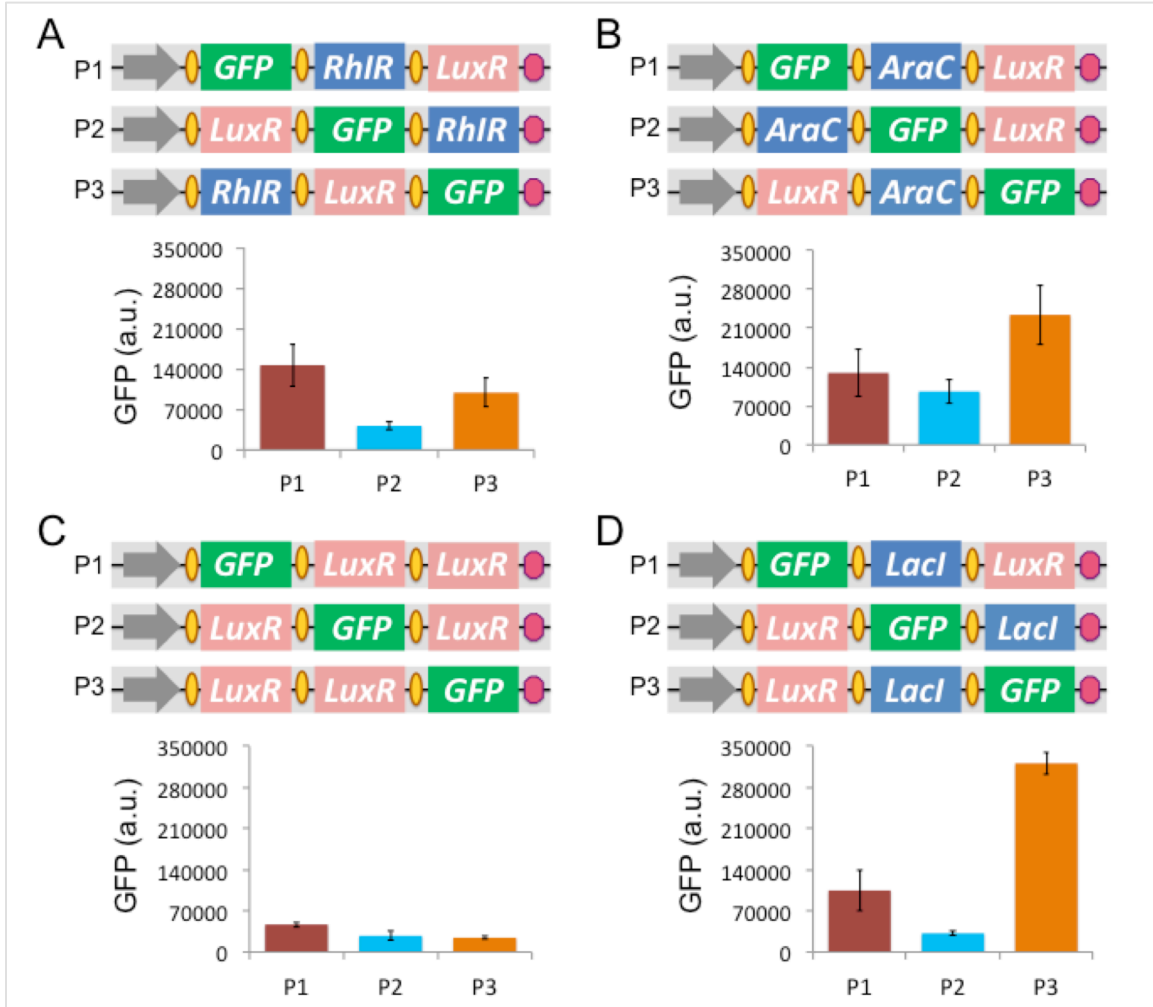


Figure 25: Gene position in the tri-cistronic circuit impacts GFP expression. (A) GFP is arranged at proximal (P1) or middle (P2) or distal (P3) positions to the constitute promoter in the tri-cistronic circuit with two more genes *LuxR* and *RhIR*. Circuit with GFP at P1 position shows the highest GFP expression. (B) GFP is arranged at P1, P2, and P3 positions in the tri-cistronic circuit with genes *LuxR* and *AraC*. Circuit with GFP at P3 position shows the highest GFP expression. (C) GFP is arranged at P1, P2, and P3 positions in the tri-cistronic circuit with two copies of *LuxR* genes. Circuit with GFP at P1 position shows the highest GFP expression. (D) GFP is arranged at P1, P2, P3 positions to the constitute promoter in the tri-cistronic circuit with genes *LuxR* and *Lacl*. Circuit with GFP at P3 position shows the highest GFP expression. Data represent the standard deviation of eight replicates. Gray arrow: constitutive promoter; Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. Rectangles with filled colors represent different genes.

when GFP is arranged right downstream of the promoter. Results from tri-cistronic constructs also indicate that GFP expression is varied for different positions in the circuit and adjacent genes (Figure 25). Moreover, for different X s with the same position, GFP shows substantial variations, consistent with results shown in Figure 23B. Altogether, these results demonstrate that a gene's sequence and position in operons have considerable effects on adjacent genes' expression.

4.2.2 Quantitative characterizations of adjacent gene regulation in synthetic operons

To quantify ATRs impacts on protein expression, we designed and constructed ~80 circuits with different neighbor genes and varying sizes (X and Y) to cover a wide range of GFP gene position and neighbor features (GC content, size, and mRNA secondary structure). These genes are commonly used in synthetic biology, including transcriptional factors, quorum-sensing components, and other functional genes. Membrane-protein coding genes were excluded from consideration due to their complex structures. To ensure experimental consistency, all circuits are constructed using the same constitutive promoter, RBS, terminator, and expression vector.

First, GFP was arranged to the distal end of synthetic operons with two or three genes (Figure 26A). It is found that GFP expression increased with the total 5'ATRs GC content (correlation coefficient $r = 0.66$), while 5'ATR length has a negative effects on GFP expression ($r = 0.53$, Figure 26A). Sliding window analysis of 5'ATR GC content suggested that the GC content of the whole 5'ATR has the highest fitting efficiency (Figure 27A). We speculate that high GC content could increase total mRNA stability, while long transcription process could consume more energy and decrease the probability

of complete GFP transcription/translation. Additionally, previous studies reported that RNA secondary structure near the RBS influences a gene's expression, so local folding energy from -70-nt to +38-nt region around GFP's RBS (GFP's translation starting site is denoted as +1) were calculated. Consistent with previous reports¹⁶⁹⁻¹⁷¹, our analysis also shows that GFP expression is negatively correlated with folding energy around RBS of GFP ($r = 0.5$, Figure 26A). Overall, these three factors together explain 63% GFP variation (Figure 26B).

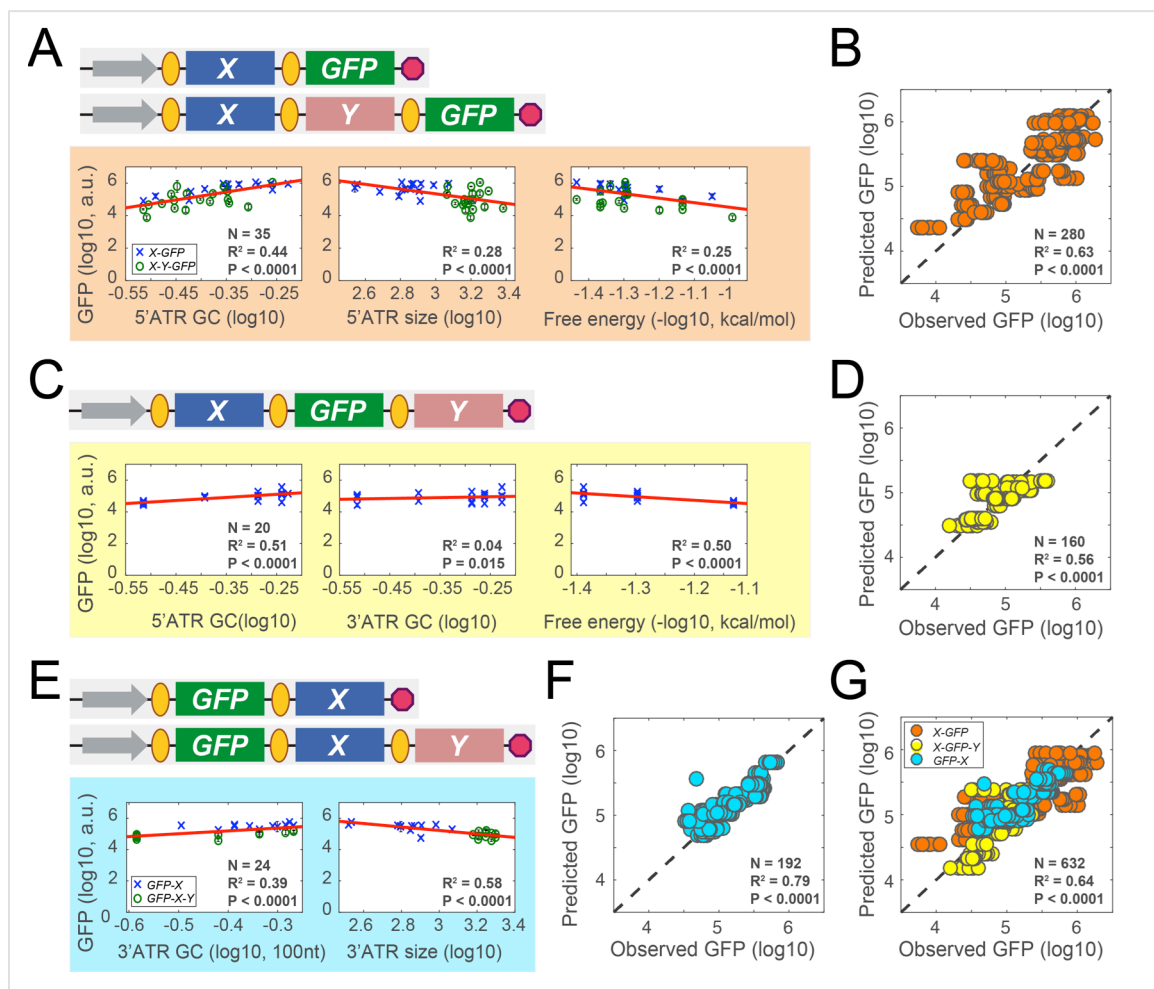


Figure 26: Quantitative characterization of adjacent gene regulation in synthetic operons. (A-B) Scenario 1: GFP is arranged distal to the promoter. Top: Schematic representation of synthetic polycistronic gene circuits X -GFP. Grey arrow: constitutive promoter; Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. X and Y represent different gene names. Bottom: GFP expression is significantly impacted by its

5'ATRs' GC content, size and local folding free energy. 35 genetic circuits with one or two genes placed in front of GFP, which are labeled with different symbols in the regression results. The red lines are the linear regression results from the data. Variables are in log scale. Variable "Free energy" is first transformed to positive values and then changed to log scale. Error bars are s.d. of eight measurements performed in three different days. (B) is the experimentally measured GFP fluorescence compared to the linear regression model predicted GFP expression with the three variables and fitted coefficients. If the model predicted values and experimentally observed values agreed perfectly ($R^2 = 100\%$), all the data points would fall on the dotted diagonal line of the squares. N is the total measurements for the 35 circuits. (C-D) Scenario 2: GFP is placed in the middle of the three-gene operons (*X-GFP-Y*). 20 circuits with different X and Y gene combinations were constructed. (D) is the experimentally measured GFP fluorescence compared to model predicted GFP expression with the three variables in (C). N is the total measurements for the 20 circuits. (E-F) Scenario 3: GFP is placed proximal to promoter (*GFP-X*). 24 circuits with different 3'ATRs (including one or two genes) were constructed, and different symbols are used to indicate bi- or tri-cistronic constructs in the regression results. (F) is the experimentally measured GFP fluorescence compared to model predicted GFP expression with the two variables of 3'ATR in (E). Only 3'ATR data are used to fit the model (without 5'ATR). (G) A combined model analysis for all three scenarios compared to five independent variables including 5'ATR GC content, 5'ATR size, 3'ATR GC content (100-nt), 3'ATR size, and local mRNA folding free energy. The five variables together explain 64% of GFP fluorescence variation. Dots with different colors indicate the data source from the above three scenarios.

Next, GFP was placed in the middle of the operon. In this case, we found that 5' and 3'ATRs GC content and local mRNA folding free energy have the most significant impacts on GFP expression (Figure 26C), and explains more than half of variation in fluorescence variations (Figure 26D).

Finally, circuits with GFP engineered proximally to the promoter were also constructed and investigated to probe the relationship between GFP expression and its 3'ATR. Similarly, results show 3'ATR GC content and size has a positive and negative correlation with GFP fluorescence, respectively (Figure 26E). Sliding window analysis further revealed that the GC content of first 100 nt of 3'ATR has the highest fitting efficiency, suggesting the rear 100 nt is important for GFP expression (Figure 27B).

These two variables of 3' ATR together can account for 79% GFP variations (Figure 26F). It is worth noting that the 5' ATR GC content has a bigger impact on GFP expression than its size, while 3' ATR size is more influential than its GC content (Figure 26A and E). On average, GFP in the middle shows the lowest expression ($\mu = 4.908$, $\sigma^2 = 0.088$) and GFP in the distal shows the highest variation ($\mu = 5.293$, $\sigma^2 = 0.334$; Figure 26B,D,E).

A combined model integrating the above three scenarios was then developed. The five variables (predictors) combined have significant correlations with the GFP expression, and explain nearly two-thirds of the variations in GFP fluorescence (Figure 26G), of which 95.6% can be accounted for by four of the five predictors, namely GC content and sizes of 5' and 3' ATRs.

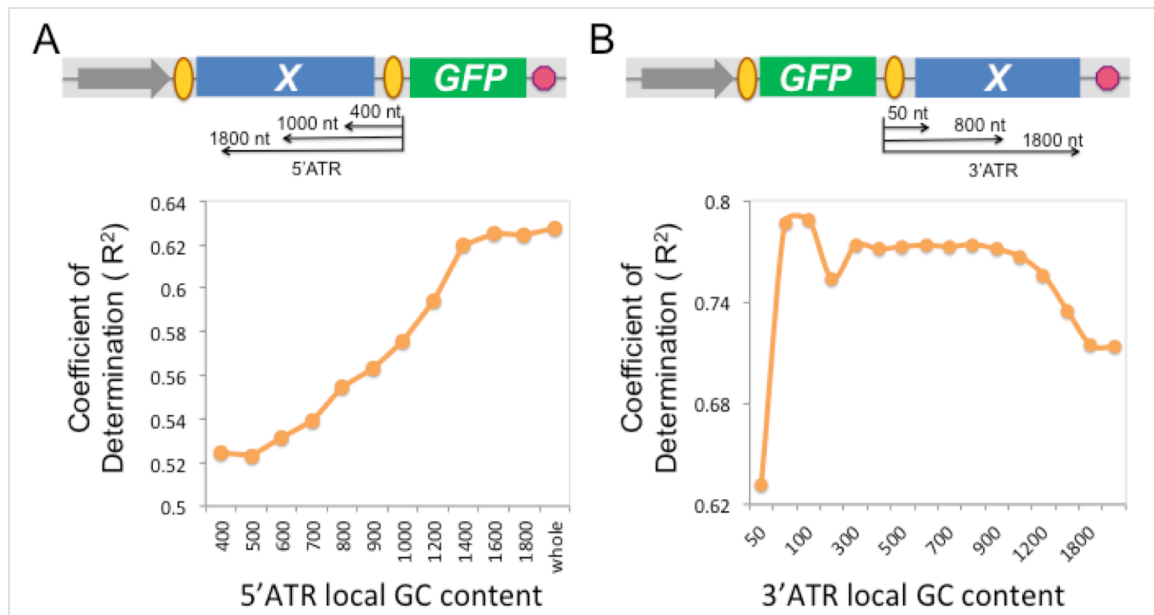


Figure 27: Sliding window analysis for local GC content and model fitting. (A) TOP: Schematic representation of constructs with GFP distal to the constitutive promoter. The black lines with arrows indicate 5' ATRs with different lengths. Bottom: GC content of 5' ATRs with different lengths from 400 nucleotides to the whole transcriptional region are calculated and then fitted to the model. The coefficients of determination (R^2) are compared to 5' ATRs with different lengths. Linear model results show that GC content with the whole 5' ATR has the highest fitting efficiency. Gray arrow: constitutive promoter; Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. *X*

represents different genes used in the circuit. (B) TOP: Schematic representation of constructs with GFP proximal to the constitutive promoter. The black lines with arrows indicate 3'ATRs with different lengths. Bottom: GC content of 3'ATRs with different lengths from 50 nucleotides to the whole transcriptional region are calculated and then fitted to the model. The coefficients of determination (R^2) are compared to 3'ATRs with different lengths. Model fitting results show that GC content of the first 100 nucleotides 3'ATR has the highest fitting efficiency.

4.2.3 Protein expression is also affected by non-coding ATR sequences

Non-coding DNA sequences also have important regulatory functions. To investigate whether non-coding sequences would similarly affect adjacent gene expression in synthetic operons, we engineered 32 synthetic circuits with 32 genes, of which 23 are transcriptional factors (common repressors/activators) and three are membrane protein genes. All of the 32 genes are placed immediately downstream of the promoter without RBS, thus they would not be translated (Figure 28). We reasoned that if GFP variations in Figure 26 were not resulted from the interference of adjacent proteins' expression, the adjacent gene regulation would be similar for non-coding ATRs.

Our results showed strong correlations between GFP expression and non-coding 5'ATR GC content, size, and local mRNA folding energy (Figure 28). Overall, with the three variables the model can explain about 65% variation of GFP expression for operons with non-coding regions (Figure 28). However, the coefficients for the three variables are slightly different from cases with RBS in Figure 26A. Higher GFP expression was observed for circuits with the same genes with RBS than those without RBS (Figure 29), suggesting the RBS of 5'ATR may be important for mRNA stabilization and expression efficacy.

Non-coding DNA sequence has been demonstrated to make up about 98% of the whole human genome and is capable of regulating gene expression and modulate chemotherapeutic resistance^{172–174}. Our findings offer direct evidence that adjacent non-coding DNA fragments can also regulate gene expression in synthetic operons. Moreover, this result also suggests that the GFP variation does not result from its neighbor proteins' expression, but more related to the mRNA transcript itself.

