Multicellular circuit design in mammalian cells

Thesis by

Ronghui Zhu

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2023

(Defended June 14, 2022)

© 2023

Ronghui Zhu
ORCID: 0000-0001-8171-482X

ACKNOWLEDGEMENTS

Earlier this month I was at my friend's PhD commencement ceremony. A quote from Jamie Williamson, executive vice president at Scripps, resonated with me, "I have one more remark to the graduates: I have good news and I have bad news. The good news is that your period of toil and trial, hard work for low pay, is at an end. The bad news is, in a short time, you will look back on this time as the best time of your life." Though I wouldn't necessarily use the word 'best', hoping that better things are yet to come, I will look back on my PhD time with fond memories. This is not because of its achievements, but because of its richness—the darkest moment and the highlights, the frustration, the surprise, the uncertainty, the satisfaction. What makes this experience rich in another dimension is that it is not a lonely journey, but filled with joy, passion, and love from people around me. I would like to express my greatest gratitude to them.

I would like to first thank my PhD advisor, Michael Elowitz. I was first attracted to his lab because of several beautiful papers. However, over the years, Michael guided me not with a few questions he was overly obsessed with, but with his genuine curiosity and passion towards science. He shows his curiosity and passion through small moments—unexpected curious questions during people's introduction slides, contagious enthusiasm for a unique project direction that no one has tried before, etc. These moments gradually and profoundly shape my scientific personality. Michael also gave me unlimited support throughout my PhD, allowing me to freely pursue my scientific passion without worrying about anything. Quite refreshingly, every time I talk with Michael, I feel that he views me as his intellectual equal,

which gives me the courage to think differently and independently. For these I feel deeply grateful and fortunate.

I am also grateful and fortunate to be in a supportive, intellectually stimulating scientific environment that centered at Elowitz lab and extended to others outside of the lab. I want to thank Joe Markson, my first official mentor in the Elowitz lab. Joe helped me start my scientific journey in the lab and patiently taught me how to give a scientific talk. Xiaojing Gao and Mark Budde were my unofficial intellectual mentors in the Elowitz lab, who taught me practical ways of thinking about projects and PhD training. Mark also contributed some useful ideas towards the MultiFate project. Leah Santat, James Linton, and Jo Leonardo are three instrumental members of Elowitz lab whom I frequently find for day-to-day help. Leah also joined force with me on the shuttling project and greatly helped with much of the tedious troubleshooting work. Sheng Wang has been a great labmate and friend, and also once my housemate. I've been enjoying the numerous conversations about science and life with him. The MultiFate story presented in Chapter 2 was a fun collaborative effort with Jordi Garcia-Ojalvo, a long-time friend of Elowitz lab, and Jesus del Rio Salgado, my summer student. The BMP shuttling story presented in Chapter 3 is built on top of numerous pioneering efforts by former labmates Pulin Li, Sandy Nandagopal, and Joe Markson. For the two projects presented in this thesis, and at various points over the past few years, I also received generous help and suggestions from present and former labmates in Elowitz lab and people outside of Elowitz lab, including Heidi Klumpe, Xun Wang, Felix Horns, Yitong Ma, Lucy Chong, Rachael Kuintzle, Michael Flynn, Zibo Chen, Bo Gu, Christina Su, Dongyang Li, Jan Gregrowicz, Sanyam Satia, Jamie Tijerina, Evan Mun, Grace Edmonds, Fangyuan Ding,

Michaela Ince, Yapeng Su, Fangzhou Xiao, Xinying Ren, and Ahmad Khalil, among others. My rotation mentors, Felipe Vieceli at Marianne Bronner lab and Hao Yuan Kueh at Ellen Rothenberg lab, as well as Professor Bronner and Professor Rothenberg, further extended my scientific scope. I would also like to give a special shout-out to Justin Bois, who teaches the 'Introduction to Biological Circuit Design' class at Caltech together with Michael Elowitz. I learned much of my mathematical modeling skills from the class, and used the codes generously shared by Justin in the two projects presented in this thesis.

The members of my thesis committee, Pamela Bjorkman, Bruce Hay, Richard Murray, and Matt Thomson, have witnessed my PhD journey along the way, provided useful advice and encouraging messages during committee meetings over the years, and helped me keep every stage of my PhD training in check. I am grateful for this committee I chose.