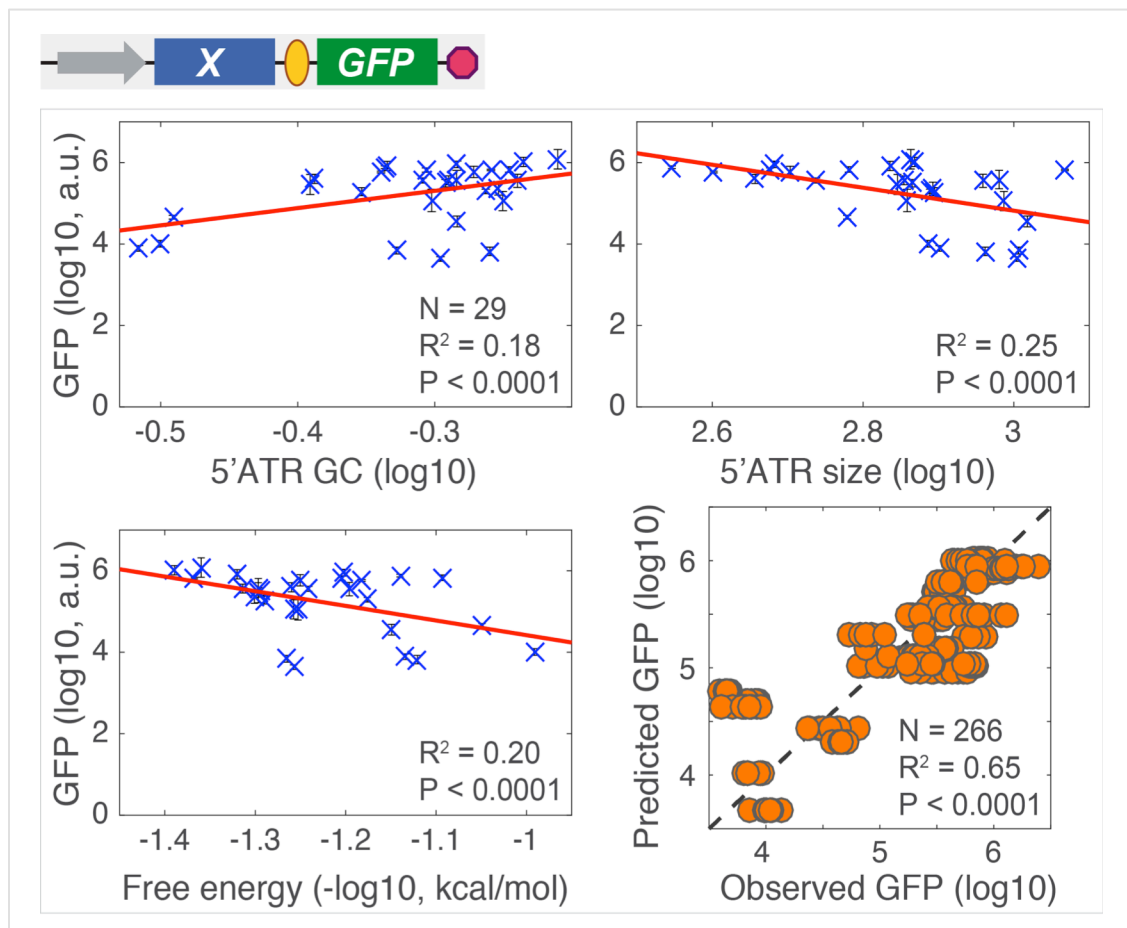


Figure 28: Quantitative characterization of non-coding ATRs regulation on gene expression. Top: Schematic representation of synthetic gene circuits with non-coding 5'ATR. Grey arrow: constitutive promoter; Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. Rectangles with filled colors represent different genes. 29 genetic circuits with different genes (*X*) were constructed. These genes would not be expressed owing to a lack of ribosome binding sites. Linear regression results indicate that 5'ATR GC content, size and local folding free energy are significantly correlated with GFP expression and the three variables together explain 65% GFP fluorescence variation.

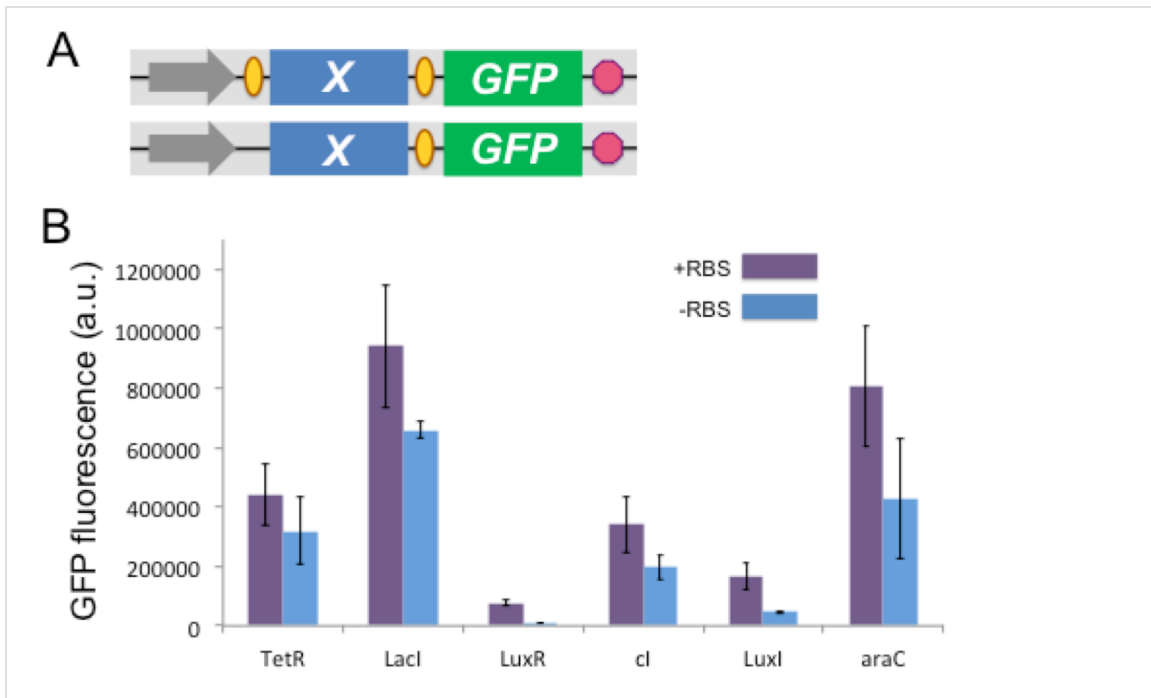


Figure 29: Comparison of GFP expression between coding ATRs and noncoding ATRs. (A) Schematic representations of synthetic bi-cistronic gene circuits. GFP reporter was expressed downstream of a coding ATR (with RBS) or noncoding ATR (without RBS), respectively. Gray arrow: constitutive promoter; Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. Rectangles with filled colors represent different genes. *X* represents different genes used in the circuit. (B) Higher GFP expression was observed for circuits with the same genes (*TetR*, or *LacI*, or *LuxR*, or *cl*, or *LuxI*, or *AraC*) with RBS than those without RBS. Fluorescence was measured by flow cytometry. Error bar represents standard deviation of eight biological replicates.

4.2.4 Comprehensive model of ATR regulation

Our results revealed that gene expression in operons is significantly regulated by its adjacent genes' sequence features and local mRNA secondary structures. The explicit mechanism of these effects remains unclear. Since the mRNA is transcribed from a single constitutive promoter and the amount of GFP mRNA is roughly equal, we attribute observed fluorescence variations to translation and degradation of mRNA, which occur simultaneously in *E. coli* through their machinery's competitively binding to mRNA after transcription initiation. Moreover, the same RBS for each construct ensures a similar

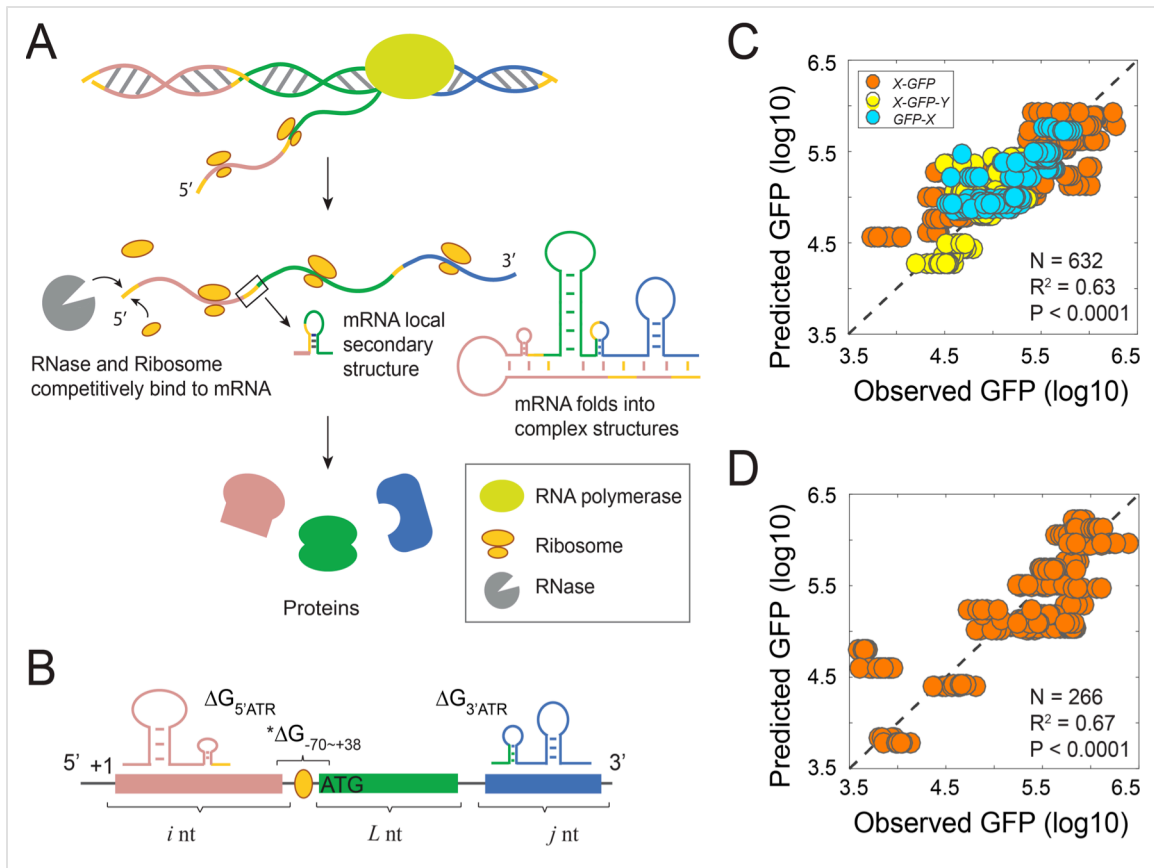


Figure 30: A general model for ATR regulation on gene expression. (A) Co-transcriptional translation and degradation. After RBS is transcribed, RNase and ribosome competitively bind to mRNA to initiate translation or degradation. Generally, gene expression is influenced by overall stability and local secondary structure. (B) Illustration of the five variables in the model: $\Delta G_{5'ATR}$, $\Delta G_{3'ATR_100}$, $\Delta G_{-70\sim+38}$, and transcriptional size (*i*, *j*). -70 and +38 is corresponding to the position of start codon (AUG) of interested gene. (C) Experimentally observed GFP expressions are plotted against the model predicted GFP values with the five energetic terms for all the three scenarios' data in Figure 26. The comprehensive model shows similar level of fitting efficiency to Figure 26G. Dots with different colors indicate the data source from the three scenarios in Figure 26 (D) is the experimentally measured GFP fluorescence from the non-coding ATR regulation data in Figure 28 compared to the linear regression model predicted GFP expression with the three energetic terms. The model shows a slightly higher fitting efficiency compared to the result in Figure 26.

efficacy of translational initiation. Thus, mRNA degradation and secondary structures play major roles in GFP expression variations. *E. coli* mRNAs are unstable in the cell with short half-lives (few minutes), and generally degraded by RNA degradosome

assembled on RNase E, which catalyzes the rate-limiting cleavage of many mRNAs from 5' to 3' direction with a preference for AU-rich sequences¹⁷⁵⁻¹⁸⁰. Previous reports have also shown that a structured mRNA with high GC content is likely to have more stable secondary structure, conferring stability for downstream sequences^{180,181}. Thus, we infer that a high GC-content 5'ATR may help stabilize the GFP transcript and slow the degradation process, leading to higher GFP expression. Our results in Figure 26 also showed that constructs with GFP distal to promoter has the highest variation ($\sigma^2 = 0.334$) while GFP proximal to promoter has the much lower variance ($\sigma^2 = 0.121$), suggesting the 5'ATR secondary structure probably influences GFP expression.

We then develop a comprehensive linear model based on measurements of sequence-dependent energetic changes during polycistronic mRNA folding/translation and costs of protein bio-synthesis¹⁸²⁻¹⁸⁴. The energetic changes are corresponding to the translation efficiency and protein abundance (c).

$$c \propto \exp(-\sum \beta_x \Delta G_x), \quad x = 1, 2, 3, \dots$$

where ΔG is the energy term and β is the scaling coefficient^{183,185}. For a given gene in an operon, the size of 5' and 3' ATR is denoted as i nt and j nt, respectively (Figure 30A-B). The minimum free energy of local GFP mRNA secondary structure around RBS is $\Delta G_{70\sim+38}$. The entire folding energy for 5'ATR is $\Delta G_{5'ATR}$, and the 100-nt 3'ATR is $\Delta G_{3'ATR_{100}}$. So the sum of energy changes can be quantified to assess the abundance of a given gene expression:

$$-\sum \beta_x \Delta G_x = \beta_0 + \beta_1 * \Delta G_{5'ATR} + \beta_2 * \Delta G_{3'ATR_{100}} + \beta_3 * i * G_m + \beta_4 * j * G_m + \beta_5 * \Delta G_{70\sim+38}$$

where the folding energy of $\Delta G_{5'ATR}$, $\Delta G_{3'ATR_{100}}$ and $\Delta G_{-70\sim+38}$ are totally sequence-dependent and G_m is an average energy cost for synthesizing a nucleotide, which here for simplicity we assume it is a constant. Based on the computed free energy and sequence length for the constructs, we found that the model has significant correlation with experimental data and explains 63% of the GFP variation for constructs with coding ATRs (Figure 30C) and 67% for circuits with non-coding ATRs (Figure 30D). With this comprehensive tool, we can evaluate the influence of the adjacent transcriptional sequences on a certain gene's expression in the operon, which provides a guide for circuit design and optimization during circuit engineering.

4.2.5 Predictor-guided circuit design for synthetic logic gates

To illustrate how the tool could be used to guide circuit design, synthetic AND logic gate was designed and tested. The gate is composed of a hybrid promoter *pLux/tet*, which has one LuxR-AHL and one TetR binding site. GFP is the output. Maximized GFP expression is achieved in presence of two inputs AHL and aTc (Figure 31A), where AHL binds with LuxR protein to activate *pLux/tet* transcription and aTc can block TetR repression to *pLux/tet*. LuxR and TetR are constitutively expressed from the same promoter.

There are two possible ways to assemble this circuit, one is LuxR-TetR (LT) combination, and the other is TetR-LuxR (TL). The GC content of *LuxR* (30.3% GC, 781 bp) is lower than *TetR* (40.4% GC, 685 bp). So in AND-gate LT, TetR expression is lowered by its 5'-low-GC-content neighbor while the impact of LuxR to TetR expression in logic TL is minor because the size of 3'ATR is a more significant factor compared to

GC content. We then calculated the equation for each circuit design and feed it into our model, results indicate that LuxR expression in TL decreases by 4.4% compared to gate LT, however, TetR expression increases by 93.6% in circuit TL (Table 6). Therefore, we infer that the basal GFP expression in circuit LT would be greater than in TL, whereas TL would harbor more dynamic responses with induction of aTc because of higher TetR

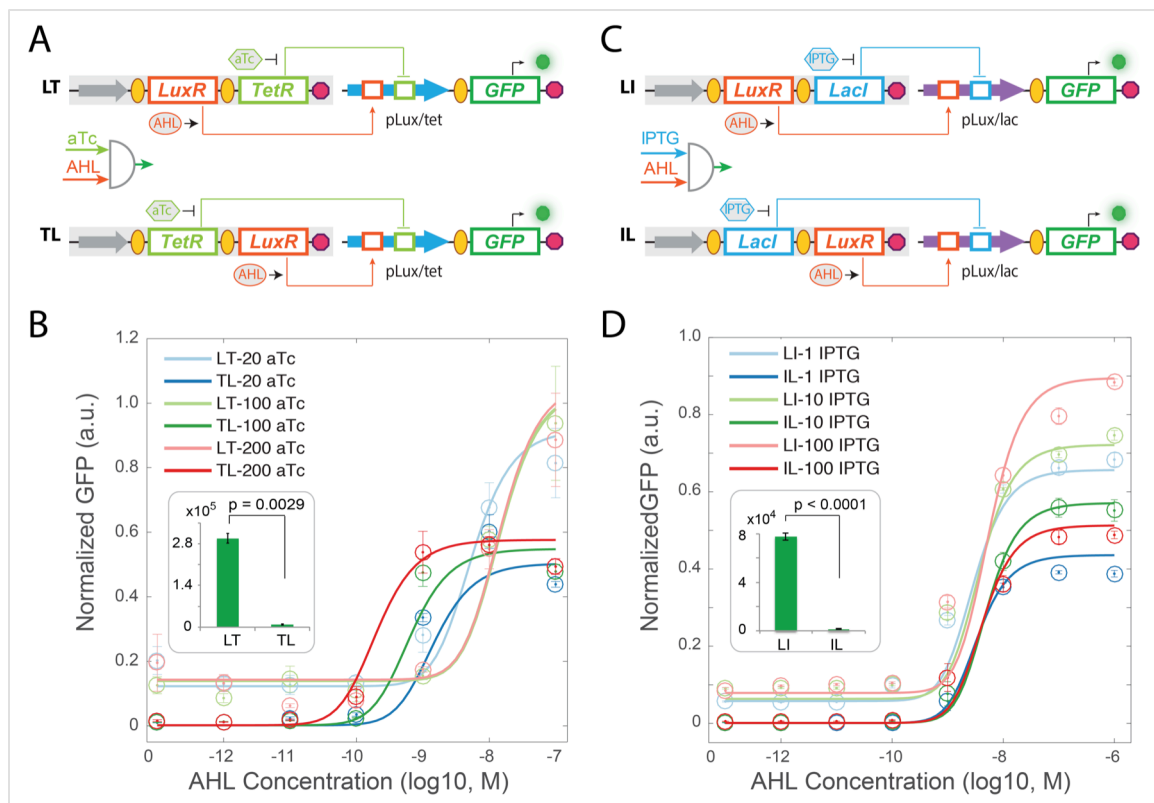


Figure 31: Model guided circuit design for synthetic logic gates. (A) Two designs for *pLux/tet*-AND logic gate. A constitutive promoter (grey arrow) drives LuxR (orange rectangle) and TetR (green rectangle) expression. *pLux/tet* is highly activated in presence of both AHL and aTc. LT and TL represent the order of LuxR and TetR positions in the operon. LuxR can bind with AHL (grey oval) to activate *pLux/tet* promoter (blue arrow), while aTc (green hexagon) can block the TetR inhibition to *pLux/tet* promoter. Lines with arrowheads indicate activation, and lines with T-bars indicate inhibition. Orange oval: ribosome binding site; Red hexagon: transcriptional terminator. (B) Dose responsive curves for different concentrations of AHL and aTc. The solid lines are from ODE model simulations based on the calculated relative changes of LuxR and TetR concentrations in LT and TL from our linear comprehensive model. Data points with error bar are

experimental results, showing good match with model predictions. Inserted diagram is the basal expression of GFP for design LT and TL. Data represent the mean \pm s.d. of three replicates. Color curves are inductions with different aTc concentrations (20 ng/ml, 100 ng/ml, and 200 ng/ml). (C) Two designs for *pLux/lac*-AND logic gate. A constitutive promoter (grey arrow) drives LuxR and LacI expression. *pLux/lac* (purple arrow) is highly activated in presence of both AHL and IPTG (blue hexagon). LuxR can bind with AHL to activate *pLux/lac* promoter, while IPTG can block the LacI inhibition to *pLux/lac* promoter. LI and IL represent the order of LuxR and LacI positions in the operon. (D) Dose responsive curves for different concentrations of AHL and IPTG. The solid lines are model simulations based on the calculated relative changes of LuxR and LacI concentrations in LI and IL from our linear comprehensive model. Experimental results (data point with error bar) show good match with model predictions. Color curves are inductions with different IPTG concentrations (1 μ M, 10 μ M and 100 μ M). Inserted diagram is the basal expression of GFP for design LI and IL. Data represent the mean \pm s.d. of three replicates. p-value is calculated from student's t-test.

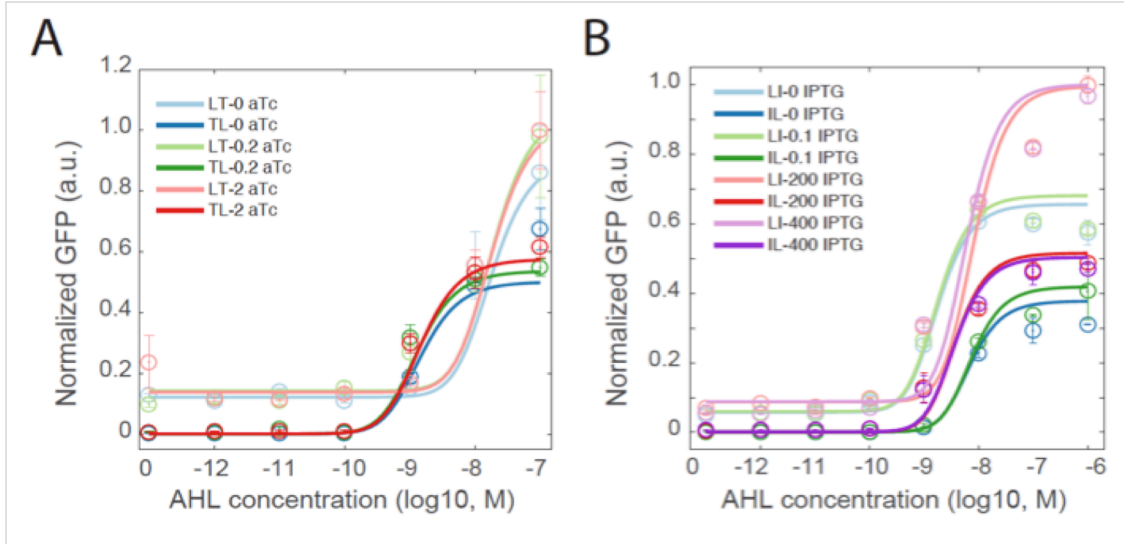


Figure 32: Model simulation and experimental validation of GFP dynamics for synthetic logic gates. (A) Dose responsive curves for different concentrations of AHL and aTc. The solid lines are from ODE model simulations based on the calculated relative changes of LuxR and TetR concentrations in LT and TL from our linear comprehensive model. Data points with error bar are experimental results, showing good match with model predictions. Color curves are inductions with different aTc concentrations (0 ng/ml, 0.2 ng/ml, and 2 ng/ml). Data represent the mean \pm s.d. of three replicates. (B) Dose responsive curves for different concentrations of AHL and IPTG. The solid lines are model simulations based on the calculated relative changes of LuxR and LacI concentrations in LI and IL from our linear comprehensive model. Experimental results (data point with error bar) show good match with model predictions. Color curves are inductions with different IPTG concentrations (0, 0.1 μ M, 200 μ M and 400 μ M). Data represent the mean \pm s.d. of three replicates.

Table 6: Model evaluation for each gene's expression in the AND logic gate.

Coefficients of the comprehensive model								
	5' ATR Size	3' ATR Size	Free energy ($\Delta G_{-70 \rightarrow +38}$)	$\Delta G_{5' ATR}$	$\Delta G_{3' ATR}$ (100nt)	Intercept		
	-3.10443	-1.15864	-1.1206	-1.80151	-0.28049	11.4635		
AND-Gates								
CP-LuxR-LacI(LacI)	815	8	-16.9	-121.6	-0.00001			
CP-LuxR-LacI(LuxR)	8	1187	-70	-0.05	-16.6			
						Predicted expression	Relative expression ($10^{(IL-LI)}$)	Overall efficiency
CP-LuxR-LacI(LacI)	2.911157609	0.903089987	-1.227886705	-2.084933575	5	5.109207328		
CP-LuxR-LacI(LuxR)	0.903089987	3.074450719	-1.84509804	1.301029996	-1.220108088	5.163765204		
CP-LacI-LuxR(LacI)	8	815	-70	-0.05	-7.1		1.38132904	2.762658081
CP-LacI-LuxR(LuxR)	1187	8	-15	-365.5	-0.00001		1.743840175	1.743840175
CP-LacI-LuxR(LacI)	0.903089987	2.911157609	-1.84509804	1.301029996	-0.851258349	5.222548745		
CP-LacI-LuxR(LuxR)	3.074450719	0.903089987	-1.176091259	-2.562887381	5	5.404555747		
CP-LuxR-TetR(TetR)	815	8	-12.5	-121.6	-0.00001			
CP-LuxR-TetR(LuxR)	8	719	-70	-0.05	-11.6			
							($10^{(TL-LT)}$)	
CP-LuxR-TetR(TetR)	2.911157609	0.903089987	-1.096910013	-2.084933575	5	4.96342505		
CP-LuxR-TetR(LuxR)	0.903089987	2.85672889	-1.84509804	1.301029996	-1.064457989	5.381238175		
CP-TetR-LuxR(TetR)	8	815	-70	-0.05	-7.1		1.936732422	1.936732422
CP-TetR-LuxR(LuxR)	719	8	-13.7	-152.5	-0.00001		0.956758533	0.956758533
CP-TetR-LuxR(TetR)	0.903089987	2.911157609	-1.84509804	1.301029996	-0.851258349	5.222548745		
CP-TetR-LuxR(LuxR)	2.85672889	0.903089987	-1.136720567	-2.183269844	5	5.35336342		

expression. An ordinary differential equation (ODE) model was then developed to simulate GFP expression based on the computed LuxR and TetR production changes in the LT and TL gates (more detail in Supplementary methods). Through tuning the relative production rates of LuxR and TetR in the model, we can predict the GFP dynamics under induction of AHL and aTc (Figure 31B and Figure 32A, solid lines).

Experimental dose-response results further supported our model analysis that for all aTc concentrations, basal expression of *pLux/tet* in circuit LT is significantly higher (~35 fold) than in circuit TL (Figure 31B and Figure 32A, data points with error bar). Moreover, the maximum GFP fluorescence is also higher in circuit LT, owing to decreased LuxR expression in gate TL. However, the sensitivity to AHL (concentration for half-maximal activation of GFP, $K_{0.5}$) is improved 2.4 ~ 64.5 fold in circuit TL compared to LT for different concentrations of aTc. And the nonlinearity (hill coefficient) is generally increased 2 ~ 5 fold with high concentrations of aTc induction. These data are in accordance with the model calculations that TetR expression is relatively increased in circuit TL than in LT, which suppresses the basal expression of *pLux/tet* and improves the sensitivity and nonlinearity of the promoter to AHL and aTc.

To further validate the method's utility, another two AND-gate gene circuits (LI and IL) with switched genes' (*LuxR* and *LacI*) position were designed (Figure 31C). Hybrid promoter *pLux/lac* was used to indicate the relative concentrations of LuxR and LacI produced from the operon. *LacI* (53.3%, 1153 bp) has a high GC content, which may increase LuxR expression. Our model calculations showed that LuxR expression increases by 74.3% and LacI increases by 38.1% in circuit IL than in LI (Table 6). Since promoter *pLux/lac* has two LacI-binding sites (one is in the region between -35 and -10,

and the other is downstream of -10 element), so the overall LacI inhibition efficiency is increased ~ 76.2% considering the importance of spacing between -35 and -10 element to RNA polymerase binding. Therefore, the basal GFP expression of logic IL would be lowered compared to LI. The ODE model also indicates higher GFP expression in gate LI (Figure 31D and Figure 32B, solid lines). Experimental results confirmed that the basal expression for circuit LI is ~ 54-fold higher than IL, and GFP expression under each inductions are higher in gate LI, which is consistent with the ODE model results (Figure 31D and Figure 32B, data points with error bar).

Taken together, the two sets of AND logic gates showed an example of applying our comprehensive model based tool to evaluate each gene's relative expression level in synthetic AND gate gene circuits, and verified that ATRs' features and local mRNA stability changes in an operon-based gene network affect genes expression and the circuit performance, including basal level, sensitivity, and nonlinearity. Furthermore, the tool could serve as a much-needed quantitative guidance to rational design and optimization of gene expression for large genetic circuits.

4.2.6 Tuning gene expression with synthetic 5' adjacent transcriptional regions

Next, we seek to use synthetic non-coding DNA fragment to fine-tune gene expression. Since GC content has been shown as a critical factor, we synthesized six short DNA fragments (with a constant size 200 bp) with varying GC content from 28% to 53%, which were inserted downstream of *LuxR* gene but upstream of *GFP* in the two-gene operon (*Promoter-LuxR-Synthetic fragment-GFP*). According to our model, synthetic fragments with varying GC content could tune GFP expression.

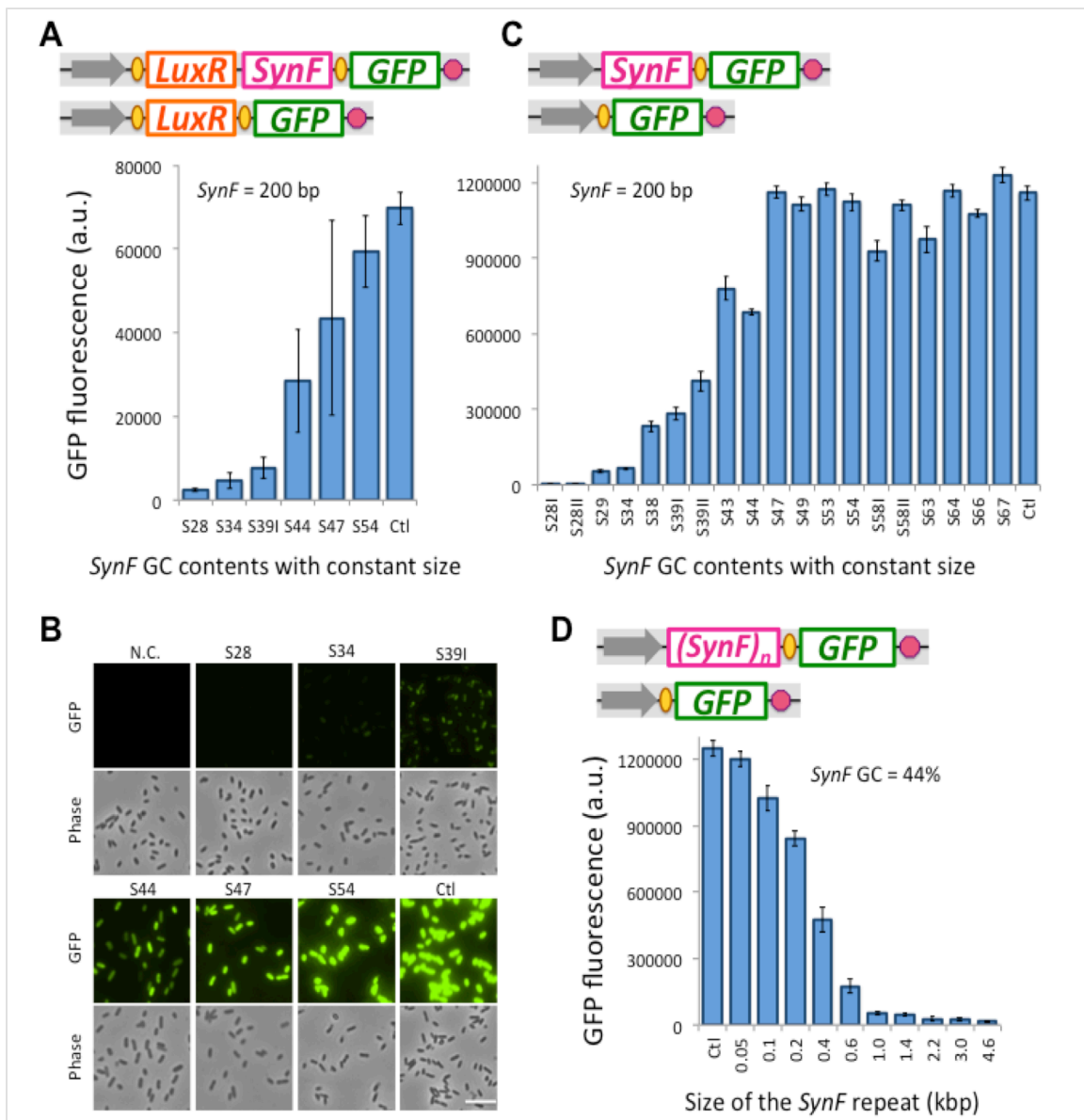


Figure 33: Tuning gene expression with synthetic 5' adjacent transcriptional regions. (A) Synthetic 5' ATRs (*SynF*) to tune GFP expression for circuit *CP-LuxR-GFP*. 200 bp ATRs were inserted between *LuxR* and *GFP* genes to tune GFP expression, and control (Ctl) is constructed without ATR insert. Flow cytometry results indicate that GFP fluorescence increases with gradually increasing 5' ATRs GC content from 28% to 54%. (B) Microscopic results of GFP fluorescence for the constructs in (A). Scar bar: 5 μ m. Magnification: 40x. (C) Synthetic 5' ATRs (*SynF*) with different GC content to tune GFP expression for circuit *CP-GFP*. All the *SynF* have the same size (200 bp) and are inserted upstream of *GFP* gene (Top). Flow cytometry results of GFP fluorescence for 5' ATRs with GC content from 28% to 67% (Bottom). (D) Circuits with different 5' ATR sizes (through shortening and adding a common sequence S44, GC: 44%, Size: 200 bp) were constructed to tune GFP expression. Flow cytometry results show that GFP fluorescence intensity gradually decreases with increasing 5' ATR sizes.

Experimental results show that GFP expression is continuously increased for synthetic fragments with increasing GC content from 28% to 53% (Figure 33A). Low GC-content fragments down-regulated GFP expression for about 25 fold. Microscopy results further confirmed flow cytometry data and visualized gradual increase of fluorescence intensity with increasing GC-content ATRs (Figure 33B). Using this strategy, we further synthesized 13 DNA fragments as 5' ATRs with varying GC content but having a constant size (200 bp), and placed downstream of the promoter (Figure 33C). Results indicate that synthetic short DNA sequences have a substantial impact on GFP expression: low GC-content ATRs largely decrease its neighbor's expression (up to 366 fold), and exhibit a gradually increasing pattern from 28% to 48%, while high GC-content (48% to 67%) ATRs drive GFP expression to a comparable level to the control (without synthetic fragments). It is possible that GFP achieves its maximum expression when the upstream ATR mRNA piece has a relatively stable structure. To further verify the ATR regulation, we changed the size of 5' ATR through shortening and adding a common sequence. Using S44 (GC: 44%; Size: 200 bp) in Figure 33C as the seed sequence, we shortened to 100 bp and 50 bp, and added to 400 bp (combined with two pieces of S44) to 4600 bp (combined with 23 pieces of S44), and all the ten fragments have the same GC content (44%, Figure 33D). Flow cytometry results show that GFP fluorescence intensity gradually decreases with increasing 5' ATR sizes (Figure 33D). Taken together, we demonstrate that synthetic non-coding 5' ATRs with designed GC content and sizes can be used to accurately tune gene expression and achieve expression levels spanning more than 300-fold.

4.2.7 Using synthetic ATRs to tune toggle switches

To further test the utility of synthetic non-coding 5' ATR in tuning gene expression, we applied synthetic 5' ATRs to tune the classical “toggle switch”. As illustrated in Figure 34A, LacI protein could inhibit TetR by binding the *pLac* promoter while TetR could bind *pTet* to block LacI expression, forming a mutually inhibitory network. Here, we positioned 200 bp synthetic ATRs with 28% and 67% GC content upstream of RBS-TetR module to tune TetR production (T_S28 and T_S67). According to our analysis above, low GC-content 5' ATR can down-regulate TetR expression while high GC-content can keep TetR at a high level.

Flow cytometry was employed to analyze the initial state and population distribution for cells harboring the three toggle switches. As shown in Figure 34B, T_WT initially shows bimodal distribution, GFP-ON and GFP-OFF populations, resulting from gene expression noise in a relatively balanced system. In contrast, both T_S28 and T_S67 exhibited unimodal distributions. Synthetic ATR S28 decreased TetR expression leading to higher LacI and GFP expression, whereas fragment with 67% GC content showed a lower GFP expression than T_S28 and slightly lower than the high-GFP population cells in T_WT (Figure 34B). The results indicated that the synthetic ATRs could tune TetR expression and initial steady states of the toggle switches.