This amazing scientific journey would not have been possible without the unconditional love and foundational support from my parents, Chunxia Huang and Changwu Zhu, my grandparents Yichuan Huang and Sanzhi Liu, my aunt Chunling Huang and my uncle Chunfeng Huang, for which I am forever grateful. Finally, Siwei Xie, my wife, has been the source of my inner peace. She has been the closest with me throughout this journey, accompanying me through ups and downs, through uncertain times like COVID. I am extremely fortunate to have her by my side, now for almost one third of my life.

ABSTRACT

Multicellular circuits control the development of multicellular organisms, through programming processes such as cell proliferation, cell differentiation, cell movement, and cell signaling. A fundamental goal of biology is to understand the design principles of these multicellular circuits, and use these principles to design synthetic multicellular systems for therapeutic purposes. Top-down approaches, for example analyzing embryos bearing genetic mutations, have identified key genes in many multicellular circuits, but are challenging to study these circuits in an isolated context and in a quantitative and systematic manner. An alternative, complementary approach is to engineer or reconstitute multicellular circuits from bottom-up, which allows us to overcome the limitations of top-down approach and gain quantitative insights into multicellular circuit design. In this thesis, we use this bottom-up approach to explore the design principles of two multicellular circuits. In the first project, we took inspiration from two prevalent features from natural multistable circuits, namely competitive protein-protein interactions and positive autoregulation, to design a synthetic multistable circuit architecture called MultiFate. Both in the model and in the experiment, MultiFate circuits generate multiple cellular states, each stable for weeks, allow control over state-switching and state stability, and can be easily expanded to generate more states. In the second project, we use a gradient reconstitution system to systematically analyze a gradient modulation circuit consisting of BMP4 and its modulators, Chordin, Twsg and BMP-1. We found that the circuit can give rise to diverse gradient modulation capabilities. In particular, the full circuit is sufficient for active ligand shuttling and generation of non-monotonic displaced gradient. These multicellular circuits could provide a foundation for engineering synthetic multicellular systems in mammalian cells.

PUBLISHED CONTENT AND CONTRIBUTIONS

R. Zhu, J.M. del Rio-Salgado, J. Garcia-Ojalvo, M.B. Elowitz Synthetic multistability in mammalian cells. *Science* 375.6578 (2022). doi: 10.1126/science.abg9765.

R. Zhu participated in the conception of the project, the design and implementation of experiments, the analysis of data, the construction of mathematical modeling, and the writing of the manuscript.

TABLE OF CONTENTS

LIST OF FIGURES, BOXES AND TABLES

Chapter 1. Introduction

1.1 Multicellular circuits control the development of a single fertilized egg into a multicellular organism.

One of the most fascinating phenomena in biology is the embryonic development of multicellular organisms. It is a highly complex, yet very reproducible process that involves many multicellular programs, such as cell proliferation, cell differentiation, cell movement, and cell signaling (*1*). These multicellular programs need to be accurately controlled to ensure a normal embryonic development. For example, to control for a proper organ size, the cell growth, division, and apoptosis are linked to each other and tightly regulated (*2*). To generate a mosaic tissue pattern, the cell fate decisions of neighboring cells are closely coordinated to give rise to mutually exclusive fates (*3*). To robustly segregate and sort cells of different lineages into distinct regions in mammalian embryos, the molecular and mechanical properties of cell surfaces need to be precisely programmed (*4*).

The precise control of these multicellular programs is realized by multicellular circuits. In these multicellular circuits, molecules encoded by a variety of different genes, interact within cells and among cells to control the tissue morphogenetic process in a spatially and temporally precise manner. For example, a multicellular circuit with Bicoid and its downstream gap genes plays a key role in the first hours of drosophila embryo development. In this circuit, Bicoid proteins translated from maternally deposited mRNAs form a concentration gradient on the anterior-posterior axis of embryo. This gradient initiates the expression pattern of its downstream gap genes. A complex

interaction network among gap genes then further develops and refines the pattern, robustly and accurately segmenting the early drosophila embryos (*5–9*). A fundamental goal of biology is to understand how these multicellular circuits are designed, and use this understanding to build synthetic multicellular systems for therapeutic purposes.

1.2 Engineering and reconstituting multicellular circuits enable one to understand the design principles of these circuits from bottom up.