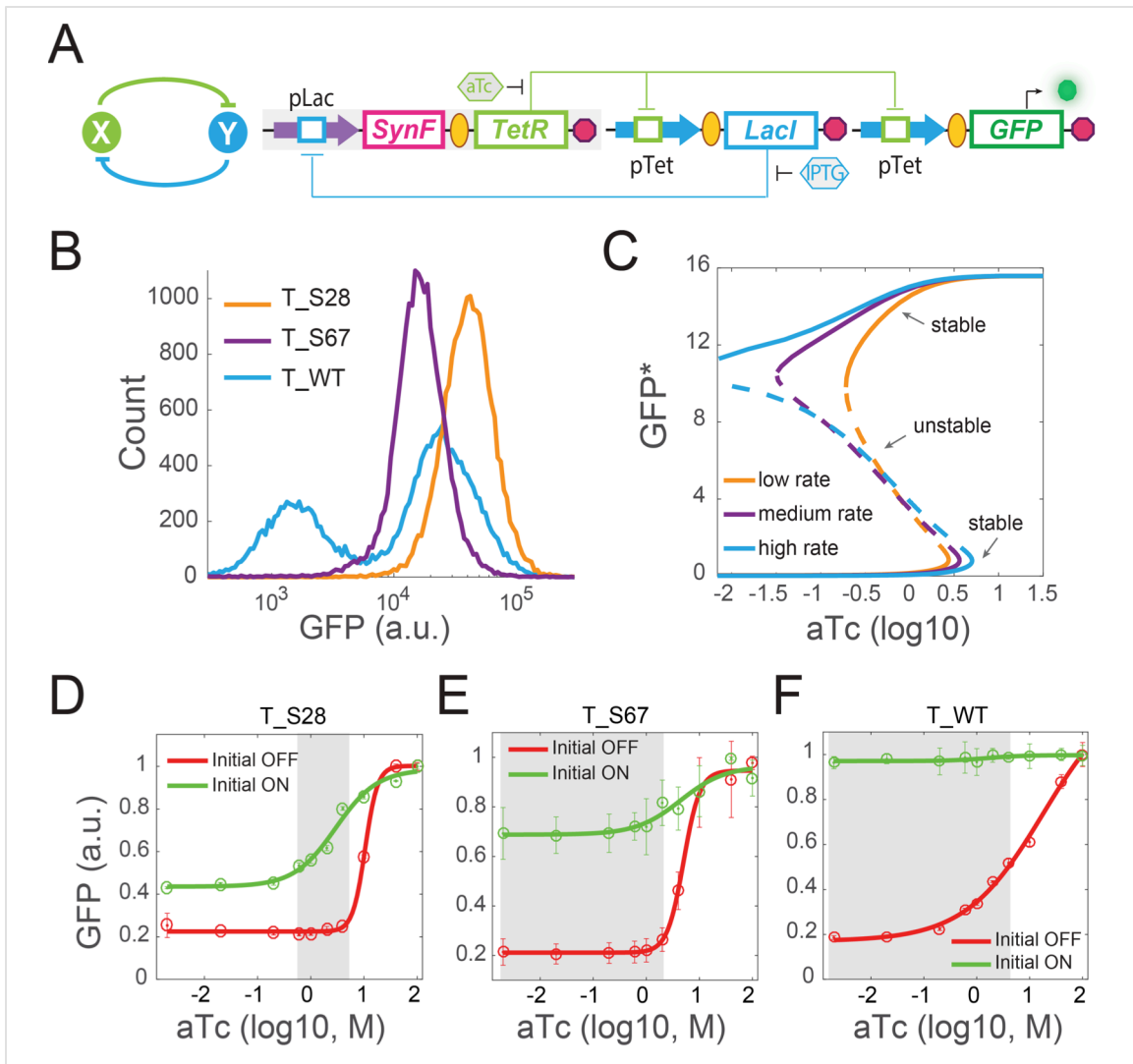


Figure 34: Using synthetic ATRs to modulate bistability of toggle switches. (A) Left: Abstract diagram of toggle switch topology, where X and Y mutually inhibit each other. Right: Molecular implementation of the toggle switch. *LacI* inhibits *TetR* by binding the *pLac* promoter while *TetR* binds *pTet* to block *LacI* expression, forming a mutually inhibitory network. Inducers IPTG and aTc (hexagon) can respectively relieve *LacI* and *TetR* inhibition. GFP serves as the readout of *pTet* promoter. Synthetic ATRs (*SynF*) were engineered upstream of *TetR* gene. (B) Initial steady states for the three toggles. Toggle without ATR insertion (T_WT) shows bimodal distribution (GFP-OFF and GFP-ON), while T_S28 (ATR with 28% GC content) shows higher GFP expression and T_S67 (ATR with 67% GC content) shows lower GFP expression than the GFP-ON population of T_WT. (C) Bifurcation analysis for GFP (*LacI*) expression with different *TetR* production rates under induction of varying concentrations of aTc. A low production rate for *TetR*, corresponding to T_S28, has the smallest bistable region, while a high rate (corresponding to T_WT) has the broadest bistable region. Solid lines represent stable steady-state solutions and dotted lines are unstable steady-state solutions. GFP* is the computed GFP abundance from the model. (D-F) Hysteresis results for

toggles T_S28, T_S67 and T_WT under induction of varying concentrations of aTc. Red lines indicate the initial OFF cells with basal GFP expression while green lines indicate the initial ON cells with high GFP expression. Data represent the mean \pm s.e. of three replicates. Grey area is the presumable bistable region for each circuit.

To achieve a quantitative understanding of the ATR's regulation on bistability, we performed bifurcation analysis from the same mathematical model as the classical toggle switch⁶. We found that the production rate of TetR has a considerable effect on bistability and bistable region (Figure 34C). A small production rate, corresponding to low-GC ATR, has a small bistable region, whereas increase of production rate leads to larger bistable regions. Experimentally, hysteresis of the three toggles were tested to verify the model analysis. Results indicated that all the three toggles exhibited hysteresis and T_WT harbors the broadest bistable region (Figure 34D-F). Moreover, consistent with model analysis, the bistable regions are gradually decreased from T_WT to T_S67 to T_S28. Collectively, these results validate a novel strategy of using synthetic ATRs to tune gene networks' initial steady states and bistability. Furthermore, this example demonstrates the feasibility of bridging ATR regulation with mathematical modeling to quantitatively understand and tune gene network dynamics.

4.3 Discussion

Circuit engineering is the first step for synthetic biologists to achieve designed functionalities with synthetic gene circuits. A successful synthetic gene circuit depends on the full characterization of biological components and the emerged interactions between modules when assembled into a complete gene network^{23,104,186,187}. Development of a reliable tool to predict protein expression in the circuit has wide applications in

biotechnology. For example, RBS Calculator is a well-developed design tool to predict and control translation initiation and protein expression in bacteria^{183,185}. Further systematic investigation revealed that mRNA translation initiation rate is controlled by coupled trade-offs between site accessibility, partial RNA unfolding and ribosomal sliding in the structured upstream standby sites, which leads to a more accurate prediction of RBS strength and protein expression¹⁸⁸.

Here, we systematically investigated how the adjacent transcriptional regions regulate gene expression in synthetic operon-based gene circuits. Through placing the GFP at different positions (proximal, middle, and distal) to the promoter, we developed a new gene expression-prediction method that takes into account the adjacent transcriptional regions' features including GC content, size and stability of mRNA folding near RBS (Figure 26), of which 5'-ATR GC content and 3'-ATR size are two most significant factors. It is necessary to point out that most ATRs' sizes in the circuits are 500 - 2000 bp, and the maximum is 2422 bp, which may undermine the contribution of ATR's size to GFP variation. Moreover, because of the limitation in sample size and available gene resources, the collected data is not perfectly normally distributed, especially for circuits with GFP in the middle (*X-GFP-Y*), which may compromise the robustness of the model. We also found that membrane protein genes cause significant inhibitions to GFP expressions (Figure 23), suggesting gene types may also have pronounced effects on the adjacent genes expression. It is possible that the secondary structure of membrane protein transcripts induced ribosome stalling¹⁸⁹ or mRNA-programmed translation pausing during membrane protein targeting¹⁹⁰ may also influence the adjacent genes expression.

Consistent with previous results that gene position in operons can affect gene expression^{166,167}, our results further demonstrated that gene position (corresponding to ATR's change) significantly altered gene network dynamics including basal expression, system sensitivity, and nonlinearity, which has profound impacts for nonlinear dynamic systems. Such adjacent gene regulation effect has been generally neglected during construction of synthetic gene networks. Furthermore, we validated that the usage of designed synthetic DNA fragments as 5'ATRs to tune gene expression and modulate bistable regions of genetic toggle switches.

The explicit mechanism about how the ATR's GC content and size regulates gene expression is not well understood. Here there are several possible explanations. Previous studies and our results share a common conclusion that the difference mostly comes from post-transcriptional regulation because of little mRNA concentration change for different transcript length and gene position in the operon¹⁶⁶⁻¹⁶⁸. Since there is a lack of complicated post-translational modifications in *E. coli*, we believe that the ATR alters the secondary or tertiary structures of mRNA locally and/or globally, which perturbs the GFP mRNA translation and degradation process. After ribosome binding site is transcribed, ribosome and RNase competitively bind to mRNA^{177,191} and a GC-rich ATR, which is likely to be more stable, could help stabilize GFP transcript and decrease the risk of degradation by RNase, and thus result in higher GFP expression. Another possible explanation is GC-rich isochores may correspond to open chromatin structure and favor the accessibility of transcriptional factors, leading to increased transcription and translation¹⁹²⁻¹⁹⁴. On the other hand, a large size of neighbor gene increases the probability of elongation pausing (RNA polymerase or ribosome drop off the template)

and degradation by RNase, and competitively recruits more protein synthesis machinery resources in the cell, resulting in decreased gene expression. It is noteworthy that during the manuscript preparation, a latest study reported that the engineered non-coding adjacent region mutations could alter transcriptional factor occupancy and local epigenetic environments to modulate gene regulation and chemotherapeutic resistance in melanoma ¹⁷⁴.

This work established a novel regulation of gene expression – adjacent transcriptional region (including 5' and 3' coding and non-coding regions) regulation, which expands current considerations of untranslated transcriptional region (UTR) or upstream open reading frame (uORF). Our quantitative results demonstrated that ATR regulation could explain 56% - 79% of gene expression variation, depending on different circuit organizations. Hence, this finding broadens our understanding of natural gene regulation strategies and helps refine the determinants for gene expression in an operon or even genome.

A central goal of synthetic biology is to develop genetic circuits to program cell behaviors in a predictable way. With the increasing complexity of integrated multi-layer circuits, specific bio-components' organization and circuitry structure design become extremely important for its functionality^{39,187,195}. The tool we provided here to evaluate each gene's expression level in a circuit would save experimentalists' time and resources to screen and test modules' combinations, and thus should greatly facilitate optimization of circuit design and accelerate the engineering of complex gene networks.

4.4 Materials and Experimental Methods

4.4.1 Strains, media and chemicals.

All cloning experiments and fluorescent measurements were performed in *Escherichia coli* DH10B (Invitrogen). Synthetic toggle switches (T_S28, T_S67 and T_WT) were tested in *E. coli* K-12 MG1655 strain with *lacI*^{-/-}¹¹⁶. Cells were cultured in liquid or solid Luria-Bertani (LB) broth medium with 100 µg/ml ampicillin at 37°C. Chemicals AHL (3oxo-C6-HSL, Sigma-Aldrich), Arabinose (Sigma-Aldrich, USA), isopropyl β-D-1-thiogalactopyranoside (IPTG, Sigma-Aldrich), and anhydrotetracycline (aTc, Sigma-Aldrich) were dissolved in ddH₂O and diluted into indicated working concentrations. Cultures were shaken in 5 mL and/or 15 mL tubes at 220 rotations per minute (r.p.m).

4.4.2 Plasmid construction.

Most genes are obtained from iGEM Registry (http://parts.igem.org/Main_Page). These genes are often used in synthetic biology projects, including transcriptional factors, quorum-sensing components, and other functional genes (Table 7). Plasmids were constructed using standard molecular biology techniques and all genetic circuits were assembled based on standardized BioBrick methods. As an example, construct *Promoter-TetR-GFP* is composed of five BioBrick standard biological parts: BBa_J23104 (constitutive promoter, CP), BBa_B0034 (ribosome binding site, RBS), BBa_C0040 (*tetR*), BBa_E0040 (green fluorescent protein, GFP) and BBa_B0015 (transcriptional terminator). To produce RBS-TetR module, plasmid containing *TetR* was digested by *XbaI* and *PstI* as the insert fragment while RBS vector was cut by *SpeI* and *PstI*. Both

fragment and vector were separated on 1% TAE agarose gel electrophoresis and purified using PureLink gel extraction Kit (Invitrogen). Purified fragment and vector were then ligated by T4 DNA ligase (New England Biolabs, NEB). The ligation products were further transformed into *E. coli* DH10B and plated on LB agar plate with 100 µg/ml ampicillin for screening. Finally, plasmids extracted by GenElute HP MiniPrep Kit (SIGMA-ALDRICH) were confirmed through gel electrophoresis (digested by *EcoRI* and *PstI*) and DNA Sequencing (Biodesign sequencing Lab, ASU). Similar steps were carried out for subsequent rounds of cloning to assemble the whole construct. All the circuits' DNA sequences are provided in the Supplementary information.

Also, 17 transcriptional factors with varying GC content and sizes were amplified from *E. coli* genome with designed primers (Table 8). Synthetic sequences were randomly generated with the same length (200 bp) but various GC contents (28%-67%). Sequences with ribosome binding site-features (AGGAGG) were redesigned to exclude its translation potential. All synthetic sequences and primers were synthesized as custom DNA oligos or gBlocks gene fragments from Integrated DNA Technologies (IDT). In order to express consistently in the cell, all constructs were finally subcloned into pSB1A3 vector prior to the test. The sequence for the constructs can be found in Supplementary information.

4.4.3 Flow cytometry measurements.

All confirmed constructs were re-transformed into DH10B strain and twelve individual colonies in total (repeated in three different days, four colonies each time) for each construct were randomly picked up to grow. Samples were then cultured in 4 ml LB

medium (100 µg/ml ampicillin) for 24hr at 37°C for testing. Flow cytometry measurements were performed using Accuri C6 flow cytometer (Becton Dickinson) and all samples were analyzed at twelve hour (data not shown) and twenty-four hour time points with 488 nm excitation and 530 ± 15 nm emission detection for GFP. 20,000 individual cells were analyzed for each sample at a slow flow rate. All fluorescence distributions were unimodal, and data with obvious errors was excluded for analysis. Data files were further analyzed by MATLAB (MathWorks).

Table 7: iGEM Registry of standard biological parts used in this study.

Biobrick number	Abbreviation in the paper	Gene description
BBa_J23104	CP	Constitutive promoter family member
BBa_B0034	RBS	Ribosome binding site
BBa_B0015	T	Transcriptional terminator (Double)
BBa_E0040	GFP	Green fluorescent protein
BBa_C0040	TetR	tetracycline repressor from transposon Tn10 (+LVA)
BBa_C0012	LacI	lacI repressor from E. coli (+LVA)
BBa_C0071	RhIR	rhlR repressor/activator from P. aeruginosa PA3477 (+LVA)
BBa_C0062	LuxR	LuxR repressor/activator
BBa_C0161	LuxI	Autoinducer synthetase for AHL (no LVA)
BBa_C0050	CI HK022	cI repressor from phage HK022
BBa_C0053	C2 P22	c2 repressor from Salmonella phage P22 (+LVA)
BBa_C2001	Zif23-GCN4	Zif23-GCN4 engineered repressor (+LVA, C2000 codon-optimized for E. coli)
BBa_C0056	CI 434	cI repressor from phage 434
BBa_C0170	RhII	Autoinducer synthetase for N-butyryl-HSL (BHL) and HHL
BBa_J45014	ATF1-1148 mutant	Alcohol acetyltransferase I; converts isoamyl alcohol to isoamyl acetate
BBa_C0080	AraC	araC arabinose operon regulatory protein (repressor/activator) from E. coli (+LVA)
BBa_C0178	LasI	Autoinducer synthetase for PAI from Pseudomonas aeruginosa
BBa_C0052	CI 434-LVA	cI repressor from phage 434 (+LVA)
BBa_C0072	Mnt	mnt repressor (strong) from Salmonella phage P22 (+LVA)
BBa_C0179	LasR	lasR activator from P. aeruginosa PAO1
BBa_K863001	BPUL	bpul laccase from Bacillus pumilus

BBa_K863006	ECOL	ecol laccase from E. coli
BBa_K863021	BHAL	bhal laccase from Bacillus halodurans
BBa_K105007	Gal4	Gal4 - DNA binding domain
BBa_K165009	LexA	LexA DNA-binding domain
BBa_K079015	LacY	LacY transporter protein from E. coli
BBa_K747000	TAL	TAL-Protein_AA1_DiRepeat
BBa_J06504	mCherry2	monomeric RFP optimized for bacteria
BBa_K629005	TrkD	trkD, a functional Kup (formerly TrkD) system took up Cs+ with a moderate rate and affinity
BBa_K118001	AppY	appY coding sequence encoding a DNA-binding transcriptional activator
BBa_J52035	dnMyD88	Dominant negative form of MyD88
BBa_K165006	Zif268-HIV	Zif268-HIV DNA-binding domain
BBa_E1010	mRFP1	Highly engineered mutant of red fluorescent protein from Discosoma striata
BBa_E0020	ECFP	Engineered cyan fluorescent protein derived from A. victoria GFP
BBa_K592009	AmilCP	amilCP, blue chromoprotein
BBa_C0074	PenI	penI repressor from Bacillus licheniformis (+LVA) Integral protein present in inner mitochondrial membrane
BBa_K141000	Ucp1	
E344015	GFP Generator	RBS-GFP-Terminator

Table 8: Transcriptional factors amplified from *E. coli* wild type strain.

Protein name	Gene size (bp)	Abbreviation in the paper	Gene description
MarA	384	Tr1	Multiple antibiotic resistance transcriptional regulator
MazF	336	Tr2	mRNA interferase toxin, antitoxin is MazE
IdnR	999	Tr3	Transcriptional repressor, 5-gluconate-binding
AsnC	459	Tr4	Transcriptional activator of <i>asnA</i> ; autorepressor
MprA	531	Tr5	Transcriptional repressor of microcin B17 synthesis and multidrug efflux
RbsR	993	Tr6	Transcriptional repressor of ribose metabolism
ArcA	717	Tr7	Response regulator in two-component regulatory system with ArcB or CpxA
CpxR	699	Tr8	Response regulator in two-component regulatory system with CpxA
SoxR	465	Tr9	Redox-sensitive transcriptional activator of <i>soxS</i> ; autorepressor
CytR	1026	Tr10	Anti-activator for CytR-CRP nucleoside utilization regulon
IscR	489	Tr11	<i>isc</i> operon transcriptional repressor; <i>suf</i> operon transcriptional activator; oxidative stress- and iron starvation-inducible; autorepressor
GlpR	758	Tr12	Repressor of the glycerol-3-phosphate regulon
CynR	900	Tr13	Transcriptional activator of <i>cyn</i> operon; autorepressor
PdhR	765	Tr14	Pyruvate dehydrogenase complex repressor; autorepressor
MetR	954	Tr15	Methionine biosynthesis regulon transcriptional regulator
IlvY	894	Tr16	Transcriptional activator of <i>ilvC</i> ; autorepressor
BetI	588	Tr17	Choline-inducible <i>betIBA-betT</i> divergent operon transcriptional repressor

4.4.4 RT-qPCR.

Total RNA was extracted from three individual cell cultures (1.5 mL exponentially growing cell cultures, fresh cultures) for each construct in Fig. 1b using Trizol (Invitron). “Control” is without *X* gene in the circuit, i.e. a constitutive promoter drives GFP expression (Constitutive promoter-RBS-GFP). DNase I (NEB) was used to remove traces of genomic DNA and then the total RNA was further purified using purelink RNA Mini Kit (Life technologies), and the eluted total RNA was quantified

using BioTek's Synergy H1multi-mode Reader (A_{260}/A_{280} are 1.91 ~ 2.13, the concentrations are 67.45 ~118.44 ng/ μ L). After DNase I treatment, 2 μ L RNA samples were loaded on 1% agarose gel (Biosciences) for electrophoresis to check the DNA contamination. cDNA was synthesized from RNA using an iScript cDNA synthesis kit and random primers (Cat.# 1708841, Bio-Rad). The reaction volume is 20 μ L and ~1 μ g RNA were used for reaction. The procedure for the reverse transcription is: 25°C 5 min for priming, then 46°C 20 min for reverse transcription, and finally 95°C 1 min for inactivation. Concentrations of cDNA are then quantified by qPCR using iTaq Universal SYBR Green Supermix (Cat.# 172-5120, Bio-Rad) with the iQ5 Real-Time PCR detection system (CFX384TM, Bio-Rad). Prokaryotic 16S rRNA was employed as endogenous control. The total reaction volume is 10 μ L, with 0.1 μ L cDNA. The procedure for the qPCR is: initial denaturation 95°C for 2 mins, followed by 40 cycles of 95°C 5 s (denaturation), and 60°C for 30 s (annealing, extension and fluorescence reading). Primers (IDT) used for amplifying 16S rRNA: 5'-GAATGCCACGGTGAATACGTT-3' (*rrnB*, forward, starting at the 1361st nucleotide), and 5'-CACAAAGTGGTAAGCGCCCT-3' (*rrnB*, reverse, starting at the 1475th nucleotide)¹⁶⁷. Two pairs of primers were designed to amplify GFP are P1: 5'-CAGTGGAGAGGGTGAAGGTGA-3' (forward, starting at the 87th nucleotide); and P2: 5'-CCTGTACATAACCTTCGGGCAT-3' (reverse, starting at the 283th nucleotide); P3: 5'-AGACACGTGCTGAAGTCAAG-3' (forward, starting at the 320th nucleotide); and P4: 5'-TCTGCTAGTTGAACGCTTCCAT-3' (reverse, starting at the 539th nucleotide). qPCR result is analyzed using Bio-rad CFX Manager software version 3.1. Each sample was performed two replicates for both 16S rRNA and GFP cDNAs, and gene expression was

normalized to 16S rRNA. Delta C_t values were calculated ($C_t^{\text{target}} - C_t^{16S}$) and compared with the biological control (*Constitutive promoter-RBS-GFP*) to calculate the relative GFP mRNA concentrations.

4.4.5 Sample preparation and microscopy.

Single colonies were picked and grew at 37°C in liquid LB medium. After 24 hours, 1 mL cells were collected and spun down at 2500 g for 5 min, washed with 1x phosphate buffer solution (PBS), and resuspended by 200 μ L 1xPBS. 5 μ L of concentrated cell solution was placed on glass microscope slides and images were captured with a Nikon Ti-Eclipse inverted microscope (magnification 40x). GFP was visualized with an excitation at 472 nm and emission at 520/35 nm using a Semrock band-pass filter. The exposure time for each sample is kept the same.

4.4.6 Hysteresis experiment.

All synthetic toggle switch plasmids (T_S28, T_S67 and T_WT) were transformed into K-12 MG1655 strain with *lacI*^{-/-}, and cells cultured overnight in LB medium. For OFF-ON experiment, samples were diluted evenly into 5 ml polypropylene round-bottom tubes (Falcon) and induced with different amounts of aTc. Fluorescence was then measured at 6, 8 and 21 hr time points to monitor the fluorescence level. In our experiment, the intensity of fluorescence became stable after ~8 hr induction. For the On-OFF experiment, cells were induced with 40 ng/ml aTc initially and fluorescence was measured at 8 hr to ensure they were fully induced. Cells were then collected by low-speed centrifugation, washed once, resuspended with LB medium, diluted and transferred

into fresh medium with various aTc concentrations at 1:100 ratio. Flow cytometry measurement was performed for each sample after 6, 10 and 18 hr culturing, respectively. Data shown in Fig. 7 are 18 hr results.

4.4.7 Minimum free energy calculation.

All minimum free energy (MFE) of mRNAs were computed on Nucleic Acid Package (NUPACK) web server (<http://www.nupack.org>). $\Delta G_{5'ATR}$ and $\Delta G_{3'ATR_{100}}$ were calculated from sequence including ATR (with or without RBS), and the two scar sequences introduced during cloning process. $\Delta G_{-70\sim+38}$ is obtained from 70 nt upstream sequence and 38 nt downstream around ATG (+1) codon of GFP gene. For analysis, the value is first transformed into positive number and then changed to log scale.

4.5 Statistical analysis and Mathematical Modeling

4.5.1 Statistical analysis and model fitting

To investigate the correlate on between GFP expression and sequence characteristics in different circuits with different genes and organizations, we performed multiple linear regression analysis using the classical statistical software SAS 9.4. Here, we mainly focused on five different independent variables including 5'- and 3'-ATR GC content (or $\Delta G_{5'ATR}$ and $\Delta G_{3'ATR_{100}}$), size, and $\Delta G_{-70\sim+38}$, all of which can be computed from the DNA sequence in each circuit. The dependent variable is GFP fluorescence measured by flow cytometry, which was transformed to log scale during analysis. At least 8 data points from three independent measurements were acquired for each construct, and error bar: mean \pm standard deviation (s.d.) unless specified. All of the

collected data points are imported to SAS for analysis.

All the information of the five variables is calculated from the specific DNA sequence. The 5'ATR includes the sequence from the scar right after the promoter to the scar right before the RBS of GFP. And the 3'ATR includes the sequence from the scar right after the GFP to the scar right before the terminator. The scar sequence is generated from the molecular cloning using biobrick modules, and the size is 6 or 8 nucleotides. GC content and size of ATRs are calculated using the web server Endmemo (<http://www.endmemo.com/index.php>). The minimum free energy of mRNAs for 5' and 3'ATR ($\Delta G_{5'ATR}$ and $\Delta G_{3'ATR_{100}}$) and $\Delta G_{-70\sim+38}$ were computed using NUPACK web tool (<http://www.nupack.org>). Since the ΔG are negative values, log transformations were performed to the absolute value of ΔG , and then set to negative value. To build a comprehensive model for all the scenarios in Figure 26 (*GFP-X*, *X-GFP-Y*, and *X-GFP*), some dummy values were introduced in the regression analyses. For example, construct *GFP-X* (Figure 26E) has no “real” 5'ATR (only 8-nt scar sequence), and the $\Delta G_{5'ATR}$ is 0, which is inappropriate for log transformation. And although the GC content for the 8-nt scar sequence is not 0, we assumed it is a small number close to 0. Similar assumptions were applied for 3'ATR. In the comprehensive model, the constant G_m is set to 1, and for cases of non-coding ATRs, the coefficients for j and $\Delta G_{3'ATR_{100}}$ are set to 0, owing to a lack of 3'ATRs.

Multiple linear regression as a standard statistical tool was then employed to discover the relationship between GFP expression and the five possible variables. The model has the form of $y = X\beta + \varepsilon$, where y is the GFP fluorescence, X is the design matrix of all predictors, and β is the slope vector including intercept, and ε is the error

vector. To find the linear model having the best prediction of dependent variable from the independent variables, we performed stepwise regression, which is an automated tool for model selection through adding the most significant variable or removing the least significant variable as needed for each step (all variables left in the model are significant at the level of 0.05).

For scenario 1 (i.e. construct *X-GFP*), results indicated that 5'ATR size, GC content and local mRNA folding energy (-70 nt ~ +38 nt region around GFP's starting codon (+1)) are necessary for the best fitting of the experimental data, and the three variables explain about 63% GFP variations (Table 9). For scenario 2 (i.e. construct *X-GFP-Y*), results indicated that only 5' and 3'ATR GC content and local mRNA folding energy are needed variables for the best fitting of the data (Table 10), but the sizes of 5' ATR and 3' ATR don't have significant effect on the model fitting, which may be because of the limited sample size and variation (5'ATR size ranging from 719 to 974 bp, and 3'ATR size is from 719 to 1187 bp). For scenario 3 (i.e. construct *GFP-X*), results indicated that 3'ATR size and GC content are necessary for the best fitting of the experimental data (Table 11).

To specify the effect of ATR scope on protein expression, we performed sliding window analysis. Our results indicated that the full length of 5'ATR or 3'ATR has the best data fitting in the comprehensive model (data not shown). For the GC content, however, we found that the GC content of the full length 5'ATR has the highest fitting efficiency (Figure 27A) and the GC content of first 100 nucleotides of 3'ATR has the

Table 9: Model selection for construct of *X-GFP* in Figure 26A.

Top: Stepwise approach was used to determine the model selection for construct *X-GFP* (all variables left in the model are significant at the level of 0.05). Result indicated that 5'ATR size, GC content and local mRNA folding energy (-70 nt ~ +38 nt) are necessary for the best fitting of the experimental data. Middle: Analysis of variance for the selected regression model, and the three variables explain 62.6% variation of GFP expression. Bottom: Parameter estimates and standard errors for the three predictors and intercept. N_ATR_GC: 5'ATR GC content; N_ATR_Size: 5'ATR size; dG: mRNA folding energy (-70 nt ~ +38 nt).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	N_ATR_GC2		1	0.4432	0.4432	134.929	221.25	<.0001
2	N_ATR_Size2		2	0.1537	0.5969	23.5085	105.60	<.0001
3	dG2		3	0.0291	0.6260	4.0000	21.51	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	58.37687	19.45896	153.99	<.0001
Error	276	34.87715	0.12637		
Corrected Total	279	93.25402			

Root MSE	0.35548	R-Square	0.6260
Dependent Mean	5.29282	Adj R-Sq	0.6219
Coeff Var	6.71628		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	8.47700	0.48576	17.45	<.0001	38.48279
N_ATR_Size2	1	-1.10829	0.10270	-10.79	<.0001	14.71730
N_ATR_GC2	1	3.35535	0.32741	10.25	<.0001	13.27176
dG2	1	-1.15293	0.24860	-4.64	<.0001	2.71795

Table 10: Model selection for construct of *X-GFP-Y* in Figure 26C.

Top: Stepwise approach was used to determine the model selection for construct *X-GFP-Y* (all variables left in the model are significant at the level of 0.05). Result indicated that only 5' and 3' ATR GC content and local mRNA folding energy (-70 nt ~ +38 nt) are necessary for the best fitting of the experimental data. 5' and 3' ATR size don't have significant effect on the model. Middle: Analysis of variance for the selected regression model, and the three variables explain 56% variation of GFP expression. Bottom: Parameter estimate and standard error for the three predictors and intercept. N_ATR_GC: 5' ATR GC content; C_ATR_GC: 3' ATR GC content; dG: mRNA folding energy (-70 nt ~ +38 nt).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	N_ATR_GC2		1	0.5127	0.5127	18.3742	166.27	<.0001
2	C_ATR_GC2		2	0.0289	0.5417	10.0171	9.91	0.0020
3	dG2		3	0.0186	0.5603	5.3454	6.61	0.0110

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7.86406	2.62135	66.27	<.0001
Error	156	6.17057	0.03955		
Corrected Total	159	14.03463			

Root MSE	0.19888	R-Square	0.5603
Dependent Mean	4.90823	Adj R-Sq	0.5519
Coeff Var	4.05206		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	3.99949	0.68062	5.88	<.0001	1.36586
N_ATR_GC2	1	1.05181	0.36866	2.85	0.0049	0.32198
C_ATR_GC2	1	0.49362	0.14939	3.30	0.0012	0.43188
dG2	1	-1.12596	0.43779	-2.57	0.0110	0.26164

Table 11: Model selection for construct of *GFP-X* in Figure 26E.

Top: Stepwise approach was used to determine the model selection for construct *GFP-X* (all variables left in the model are significant at the level of 0.05). Result indicated that 3'ATR size and GC content are necessary for the best fitting of the experimental data. Bottom left: Regression analysis performed with 3'ATR size and the whole 3'ATR GC content. The two variables explain 71.5% variation of GFP expression. Bottom right: Regression analysis performed with 3'ATR size and the 3'ATR GC content for the first 100 nt. The two predictors explain 78.6% GFP variation. Parameter estimates and standard errors for the two predictors and intercept are also listed. C_ATR_Size: 3'ATR size; C_ATR_GC: 3'ATR GC content. 3'ATR GC content (the closest 100 nt after GFP sequence).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	C_ATR_Size2		1	0.5752	0.5752	94.1523	257.27	<.0001
2	C_ATR_GC2		2	0.1402	0.7154	3.0000	93.15	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	16.46672	8.23336	237.60	<.0001
Error	189	6.54925	0.03465		
Corrected Total	191	23.01596			

Root MSE	0.18615	R-Square	0.7154
Dependent Mean	5.18638	Adj R-Sq	0.7124
Coeff Var	3.58922		

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	18.08752	9.04376	346.82	<.0001
Error	189	4.92844	0.02608		
Corrected Total	191	23.01596			

Root MSE	0.16148	R-Square	0.7859
Dependent Mean	5.18638	Adj R-Sq	0.7836
Coeff Var	3.11358		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	9.01666	0.18100	49.81	<.0001	85.99013
C_ATR_Size2	1	-1.07864	0.05815	-18.55	<.0001	11.92244
C_ATR_GC2	1	1.57954	0.16366	9.65	<.0001	3.22792

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	8.79089	0.15348	57.28	<.0001	85.54765
C_ATR_Size2	1	-0.96731	0.05164	-18.73	<.0001	9.14805
C_ATR_100_GC2	1	1.64200	0.12042	13.64	<.0001	4.84872

best fitting efficiency (Figure 27B). These results suggest that the protein expression is influenced by ATRs, globally for the transcription size and partially for the GC content. We also performed the regression with the first 100 nucleotides of 3'ATR GC content for construct *GFP-X*, and found the two variables (3'ATR size and 3'ATR GC content (100 bp) can explain 79% variation of GFP expression (Table 11).

Finally, a combined model combining the three scenarios was then constructed with the variables 5' ATR size and GC content, 3' ATR size and GC content (100 bp), and local mRNA free folding energy. Results showed that all the five variables are needed for the best prediction of GFP expression in the combined model, and explains nearly two-thirds of GFP variation in those synthetic circuits (Figure 26G and Table 12). The fitting diagnostics also indicated that there is no apparent trend for the residuals, and the data is roughly normally distributed, and the variables in the model explain most variation in the response variable from the residual-fit result (Figure 35).

The predicted value by observed GFP plot (Predicted Value - logGFP) reveals a reasonably successful model for explaining the variation in GFP for most of the circuits (Figure 26G and Figure 35). The predicted responses (logGFP value) are calculated according to the generated linear regression model, with the corresponding inputs from each circuit. And the plots of predicted GFP against experimentally observed GFP values are then generated to evaluate and visualize the model-fitting efficacy (Figure 26B,D,F,G, and Figure 28, and Figure 30C-D). If the model predicted values and observed values agreed perfectly ($R^2 = 100\%$), all the data points would fall on the dotted diagonal line of the squares. However, several outliers are also observed and some observations with high leverages might also be overly influencing the fit result (Figure 36). Of the outliers, most of them are corresponding to specific circuits, such as outliers 217~224 corresponding to the tricistronic circuit (*promoter-luxR-appY-GFP*, has 8 data points). Observations with high leverages such as 113~120 are corresponding to the circuit *promoter-mnt-GFP*. Moreover, some outliers are also high-leverage observations. Given the data sample size ($N = 632$), the original data collection, and the overall data-fitting

Table 12: A combined linear model for all the constructs in Figure 28.

Top: Stepwise approach was used to determine the model selection for all the three scenarios (variables left in the model are significant at the level of 0.05). Result indicated that 5'ATR size and GC content, 3'ATR size and GC content (the first 100 nt of 3'ATR), and local mRNA folding energy (-70 nt ~ +38 nt) are necessary for the best fitting of the experimental data. Middle: Analysis of variance for the selected regression model, and the five variables explain 64% variation of GFP expression. Bottom: Parameter estimates and standard errors for the five predictors and intercept. N_ATR_GC: 5'ATR GC content; N_ATR_Size: 5'ATR size; C_ATR_Size: 3'ATR size; C_ATR_100_GC: 3'ATR GC content (100 nt); dG: mRNA folding energy (-70 nt ~ +38 nt).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	C_ATR_Size2		1	0.0725	0.0725	996.229	49.25	<.0001
2	dG2		2	0.1576	0.2301	722.263	128.75	<.0001
3	N_ATR_GC2		3	0.1639	0.3940	437.310	169.80	<.0001
4	N_ATR_Size2		4	0.1725	0.5664	137.305	249.38	<.0001
5	C_ATR_100_GC2		5	0.0761	0.6425	6.0000	133.31	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	93.49738	18.69948	225.04	<.0001
Error	626	52.01689	0.08309		
Corrected Total	631	145.51427			

Root MSE	0.28826	R-Square	0.6425
Dependent Mean	5.16312	Adj R-Sq	0.6397
Coeff Var	5.58307		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	11.25218	0.41710	26.98	<.0001	60.47397
N_ATR_Size2	1	-1.17592	0.06925	-16.98	<.0001	23.95994
C_ATR_Size2	1	-0.93803	0.05537	-16.94	<.0001	23.84512
N_ATR_GC2	1	2.38728	0.10323	23.13	<.0001	44.43909
C_ATR_100_GC2	1	1.28908	0.11165	11.55	<.0001	11.07687
dG2	1	-0.92468	0.13171	-7.02	<.0001	4.09531

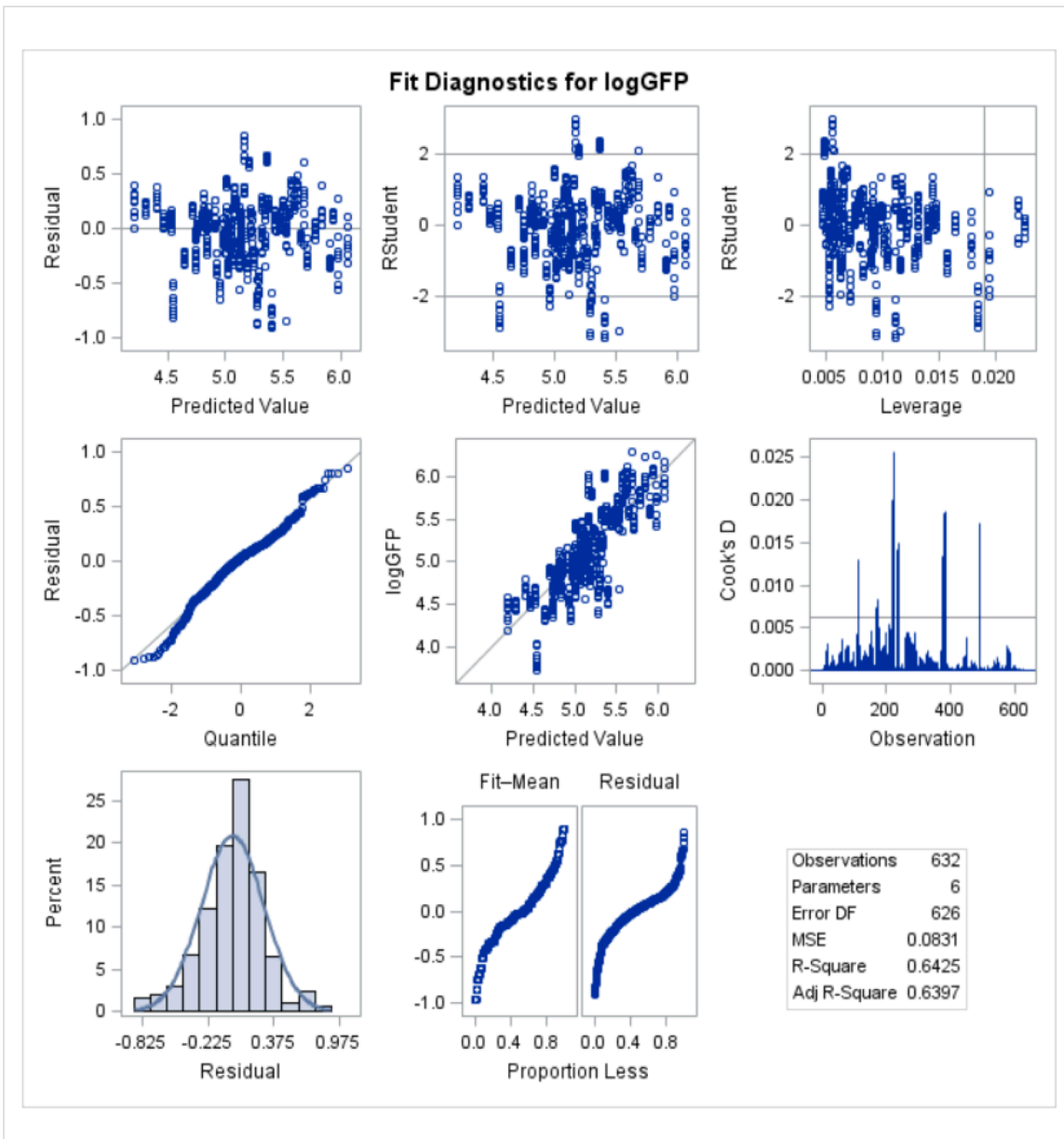


Figure 35: Fit diagnostics for the GFP in the combined model integrating the three scenarios in Figure 26. The Predicted value-Residual plot indicates that there is no apparent trend for the residuals, and the data is roughly normally distributed (Quantile-Residual plot and histogram), and the variables in the model explain most variation in the response variable from the residual-fit result (Fit-Mean and Residual). Leverage-RStudent plot and Cook's D value indicate there are some outliers and high-leverage observations, which may influence the model. Overall, the generated model has a good fitting of the experimental data.

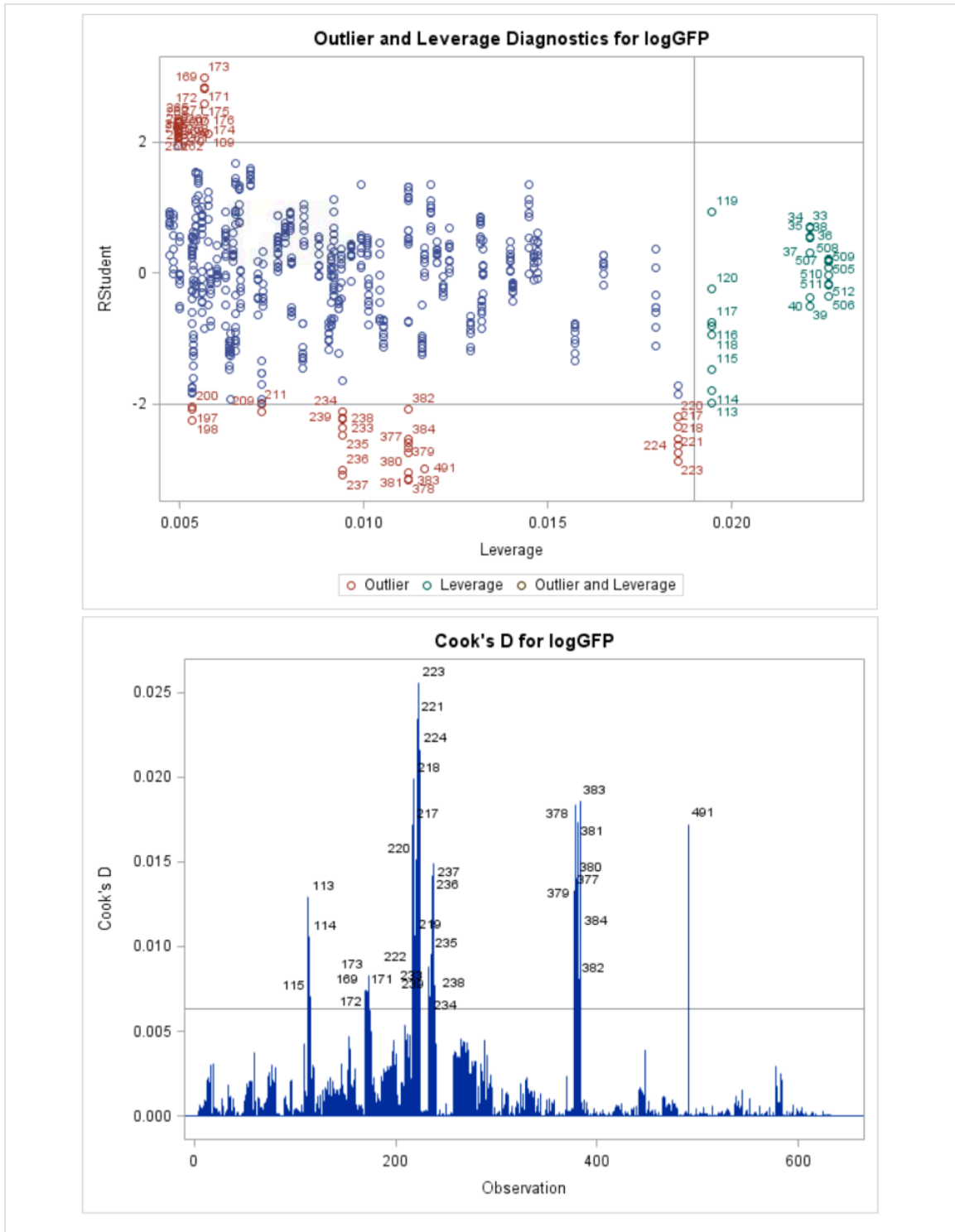


Figure 36: Outlier and leverage diagnostics for the response (GFP) in the combined model. High-leverage data points and outliers are labeled out. Of the outliers, most of them are corresponding to a specific circuit, such as outliers 217~224 corresponding to the tricistronic circuit (*promoter-luxR-appY-GFP*, has 8 data points). Observations with high leverage such as 113~120 are corresponding to the circuit *promoter-mnt-GFP*. Moreover, some outliers are also high-leverage observations.

efficacy, we here didn't exclude the outliers or data with very high leverages (although that would improve the model-fitting efficacy).

For the noncoding ATR regulation in Figure 28, similar statistical analysis was performed. Stepwise regression showed the 5'ATR size and GC content, and the local mRNA folding energy are crucial for the model fitting, and the three could explain 64.8% variation of GFP expressions in those synthetic circuits (Table 13).

To have a general understanding of ATR regulation, we next developed a comprehensive linear model based on the sequence-dependent energetic changes during the polycistronic mRNA folding and translation and the costs of protein biosynthesis. The biophysical model was based on previous pioneer work characterizing the relationship between free energy changes and protein translation initiation^{203,205}. We calculated the free energies for 5'ATR and the first 100 nucleotides of 3'ATR ($\Delta G_{5'ATR}$ and $\Delta G_{3'ATR_{100}}$) using NUPACK. Since all the energy terms are negative values, absolute values were first acquired for each of them and then set to negative values for data analysis. Stepwise regression was performed with the five variables: $\Delta G_{5'ATR}$, $\Delta G_{3'ATR_{100}}$, 5'ATR size, 3'ATR size and $\Delta G_{-70\sim+38}$. Results showed that all the five variables are necessary for the best fitting of experimental data with R^2 0.64 (Table 14). It is necessary to note that the negative correlation between protein abundance (c) and the sum of energetic terms ($\sum \beta_x \Delta G_x$) in the equation is already reflected in the coefficients of each term. Similar analysis was also applied to the data with noncoding ATR, and results showed that 5'ATR size and folding energy $\Delta G_{5'ATR}$, local mRNA folding energy $\Delta G_{-70\sim+38}$ are crucial for the best fitting of the experimental data (Table 15).

Table 13: Model selection for noncoding ATR constructs in Figure 28.

Top: Stepwise approach was used to determine the model selection for the noncoding 5'ATR regulation (variables left in the model are significant at the level of 0.05). Result indicated that 5'ATR size and GC content, local mRNA folding energy (-70 nt ~ +38 nt) are crucial for the best fitting of the experimental data. Middle: Analysis of variance for the selected regression model, and the three variables explain 64.8% variation of GFP expression. Bottom: Parameter estimates and standard errors for variables and intercept. N_ATR_GC: 5'ATR GC content; N_ATR_Size: 5'ATR size; dG: mRNA folding energy (-70 nt ~ +38 nt).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	N_ATR_Size2		1	0.2766	0.2766	276.941	100.94	<.0001
2	dG2		2	0.2832	0.5598	67.9446	169.21	<.0001
3	N_ATR_GC2		3	0.0885	0.6483	4.0000	65.94	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	96.58107	32.19369	161.00	<.0001
Error	262	52.38941	0.19996		
Corrected Total	265	148.97048			

Root MSE	0.44717	R-Square	0.6483
Dependent Mean	5.24196	Adj R-Sq	0.6443
Coeff Var	8.53055		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	13.36939	0.75475	17.71	<.0001	62.74170
N_ATR_Size2	1	-3.84724	0.21404	-17.97	<.0001	64.60441
N_ATR_GC2	1	3.50376	0.43146	8.12	<.0001	13.18626
dG2	1	-3.18689	0.35054	-9.09	<.0001	16.52706

Table 14: A comprehensive model with energetic changes for the constructs in Figure 26. Top: Stepwise approach was used to determine the model selection for all the 79 constructs (variables left in the model are significant at the level of 0.05). Result indicated that $\Delta G_{5'ATR}$, $\Delta G_{3'ATR_{100}}$, 5'ATR size, 3'ATR size and $\Delta G_{-70\sim+38}$ are necessary for the best fitting of the experimental data. Middle: Analysis of variance for the selected regression model, and the five variables explain 63% variation of GFP expression. Bottom: Parameter estimates and standard errors for the five predictors and intercept. N_ATR_dG: $\Delta G_{5'ATR}$; N_ATR_Size: 5'ATR size; C_ATR_Size: 3'ATR size; C_ATR_100_dG: $\Delta G_{3'ATR_{100}}$; dG: mRNA folding energy ($\Delta G_{-70\sim+38}$).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	C_ATR_Size2		1	0.0725	0.0725	946.159	49.25	<.0001
2	dG2		2	0.1576	0.2301	680.700	128.75	<.0001
3	C_ATR_100_dG2		3	0.0903	0.3203	529.525	83.39	<.0001
4	N_ATR_dG2		4	0.0566	0.3769	435.542	56.91	<.0001
5	N_ATR_Size2		5	0.2543	0.6312	6.0000	431.54	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	91.84285	18.36857	214.24	<.0001
Error	626	53.67142	0.08574		
Corrected Total	631	145.51427			

Root MSE	0.29281	R-Square	0.6312
Dependent Mean	5.16312	Adj R-Sq	0.6282
Coeff Var	5.67116		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	11.46350	0.43155	26.56	<.0001	60.49930
N_ATR_Size2	1	-3.10443	0.14944	-20.77	<.0001	36.99917
C_ATR_Size2	1	-1.15864	0.07953	-14.57	<.0001	18.19507
N_ATR_dG2	1	-1.80151	0.07964	-22.62	<.0001	43.86968
C_ATR_100_dG2	1	-0.28049	0.02672	-10.50	<.0001	9.44495
dG2	1	-1.12060	0.13274	-8.44	<.0001	6.10990

Table 15: A comprehensive model with energetic changes for noncoding ATR constructs in Figure 28. Top: Stepwise approach was used to determine the model selection for the noncoding 5'ATR regulation (variables left in the model are significant at the level of 0.05). Result indicated that 5'ATR size and folding energy, local mRNA folding energy (-70 nt ~ +38 nt) are crucial for the best fitting of the experimental data. Middle: Analysis of variance for the selected regression model, and the three variables explain 67.4% variation of GFP expression. Bottom: Parameter estimate and standard error for variables and intercept. N_ATR_dG: 5'ATR folding energy; N_ATR_Size: 5'ATR size; dG: mRNA folding energy (-70 nt ~ +38 nt).

Summary of Stepwise Selection								
Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	N_ATR_Size2		1	0.2766	0.2766	318.842	100.94	<.0001
2	dG2		2	0.2832	0.5598	93.4417	169.21	<.0001
3	N_ATR_dG2		3	0.1139	0.6737	4.0000	91.44	<.0001

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	100.36041	33.45347	180.31	<.0001
Error	262	48.61007	0.18553		
Corrected Total	265	148.97048			

Root MSE	0.43074	R-Square	0.6737
Dependent Mean	5.24196	Adj R-Sq	0.6700
Coeff Var	8.21710		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type II SS
Intercept	1	14.45056	0.76488	18.89	<.0001	66.22312
N_ATR_Size2	1	-7.27722	0.42999	-16.92	<.0001	53.14271
N_ATR_dG2	1	-3.23089	0.33787	-9.56	<.0001	16.96561
dG2	1	-3.39095	0.32107	-10.56	<.0001	20.69459

In summary, we have demonstrated that the 5' and 3' adjacent transcription regions have remarkable effects on GFP expression from the results in Figure 26 and Figure 28. Furthermore, we can use a general model with sequence-dependent energetic changes to explain the ATR regulation on gene expression. In this study, we mainly investigated five factors involved in ATR regulation: 5' and 3' ATRs free energies $\Delta G_{5'ATR}$ and $\Delta G_{3'ATR_{100}}$, transcriptional sizes and the mRNA folding energy near the GFP starting codon. It is possible that there are some other unknown or uncharacterized factors influencing GFP expression, such as the codon degeneracy for the coding ATRs. Furthermore, there may have some special local secondary or higher structures in some ATRs, which may impact the degradation or translation of GFP.

4.5.2 Deterministic Model construction and prediction for the logic gate

In the four logic gates, GFP expression depends on the relative concentrations of activator (LuxR) and repressor (TetR or LacI) produced from a constitutive promoter. AHL binds with LuxR protein to activate *pLux/tet* transcription and aTc can block TetR repression to *pLux/tet*. Since the two sets of logic gates (LT/TL and LI and IL) are constructed similarly and described by the same deterministic equations, we here only explain the technical details for the gate LT. The model was built based on our previous work³⁹. From the biochemical reactions depicted in Fig. 5a, we derived the following ordinary differential equations for intercellular concentrations of LuxR (U), TetR (R) and GFP (G):

$$\frac{dU}{dt} = (k_0 + \alpha_1) - d_1 \cdot U \quad (1)$$

$$\frac{dR}{dt} = (k_0 + \alpha_2) - d_2 \cdot R \quad (2)$$

$$\frac{dG}{dt} = \left(c_1 + \frac{K_1 C}{C + K_n} \right) \cdot \frac{1}{K_r^{nt} + (R \cdot F)^{nt}} - d_3 \cdot G \quad (3)$$

and

$$f = \frac{AHL^{ni}}{AHL^{ni} + K_i^{ni}} \quad (4)$$

$$C = \frac{(f \cdot U)^2}{K_d} \quad (5)$$

$$F = \frac{1}{K_t^{nr} + ATC^{nr}} \quad (6)$$

The first two equations describe the concentrations of LuxR and TetR, both of which are driven by a constitutive promoter at a constant level (k_0). α_1 and α_2 are constants used to describe the relative changes of LuxR and TetR production, owing to the position changes in the And-gate circuit. d_1 and d_2 are the degradation rates for the LuxR and TetR protein, respectively. The third equation describes the concentration of GFP, which is determined by the relative concentrations of LuxR and TetR. LuxR binds to AHL molecules and forms the active LuxR monomers in the form of (LuxR-AHL), when the AHL concentration reaches a certain threshold (quorum-sensing mechanism). So the fraction of LuxR monomers (f) bound by AHL can be described by Eq4, where ni is the binding cooperativity (Hill coefficient) between LuxR and AHL, and K_i represents the dissociation constant between LuxR and AHL. LuxR needs to form a dimer to bind

the promoter and activate transcription, so the concentration of the functional LuxR dimer (C) that binds to the hybrid promoter $pLux/tet$ and activates its transcription can be described by Eq5, where K_d is the dissociation constant for dimerization. Thus, GFP expression driven by LuxR and inducer AHL is represented by the first part of Eq3. c_l is the basal mRNA expression without LuxR protein; K_l is the production rate; and K_n is the dissociation constant between C and $pLux/tet$ promoter. TetR protein can bind and inhibit GFP transcription, and the inhibition can be repressed by inducer aTc. So high GFP expression is achieved in presence of high doses of aTc, and vice versa (Eq6). The second part of Eq3 describes TetR inhibition to GFP expression, under induction of aTc. And the third part of Eq3 is the degradation of GFP.

The three ordinary differential equations were used to model the two sets of AND-gate circuits: LT and TL, LI and IL. For each of the two sets, most parameters should be the same except α_1 , α_2 , c_l , and K_i . Based on the parameter used in our previous studies³⁹, we used the following parameters in our simulations: $k_0 = 1.0$, $d_l = 0.2$, $d_2 = 0.2$, $d_3 = 0.2$, $c_l = 0.002$ (for TL) or 0.08 (for LT), $K_l = 1.7$, $K_n = 4.4$, $K_d = 13$, $K_t = 400$, $K_r = 3.2$, $n_i = 1.2$, $n_t = 2$, $n_i = 1.2$, $n_r = 2$. For circuits LI and IL, $c_l = 0.002$ (for IL) or 0.05 (for LI), $K_t = 1000$, and the other parameters are the same.

From our comprehensive linear model, we calculated that LT has more LuxR than TetR production (Table 6), so the basal expression c_l is set to a bigger value in LT model. K_i has little effect on the shape of the GFP dynamic curves, but determines the AHL concentration producing half conversion of LuxR monomers into LuxR-HSL complexes (half GFP activation). So the K_i value in the model is acquired from the experimental data. Through changing the relative expression of LuxR and TetR (i.e. α_1 and α_2), we can

modulate GFP production dynamics (Figure 31). To predict the GFP responses in circuit TL with AHL and aTc inductions, we use the parameter α_1 and α_2 in LT as a control to tune the parameter α_1 and α_2 in TL. According to the linear model calculations, the production rate for LuxR in LT and TL almost doesn't change, but production rate of TetR in TL increases by ~93% (Table 6). For example, we set the production rates for LuxR and TetR in circuit LT to $1.0 (k_0 + \alpha_1)$ and $0.6 (k_0 + \alpha_2)$, respectively. So in the circuit TL, the two rates should be $1.0 (k_0 + \alpha_1)$ and $1.15 (k_0 + \alpha_2)$ based on calculations. For different doses of aTc induction, we allowed ~10% parameter variations for α_1 and α_2 . We found that the model simulations have a good match with our experimental data. The parameters for α_1 and α_2 in TL and LT under different doses of aTc are listed below:

Circuit	aTc (0 ng/ml)	aTc (0.2 ng/ml)	aTc (2 ng/ml)	aTc (20 ng/ml)	aTc (100 ng/ml)	aTc (200 ng/ml)
LT	$\alpha_1 = 0$ $\alpha_2 = -0.3$	$\alpha_1 = 0$ $\alpha_2 = -0.4$	$\alpha_1 = 0$ $\alpha_2 = -0.38$	$\alpha_1 = 0$ $\alpha_2 = -0.3$	$\alpha_1 = 0$ $\alpha_2 = -0.35$	$\alpha_1 = 0$ $\alpha_2 = -0.25$
TL	$\alpha_1 = 0.1$ $\alpha_2 = 0.1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.05$	$\alpha_1 = 0.1$ $\alpha_2 = 0$	$\alpha_1 = 0.1$ $\alpha_2 = 0.1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.1$	$\alpha_1 = 0.1$ $\alpha_2 = 0.25$

Compared to circuit LI, the production rate for LuxR in IL increases by ~74%, and ~38% for LacI (the overall inhibition efficiency may increase by ~76%, Supplementary Table 12). For example, we set the production rates for LuxR and LacI in circuit LI to $1.0 (k_0 + \alpha_1)$ and $0.8 (k_0 + \alpha_2)$, respectively. So in the circuit IL, the two rates should be $1.74 (k_0 + \alpha_1)$ and $\sim 1.41 (k_0 + \alpha_2)$ based on calculations. For different doses of IPTG induction, we allowed ~10% parameter variations for α_1 and α_2 . And the parameters for α_1 and α_2 in LI and IL under different doses of IPTG are listed below:

Circuit	IPTG (0 μM)	IPTG (0.1 μM)	IPTG (1 μM)	IPTG (10 μM)	IPTG (100 μM)	IPTG (200 μM)	IPTG (400 μM)
LI	$\alpha_1 = 0$ $\alpha_2 = -0.1$	$\alpha_1 = 0$ $\alpha_2 = -0.12$	$\alpha_1 = 0$ $\alpha_2 = -0.1$	$\alpha_1 = 0$ $\alpha_2 = -0.15$	$\alpha_1 = 0$ $\alpha_2 = -0.25$	$\alpha_1 = 0$ $\alpha_2 = -0.28$	$\alpha_1 = 0$ $\alpha_2 = -0.2$
IL	$\alpha_1 = 0.57$ $\alpha_2 = 0.57$	$\alpha_1 = 0.6$ $\alpha_2 = 0.5$	$\alpha_1 = 0.69$ $\alpha_2 = 0.5$	$\alpha_1 = 0.82$ $\alpha_2 = 0.35$	$\alpha_1 = 0.7$ $\alpha_2 = 0.4$	$\alpha_1 = 0.87$ $\alpha_2 = 0.5$	$\alpha_1 = 0.57$ $\alpha_2 = 0.56$

4.5.3 Parameter and bifurcation analysis for the synthetic toggle switch circuit.

For the toggle switch model in Figure 29, we used the same mathematical model and most parameters in the Gardner *et al.* paper⁶. Here we think the synthetic ATRs mainly influenced the tetR production rate, with low rate corresponding to T_S28 ($\alpha_1 = 400$, $\beta = 2.7$), medium rate corresponding to T_S67 ($\alpha_1 = 600$, $\beta = 3.0$), and high rate corresponding to T_WT ($\alpha_1 = 1000$, $\beta = 3.245$). All the other parameters are set the same as in Gardner *et al.* paper. Bifurcation analyses are performed using XPP-AUTO software (www.math.pitt.edu).

CONCLUSIONS AND OUTLOOK

5.1 Conclusions

This dissertation predominantly focuses on synthetic biology projects centered on the theme of engineering synthetic gene networks with mathematical modeling to probe the underlying principles for cell fate determination in multistable systems and develop a tool to pre-evaluate protein expression levels in synthetic circuits. In the very last chapter, I am trying to summarize all the projects and research contributions to synthetic biology field, and discuss future directions of synthetic gene networks.

In the first chapter, I highlighted the main applications of synthetic gene networks in various aspects in understanding the fundamental cellular design principles, development of diagnostic bio-devices, and production of industrial and biomedical compounds. This brief introduction offers an insight into the history of synthetic biology, and helps to guide and inspire future research in the field.

My first project is to investigate quorum-sensing crosstalk and it induced distinct dynamics (Chapter 2). I systematically studied the crosstalk between LuxR/I and LasR/I systems and found that QS crosstalk can be dissected into signal crosstalk and promoter crosstalk. Further investigations using synthetic positive feedback circuits revealed that signal crosstalk significantly decreases the circuit's bistable potential while maintaining unimodality. Promoter crosstalk, however, reproducibly generates a complex trimodal response resulting from noise-induced state transitions and host-circuit interactions. This

work reveals a new hybrid multimodality arising from gene circuit, gene expression stochasticity and host-circuit interactions, which could be exploited for therapeutics and biotechnology.

The second project is about engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination (Chapter 3). In this work, I engineered and tested different synthetic two-node network topologies and verified a synthetic mutual inhibition network with positive auto-activations (MINPA) to be quadrastable. We show that cells indeed gravitate towards local minima and signal inductions dictate cell fates through modulating the shape of the multistable landscape. Experiments, guided by model predictions, reveal that sequential inductions generate distinct cell fates by changing landscape in sequence and hence navigating cells to different final states. This study serves as a proof-of-principle demonstration of the Waddington landscape and provides a synthetic biology framework to understand cell differentiation process and suggests a landscape-based explanation of fixed induction sequences for directed differentiation.

The third project is to develop a synthetic circuit performance evaluator, taking into account the properties of adjacent transcriptional regions, to calculate each gene's relative expression level in a synthetic operon-based gene circuit (Chapter 4). We demonstrated the tool's utility in guiding synthetic logic gate design, and tuning gene expression and nonlinear dynamics of bistable gene networks. This tool can save experimentalists' time and resources to test different biomodules' combinations, and thus should greatly facilitate optimization of circuit design and accelerate the engineering of complex gene networks.

5.2 Outlook

The work in this thesis illustrated applications of designed synthetic gene networks to understand cellular multistability in a predictable way. Synthetic circuitry engineering as one of the foundation technologies has helped advance the burgeoning development of synthetic biology for about twenty years. In the next decade, we expect that synthetic gene networks will become increasingly complex for more sophisticated tasks. Based on my research, I think synthetic gene networks will progress toward at least the following three directions.

First, synthetic gene networks are becoming more complex. Our results indicated that the observed multistability could be generated either from a combination of a bistable gene circuit and host-circuit interactions or from a coherently regulated MINPA circuit. Positive feedback loop and MINPA are recurring motifs in natural gene regulatory networks, especially in developmental gene regulation. Although these synthetic networks including MINPA can serve as suitable prototypes to probe the fundamental principles of cellular decision-making in complex networks, most of them are relatively simple both topologically and functionally. So how to construct experimentally complex and multi-node networks with predicted functions is a big challenge. For example, it is still difficult experimentally to build gene networks of fully connected triads with complete auto-activation (such as the Oct4-Sox2-Nanog triad), which are important to understand multistability in cell differentiation and human development¹⁹⁸. To achieve this, more standardized and well-characterized biobricks and devices will be necessary for advanced multi-layer circuits engineering. Recent study

indicated that we might use a programming language to realize the automation of synthetic circuit design and performance simulation¹⁸⁷.

Second, synthetic gene networks will find a broader application in mammalian cells. Complementary to the synthetic gene networks in single-cell organisms (bacteria and yeast cells), synthetic mammalian gene networks is essential to elucidate the complex mechanisms for cellular regulations and behaviors in higher organisms^{199,200}. Compared to prokaryotic systems, mammalian synthetic biology enables us to investigate more biological events such as epigenetics, signaling transductions, and alternative splicing. For example, synthetic biologists designed synthetic transcription factors Pc-TFs to target trimethyl-histone H3 lysine 27 and control the chromatin dynamics from a closed state to open state²⁰¹. Furthermore, mammalian synthetic biology enables to develop therapeutic interventions for human diseases. Compared to traditional drugs, synthetic gene networks have unique advantages such as sensing disease states and making corresponding therapeutic decisions in a self-sufficient manner, easy engineering and manipulation, and providing sustainable protection^{66,202,203}. Recent studies indicated that T cells engineered with chimeric antigen receptors or T cell receptors could be designed to attack cancer cells. The newly developed synthetic Notch receptors have been demonstrated higher therapeutic efficacy in precisely recognizing and clearing tumor cells^{152,204}.

Nonlinear multistability has long been proposed as the underlying mechanism that cells used to maintain pluripotency and guide differentiation^{31,137}. Although we here tried to understand the principles of achieving multistability and cell state transitions in *E. coli* cells, we are also interested in engineering synthetic multistable circuits in mammalian cells and pluripotent stem cells to investigate the fate determination process in a more

complex cell context. Also, synthetic gene circuits can be engineered to rewire the GRNs in stem cells to modulate cells' self-renewal and differentiation.

Last but not the least, synthetic gene networks will accelerate the process towards practical applications, especially in the industrial and biomedical fields. So far, engineering of biosynthetic pathways using simple microbes for large-scale and cost-effective renewable energy and pharmaceuticals is still immature and awaits further development. This will rely on our complete understanding of the natural biosynthetic pathways, metabolic flux and regulatory mechanisms²⁰⁵. On the other hand, the overexpressed networks place a metabolic burden on the host cells, so synthetic biologists need to balance the final output and the microbes' stresses for product maximization.

As a summary, synthetic biology holds promise for understanding of cell design principles, biomedical and industrial products, and development of effective diseases therapies including cancer. And synthetic gene networks will be an essential part to advance these applications and realize the full potential of synthetic biology.

REFERENCES

1. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
2. Elowitz, M. & Lim, W. A. Build life to understand it. *Nature* **468**, 889–890 (2010).
3. Oppenheim, A. B., Kobilier, O., Stavans, J., Court, D. L. & Adhya, S. Switches in Bacteriophage Lambda Development. *Annu. Rev. Genet.* **39**, 409–429 (2005).
4. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379 (2010).
5. Lu, T. K., Khalil, A. S. & Collins, J. J. Next-generation synthetic gene networks. *Nat. Biotechnol.* **27**, 1139–1150 (2009).
6. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
7. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
8. Nevozhay, D., Adams, R. M., Murphy, K. F., Josić, K. & Balázsi, G. Negative autoregulation linearizes the dose–response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci.* pnas.0809901106 (2009). doi:10.1073/pnas.0809901106
9. Friedland, A. E. *et al.* Synthetic Gene Networks That Count. *Science* **324**, 1199–1202 (2009).
10. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
11. Sohka, T. *et al.* An externally tunable bacterial band-pass filter. *Proc. Natl. Acad. Sci.* **106**, 10135–10140 (2009).
12. Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6355–6360 (2004).
13. Tabor, J. J. *et al.* A synthetic genetic edge detection program. *Cell* **137**, 1272–1281 (2009).
14. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013).
15. Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic

- gates. *Nat. Chem. Biol.* **10**, 99–105 (2014).
16. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
 17. Green, A. A., Silver, P. A., Collins, J. J. & Yin, P. Toehold switches: de-novo-designed regulators of gene expression. *Cell* **159**, 925–939 (2014).
 18. Bashor, C. J., Helman, N. C., Yan, S. & Lim, W. A. Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics. *Science* **319**, 1539–1543 (2008).
 19. Deans, T. L., Cantor, C. R. & Collins, J. J. A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. *Cell* **130**, 363–372 (2007).
 20. Padirac, A., Fujii, T. & Rondelez, Y. Bottom-up construction of in vitro switchable memories. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E3212–3220 (2012).
 21. Karlebach, G. & Shamir, R. Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* **9**, 770–780 (2008).
 22. Mukherji, S. & van Oudenaarden, A. Synthetic biology: understanding biological design from synthetic circuits. *Nat. Rev. Genet.* **10**, 859–871 (2009).
 23. Wu, M. *et al.* Engineering of regulated stochastic cell fate determination. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10610–10615 (2013).
 24. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
 25. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).
 26. Schultz, D., Wolynes, P. G., Jacob, E. B. & Onuchic, J. N. Deciding fate in adverse times: Sporulation and competence in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **106**, 21027–21034 (2009).
 27. Pomerening, J. R., Sontag, E. D. & Ferrell, J. E. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat. Cell Biol.* **5**, 346–351 (2003).
 28. Jolly, M. K. *et al.* Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7**, 27067–27084 (2016).

29. Lee, J. *et al.* Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E364–373 (2014).
30. Laurent, M. & Kellershohn, N. Multistability: a major means of differentiation and evolution in biological systems. *Trends Biochem. Sci.* **24**, 418–422 (1999).
31. Guantes, R. & Poyatos, J. F. Multistable Decision Switches for Flexible Control of Epigenetic Differentiation. *PLoS Comput Biol* **4**, e1000235 (2008).
32. Xiong, W. & Ferrell, J. E. A positive-feedback-based bistable ‘memory module’ that governs a cell fate decision. *Nature* **426**, 460–465 (2003).
33. Kramer, B. P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004).
34. Ferrell Jr., J. E. *et al.* Simple, realistic models of complex biological processes: Positive feedback and bistability in a cell fate switch and a cell cycle oscillator. *FEBS Lett.* **583**, 3999–4005 (2009).
35. Kueh, H. Y., Champhekar, A., Nutt, S. L., Elowitz, M. B. & Rothenberg, E. V. Positive Feedback Between PU.1 and the Cell Cycle Controls Myeloid Differentiation. *Science* **341**, 670–673 (2013).
36. Lee, K. E. *et al.* Positive feedback loop between Sox2 and Sox6 inhibits neuronal differentiation in the developing central nervous system. *Proc. Natl. Acad. Sci.* **111**, 2794–2799 (2014).
37. Osborn, D. P. S., Li, K., Hinitz, Y. & Hughes, S. M. Cdkn1c drives muscle differentiation through a positive feedback loop with Myod. *Dev. Biol.* **350**, 464–475 (2011).
38. Chang, D.-E. *et al.* Building biological memory by linking positive feedback loops. *Proc. Natl. Acad. Sci.* **107**, 175–180 (2010).
39. Wu, F., Menn, D. J. & Wang, X. Quorum-sensing crosstalk-driven synthetic circuits: from unimodality to trimodality. *Chem. Biol.* **21**, 1629–1638 (2014).
40. Brenner, K., Karig, D. K., Weiss, R. & Arnold, F. H. Engineered bidirectional communication mediates a consensus in a microbial biofilm consortium. *Proc. Natl. Acad. Sci.* **104**, 17300–17304 (2007).
41. Becskei, A. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
42. Sanders, T. A., Llagostera, E. & Barna, M. Specialized filopodia direct long-range transport of SHH during vertebrate tissue patterning. *Nature* **497**, 628–632 (2013).

43. Lee, A. J. & You, L. Cells Listen to Their Inner Voice. *Science* **343**, 624–625 (2014).
44. Youk, H. & Lim, W. A. Secreting and Sensing the Same Molecule Allows Cells to Achieve Versatile Social Behaviors. *Science* **343**, 1242782 (2014).
45. Srimani, J. K. *et al.* Linear Population Allocation by Bistable Switches in Response to Transient Stimulation. *PLoS ONE* **9**, e105408 (2014).
46. Prindle, A. *et al.* Rapid and tunable post-translational coupling of genetic circuits. *Nature* **508**, 387–391 (2014).
47. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial Synthesis of Genetic Networks. *Science* **296**, 1466–1470 (2002).
48. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **3**, 145 (2007).
49. Reeves, G. T., Muratov, C. B., Schüpbach, T. & Shvartsman, S. Y. Quantitative Models of Developmental Pattern Formation. *Dev. Cell* **11**, 289–300 (2006).
50. Liu, C. *et al.* Sequential Establishment of Stripe Patterns in an Expanding Cell Population. *Science* **334**, 238–241 (2011).
51. Balagaddé, F. K. *et al.* A synthetic Escherichia coli predator–prey ecosystem. *Mol. Syst. Biol.* **4**, (2008).
52. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–4 (2005).
53. Payne, S. *et al.* Temporal control of self-organized pattern formation without morphogen gradients in bacteria. *Mol. Syst. Biol.* **9**, (2013).
54. Cao, Y. *et al.* Collective Space-Sensing Coordinates Pattern Scaling in Engineered Bacteria. *Cell* **165**, 620–630 (2016).
55. Cheng, A. A. & Lu, T. K. Synthetic biology: an emerging engineering discipline. *Annu. Rev. Biomed. Eng.* **14**, 155–178 (2012).
56. Wang, B., Barahona, M. & Buck, M. A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosens. Bioelectron.* **40**, 368–376 (2013).
57. Kotula, J. W. *et al.* Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proc. Natl. Acad. Sci.* **111**, 4838–4843 (2014).
58. Pardee, K. *et al.* Paper-Based Synthetic Gene Networks. *Cell* **159**, 940–954 (2014).

59. Wu, M.-C., Law, B., Wilkinson, B. & Micklefield, J. Bioengineering natural product biosynthetic pathways for therapeutic applications. *Curr. Opin. Biotechnol.* **23**, 931–940 (2012).
60. Atsumi, S. *et al.* Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).
61. Valle-Rodríguez, J. O., Shi, S., Siewers, V. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid ethyl esters, an advanced biofuel, by eliminating non-essential fatty acid utilization pathways. *Appl. Energy* **115**, 226–232 (2014).
62. Savage, D. F., Way, J. & Silver, P. A. Defossilizing fuel: how synthetic biology can transform biofuel production. *ACS Chem. Biol.* **3**, 13–16 (2008).
63. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B. & Keasling, J. D. Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320–328 (2012).
64. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
65. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
66. Wu, F. & Wang, X. Applications of synthetic gene networks. *Sci. Prog.* **98**, 244–252 (2015).
67. Ruder, W. C., Lu, T. & Collins, J. J. Synthetic Biology Moving into the Clinic. *Science* **333**, 1248–1252 (2011).
68. Nissim, L. & Bar-Ziv, R. H. A tunable dual-promoter integrator for targeting of cancer cells. *Mol. Syst. Biol.* **6**, 444 (2010).
69. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).
70. Chen, Y. Y., Jensen, M. C. & Smolke, C. D. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc. Natl. Acad. Sci.* **107**, 8531–8536 (2010).
71. Wei, P. *et al.* Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells. *Nature* **488**, 384–388 (2012).
72. Ng, W.-L. & Bassler, B. L. Bacterial Quorum-Sensing Network Architectures. *Annu. Rev. Genet.* **43**, 197–222 (2009).

73. Miller, M. B. & Bassler, B. L. Quorum Sensing in Bacteria. *Annu. Rev. Microbiol.* **55**, 165–199 (2001).
74. LaSarre, B. & Federle, M. J. Exploiting Quorum Sensing To Confuse Bacterial Pathogens. *Microbiol. Mol. Biol. Rev.* **77**, 73–111 (2013).
75. Jayaraman, A. & Wood, T. K. Bacterial quorum sensing: signals, circuits, and implications for biofilms and disease. *Annu. Rev. Biomed. Eng.* **10**, 145–167 (2008).
76. Solano, C., Echeverz, M. & Lasa, I. Biofilm dispersion and quorum sensing. *Curr. Opin. Microbiol.* **18**, 96–104 (2014).
77. Hong, S. H. *et al.* Synthetic quorum-sensing circuit to control consortial biofilm formation and dispersal in a microfluidic device. *Nat. Commun.* **3**, 613 (2012).
78. Kobayashi, H. *et al.* Programmable cells: interfacing natural and engineered gene networks. *Proc Natl Acad Sci U S A* **101**, 8414–9 (2004).
79. Chuang, J. S., Rivoire, O. & Leibler, S. Simpson’s paradox in a synthetic microbial system. *Science* **323**, 272–275 (2009).
80. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires’. *Nature* **469**, 212–215 (2011).
81. Stevens, A. M. & Greenberg, E. P. Quorum sensing in *Vibrio fischeri*: essential elements for activation of the luminescence genes. *J. Bacteriol.* **179**, 557–562 (1997).
82. Schuster, M., Urbanowski, M. L. & Greenberg, E. P. Promoter specificity in *Pseudomonas aeruginosa* quorum sensing revealed by DNA binding of purified LasR. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15833–15839 (2004).
83. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nat Biotechnol* **26**, 787–93 (2008).
84. Kaplan, H. B. & Greenberg, E. P. Diffusion of autoinducer is involved in regulation of the *Vibrio fischeri* luminescence system. *J. Bacteriol.* **163**, 1210–1214 (1985).
85. Seed, P. C., Passador, L. & Iglewski, B. H. Activation of the *Pseudomonas aeruginosa* lasI gene by LasR and the *Pseudomonas* autoinducer PAI: an autoinduction regulatory hierarchy. *J. Bacteriol.* **177**, 654–659 (1995).
86. Pestova, E. V., Håvarstein, L. S. & Morrison, D. A. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol. Microbiol.* **21**, 853–862 (1996).

87. de Kievit, T. R. & Iglewski, B. H. Bacterial quorum sensing in pathogenic relationships. *Infect. Immun.* **68**, 4839–4849 (2000).
88. Ji, G., Beavis, R. C. & Novick, R. P. Cell density control of staphylococcal virulence mediated by an octapeptide pheromone. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12055–12059 (1995).
89. Piper, K. R., Beck von Bodman, S. & Farrand, S. K. Conjugation factor of *Agrobacterium tumefaciens* regulates Ti plasmid transfer by autoinduction. *Nature* **362**, 448–450 (1993).
90. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–40 (2004).
91. Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435**, 228–232 (2005).
92. Kovarik, A., Matzke, M. A., Matzke, A. J. & Koulaková, B. Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. *Mol. Genet. Genomics MGG* **266**, 216–222 (2001).
93. Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 2340–2361 (1977).
94. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5**, 842–848 (2009).
95. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
96. Sousa, A., Bourgard, C., Wahl, L. M. & Gordo, I. Rates of transposition in *Escherichia coli*. *Biol. Lett.* **9**, 20130838 (2013).
97. Ohtsubo, Y., Genka, H., Komatsu, H., Nagata, Y. & Tsuda, M. High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC 17616. *Appl. Environ. Microbiol.* **71**, 1822–1828 (2005).
98. Ferry, M. S., Razinkov, I. A. & Hasty, J. Microfluidics for synthetic biology: from design to execution. *Methods Enzym.* **497**, 295–372 (2011).
99. Prindle, A. *et al.* A sensing array of radically coupled genetic ‘biopixels’. *Nature* **481**, 39–44 (2012).
100. Chen, A. Y. *et al.* Synthesis and patterning of tunable multiscale materials with engineered cells. *Nat. Mater.* **13**, 515–523 (2014).

101. Pérez, P. D., Weiss, J. T. & Hagen, S. J. Noise and crosstalk in two quorum-sensing inputs of *Vibrio fischeri*. *BMC Syst. Biol.* **5**, 153 (2011).
102. Winzer, K. *et al.* The *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL are controlled by quorum sensing and by RpoS. *J. Bacteriol.* **182**, 6401–6411 (2000).
103. Gray, K. M., Passador, L., Iglewski, B. H. & Greenberg, E. P. Interchangeability and specificity of components from the quorum-sensing regulatory systems of *Vibrio fischeri* and *Pseudomonas aeruginosa*. *J. Bacteriol.* **176**, 3076–3080 (1994).
104. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).
105. Ellis, T., Wang, X. & Collins, J. J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat. Biotechnol.* **27**, 465–471 (2009).
106. Hussain, F. *et al.* Engineered temperature compensation in a synthetic genetic clock. *Proc. Natl. Acad. Sci.* 201316298 (2014). doi:10.1073/pnas.1316298111
107. Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S. & Collins, J. J. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat. Methods* **9**, 1077–1080 (2012).
108. Nevozhay, D., Zal, T. & Balazsi, G. Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat. Commun.* **4**, 1451 (2013).
109. Zhou, J. X., Aliyu, M. D. S., Aurell, E. & Huang, S. Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface R. Soc.* **9**, 3539–3553 (2012).
110. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750-753 (2010).
111. Craig, N. L. *Mobile DNA II*. (ASM Press, 2002).
112. Sayut, D. J. & Sun, L. Slow activator degradation reduces the robustness of a coupled feedback loop oscillator. *Mol. Biosyst.* **6**, 1469–1474 (2010).
113. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol. Microbiol.* **85**, 21–38 (2012).
114. Lee, J. *et al.* Unraveling the regulatory connections between two controllers of breast cancer cell fate. *Nucleic Acids Res.* **42**, 6839–6849 (2014).
115. Waddington, C. H. *The strategy of the genes; a discussion of some aspects of*

theoretical biology. (Allen & Unwin, 1957).

116. Zhang, B. & Wolynes, P. G. Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10185–10190 (2014).
117. Li, C. & Wang, J. Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. *PLoS Comput Biol* **9**, e1003165 (2013).
118. Ferrell, J. E. Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol.* **22**, R458–R466 (2012).
119. Huang, S., Guo, Y.-P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).
120. Wang, J. Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Adv. Phys.* **64**, 1–137 (2015).
121. Xu, L., Zhang, K. & Wang, J. Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation. *PloS One* **9**, e105216 (2014).
122. Li, C. & Wang, J. Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *J. R. Soc. Interface* **10**, 20130787 (2013).
123. Schmiedel, J. M. *et al.* Gene expression. MicroRNA control of protein expression noise. *Science* **348**, 128–132 (2015).
124. Tanouchi, Y. *et al.* A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* **523**, 357–360 (2015).
125. Chalancon, G. *et al.* Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28**, 221–232 (2012).
126. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587–594 (2009).
127. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation. *Cell* **123**, 917–929 (2005).
128. Bessonard, S. *et al.* Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**, 3637–3648 (2014).
129. MacArthur, B. D., Please, C. P. & Oreffo, R. O. C. Stochasticity and the Molecular Mechanisms of Induced Pluripotency. *PLoS ONE* **3**, e3086 (2008).

130. Rabajante, J. F. & Babierra, A. L. Branching and oscillations in the epigenetic landscape of cell-fate determination. *Prog. Biophys. Mol. Biol.* **117**, 240–249 (2015).
131. Angeli, D., Ferrell, J. E. & Sontag, E. D. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1822–1827 (2004).
132. Lee, M. J. *et al.* Sequential Application of Anticancer Drugs Enhances Cell Death by Rewiring Apoptotic Signaling Networks. *Cell* **149**, 780–794 (2012).
133. Paşca, A. M. *et al.* Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671–678 (2015).
134. Pagliuca, F. W. *et al.* Generation of Functional Human Pancreatic β Cells In Vitro. *Cell* **159**, 428–439 (2014).
135. Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat. Biotechnol.* **26**, 443–452 (2008).
136. Liu, X. *et al.* Sequential introduction of reprogramming factors reveals a time-sensitive requirement for individual factors and a sequential EMT–MET mechanism for optimal reprogramming. *Nat. Cell Biol.* **15**, 829–838 (2013).
137. Wu, F., Su, R.-Q., Lai, Y.-C. & Wang, X. Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. *eLife* **6**, e23702 (2017).
138. Zhang, B. W., Jasnow, D. & Zuckerman, D. M. The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* **132**, (2010).
139. Palani, S. & Sarkar, C. A. Integrating Extrinsic and Intrinsic Cues into a Minimal Model of Lineage Commitment for Hematopoietic Progenitors. *PLOS Comput Biol* **5**, e1000518 (2009).
140. Narula, J., Smith, A. M., Gottgens, B. & Igoshin, O. A. Modeling Reveals Bistability and Low-Pass Filtering in the Network Module Determining Blood Stem Cell Fate. *PLOS Comput Biol* **6**, e1000771 (2010).
141. Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci.* **108**, 8257–8262 (2011).
142. Szathmáry, E., Jordán, F. & Pál, C. Can Genes Explain Biological Complexity? *Science* **292**, 1315–1316 (2001).

143. Kohles, S. S., Nève, N., Zimmerman, J. D. & Tretheway, D. C. Mechanical Stress Analysis of Microfluidic Environments Designed for Isolated Biological Cell Investigations. *J. Biomech. Eng.* **131**, 121006 (2009).
144. Shen, F., Li, X. & Li, P. C. H. Study of flow behaviors on single-cell manipulation and shear stress reduction in microfluidic chips using computational fluid dynamics simulations. *Biomicrofluidics* **8**, (2014).
145. Shemesh, J. *et al.* Flow-induced stress on adherent cells in microfluidic devices. *Lab. Chip* **15**, 4114–4127 (2015).
146. Smanski, M. J. *et al.* Synthetic biology to access and expand nature’s chemical diversity. *Nat. Rev. Microbiol.* **14**, 135–149 (2016).
147. Wright, G. Perspective: Synthetic biology revives antibiotics. *Nature* **509**, S13 (2014).
148. Cai, M. & Yang, Y. Targeted genome editing tools for disease modeling and gene therapy. *Curr. Gene Ther.* **14**, 2–9 (2014).
149. Jusiak, B., Cleto, S., Perez-Piñera, P. & Lu, T. K. Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends Biotechnol.* **34**, 535–547 (2016).
150. Rusk, N. Synthetic biology: CRISPR circuits. *Nat. Methods* **11**, 710–711 (2014).
151. Chakravarti, D. & Wong, W. W. Synthetic biology in cell-based cancer immunotherapy. *Trends Biotechnol.* **33**, 449–461 (2015).
152. Morsut, L. *et al.* Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **164**, 780–791 (2016).
153. Roybal, K. T. *et al.* Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits. *Cell* **164**, 770–779 (2016).
154. Pardee, K. *et al.* Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* **165**, 1255–1266 (2016).
155. Din, M. O. *et al.* Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**, 81–85 (2016).
156. Mus, F. *et al.* Symbiotic Nitrogen Fixation and Challenges to Extending it to Non-Legumes. *Appl. Environ. Microbiol.* AEM.01055-16 (2016). doi:10.1128/AEM.01055-16
157. Rogers, C. & Oldroyd, G. E. D. Synthetic biology approaches to engineering the nitrogen symbiosis in cereals. *J. Exp. Bot.* eru098 (2014). doi:10.1093/jxb/eru098

158. König, H., Frank, D., Heil, R. & Coenen, C. Synthetic Genomics and Synthetic Biology Applications Between Hopes and Concerns. *Curr. Genomics* **14**, 11–24 (2013).
159. Zhang, W. & Nielsen, D. R. Synthetic biology applications in industrial microbiology. *Front. Microbiol.* **5**, (2014).
160. Rocha, E. P. C. The Organization of the Bacterial Genome. *Annu. Rev. Genet.* **42**, 211–233 (2008).
161. Yang, L. *et al.* Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).
162. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32**, 1276–1281 (2014).
163. Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol. Syst. Biol.* **10**, 731 (2014).
164. Lee, J. W. *et al.* Creating Single-Copy Genetic Circuits. *Mol. Cell* **63**, 329–336 (2016).
165. Ma, K. C., Perli, S. D. & Lu, T. K. Foundations and Emerging Paradigms for Computing in Living Cells. *J. Mol. Biol.* **428**, 893–915 (2016).
166. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. *ACS Synth. Biol.* **3**, 363–371 (2014).
167. Lim, H. N., Lee, Y. & Hussein, R. Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci.* **108**, 10626–10631 (2011).
168. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
169. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science* **324**, 255–258 (2009).
170. Mao, Y., Liu, H., Liu, Y. & Tao, S. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **42**, 4813–4822 (2014).
171. Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci.* **107**, 3645–3650 (2010).
172. Kaikkonen, M. U., Lam, M. T. Y. & Glass, C. K. Non-coding RNAs as regulators of

- gene expression and epigenetics. *Cardiovasc. Res.* **90**, 430–440 (2011).
173. Mihailescu, R. Gene expression regulation: lessons from noncoding RNAs. *RNA* **21**, 695–696 (2015).
174. Sanjana, N. E. *et al.* High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545–1549 (2016).
175. Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453–1462 (1997).
176. Carpousis, A. J. The RNA Degradosome of Escherichia coli: An mRNA-Degrading Machine Assembled on RNase E. *Annu. Rev. Microbiol.* **61**, 71–87 (2007).
177. Chen, H., Shiroguchi, K., Ge, H. & Xie, X. S. Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Mol. Syst. Biol.* **11**, 781 (2015).
178. Hui, M. P., Foley, P. L. & Belasco, J. G. Messenger RNA degradation in bacterial cells. *Annu. Rev. Genet.* **48**, 537–559 (2014).
179. Salvador, M. L., Suay, L. & Klein, U. Messenger RNA degradation is initiated at the 5' end and follows sequence- and condition-dependent modes in chloroplasts. *Nucleic Acids Res.* **39**, 6213–6222 (2011).
180. Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M. & Rosenow, C. Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res.* **13**, 216–223 (2003).
181. Emory, S. A., Bouvet, P. & Belasco, J. G. A 5'-terminal stem-loop structure can stabilize mRNA in Escherichia coli. *Genes Dev.* **6**, 135–148 (1992).
182. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940 (1999).
183. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated Design of Synthetic Ribosome Binding Sites to Precisely Control Protein Expression. *Nat. Biotechnol.* **27**, 946–950 (2009).
184. Xia, T. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry (Mosc.)* **37**, 14719–14735 (1998).
185. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* **138**, 7016–7023 (2016).

186. Bennett, M. R. & Hasty, J. Overpowering the component problem. *Nat. Biotechnol.* **27**, 450–451 (2009).
187. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science* **352**, aac7341 (2016).
188. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42**, 2646–2659 (2014).
189. Campo, C. D., Bartholomäus, A., Fedyunin, I. & Ignatova, Z. Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLOS Genet.* **11**, e1005613 (2015).
190. Fluman, N., Navon, S., Bibi, E. & Pilpel, Y. mRNA-programmed translation pauses in the targeting of *E. coli* membrane proteins. *eLife* **3**, e03440 (2014).
191. Mackie, G. A. RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* **11**, 45–57 (2013).
192. Konu, O. & Li, M. D. Correlations between mRNA expression levels and GC contents of coding and untranslated regions of genes in rodents. *J. Mol. Evol.* **54**, 35–41 (2002).
193. Kudla, G., Lipinski, L., Caffin, F., Helwak, A. & Zylicz, M. High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells. *PLoS Biol.* **4**, (2006).
194. Sémon, M., Mouchiroud, D. & Duret, L. Relationship between gene expression and GC-content in mammals: statistical significance and biological relevance. *Hum. Mol. Genet.* **14**, 421–427 (2005).
195. Chen, Y., Kim, J. K., Hirning, A. J., Josić, K. & Bennett, M. R. Emergent genetic oscillations in a synthetic microbial consortium. *Science* **349**, 986–989 (2015).
196. Smit, M. H. de & Duin, J. van. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci.* **87**, 7668–7672 (1990).
197. Espah Borujeni, A. *et al.* Precise quantification of translation inhibition by mRNA structures that overlap with the ribosomal footprint in N-terminal coding sequences. *Nucleic Acids Res.* doi:10.1093/nar/gkx061
198. Faucon, P. C. *et al.* Gene Networks of Fully Connected Triads with Complete Auto-Activation Enable Multistability and Stepwise Stochastic Transitions. *PLoS ONE* **9**, (2014).

199. Mathur, M., Xiang, J. S. & Smolke, C. D. Mammalian synthetic biology for studying the cell. *J Cell Biol* jcb.201611002 (2016). doi:10.1083/jcb.201611002
200. Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat. Rev. Mol. Cell Biol.* **15**, 95–107 (2014).
201. Haynes, K. A. & Silver, P. A. Synthetic reversal of epigenetic silencing. *J. Biol. Chem.* **286**, 27176–27182 (2011).
202. Ye, H., Auel, D. & Fussenegger, M. Synthetic mammalian gene circuits for biomedical applications. *Curr. Opin. Chem. Biol.* **17**, 910–917 (2013).
203. Ye, H. & Fussenegger, M. Synthetic therapeutic gene circuits in mammalian cells. *FEBS Lett.* **588**, 2537–2544 (2014).
204. Roybal, K. T. *et al.* Engineering T Cells with Customized Therapeutic Response Programs Using Synthetic Notch Receptors. *Cell* **167**, 419–432.e16 (2016).
205. Yadav, V. G., De Mey, M., Giaw Lim, C., Kumaran Ajikumar, P. & Stephanopoulos, G. The future of metabolic engineering and synthetic biology: Towards a systematic practice. *Metab. Eng.* **14**, 233–241 (2012).