Much of our knowledge of these multicellular circuits comes from analyzing cells or embryos bearing genetic mutations or chemically perturbed. Through these top-down approaches, many key genes have been identified for multicellular programs that control cell proliferation (*10–12*), cell death (*13–16*), cell differentiation (*17–21*), cell movement and adhesion (*22*, *23*), and cell signaling (*24–27*). In a more ambitious effort, how these key developmental genes interact with each other are comprehensively mapped in model organisms like sea urchin (*28*). Recent technological advancement further reveals more relevant genes through profiling developmental systems at single cell resolution in larger scales (single-cell omics) (*29*), with preserved spatial information (*30*) and at various developmental stages (*31*).

Despite continued methodological and technological advancements, we are still far away from fully understanding the design principles of these multicellular circuits, and using this understanding to engineer synthetic multicellular systems for therapeutic purposes. Descriptive measurements such as omics experiments are usually not enough for a mechanistic understanding of multicellular circuits (*32*), and top-down approaches, such

as using genetic perturbations to dissect multicellular circuits, face several challenges. These challenges are:

1. Natural circuits are interconnected, and therefore one cannot isolate one individual circuit to study from the rest of the gene network. Perturbing one gene in a circuit usually has propagating effects on multiple connecting circuits, and distinguishing primary effects from secondary ones can be nontrivial. Frequent compensatory effects from paralog genes further complicate the analysis (*33*). Multicellular circuits are also highly interconnected, and this interconnectedness often spans multiple cells, allowing for the coordination of cellular programs across these cells. For example, during drosophila wing disc development, Dpp morphogen signaling both induce cell differentiation and pattern the tissue, and control cell proliferation. Thus tissue patterning is coupled with cell proliferation, which has been suggested to enable the scaling of Dpp morphogen concentrations with growing tissue size (*34*, *35*). This kind of interconnectedness makes it hard to study individual multicellular circuits and gain insights into their design principles.

2. It remains hard to gain a quantitative understanding of natural circuits using a top-down approach. Quantitative understanding requires one to perform controlled perturbation and quantitative measurements. Measurement results can then be used to construct a mathematical model. Our understanding of a circuit increases as its mathematical model becomes increasingly predictive. For example, to understand how mouse embryonic stem cells achieve a robust response to differentiation signals, Sokolik et al., engineered mouse embryonic

stem cells to allow for quantitative and temporal control of Brn2 expression using optogenetics (*36*). Using this system, the authors quantitatively measured the differentiation outcomes with different durations of Brn2 induction, and used these measurements to construct a mathematical model incorporating the interactions of Brn2 with pluripotent factors. The model predicted that the pluripotent gene regulatory network can filter out the transient fluctuations of Brn2 expression, and this filtering capability is controlled by the intrinsic half-life of Nanog, which was later confirmed by additional experiments. However, this kind of quantitative study is generally hard to perform. One challenge comes from the interconnectedness of natural circuits. Because of interconnectedness of circuits, a mathematical model may need to encompass other related circuits and can thus become overwhelmingly complicated and intractable. Furthermore, controlled and quantitative perturbations are easier when only one cell type is involved (*36*–*39*), but become much more challenging with multicellular circuits. Since multicellular circuits often involve multiple distinct cell types, the same component may operate in a different regime in each cell type. Thus cell-type specific perturbation may be required, but tools to target specific cell types (e.g. cell-type specific promoters) are not always available.

3. With top-down approaches, observable behaviors of a circuit are limited to those possible in specific contexts, but one circuit design may perform differently in different developmental stages, in different tissues, in different model organisms, or in different environments. One may not be able to find a general design principle of a circuit if only studying it in one context using top-down approaches.

For example, a circuit with BMP, Chordin, Twsg, and BMP-1 has been suggested to enable ligand shuttling and reshape BMP gradients in the embryos of Drosophila (*40*) and Xenopus (*41*). However, two recent papers showed that ligand shuttling is not playing a role in zebrafish early embryonic development, despite the presence of all circuit components (*42*, *43*). Nevertheless, these studies do not rule out the possibility that it is the same circuit performing in very different parameter regimes in different model organisms.

4. Top-down approaches do not allow one to compare alternative circuit designs. Testing whether alternative circuit designs are permitted, and the tradeoffs among designs can provide valuable insights into the evolutionary origins of biological circuits (*44*). To test an alternative circuit design using a top-down approach, one needs to massively rewire the natural circuit, for example by introducing mutations to change regulatory interactions among genes (*45*). This is a daunting task even with CRISPR-Cas9 and other new genome engineering methods (*46–48*), since mutations sufficient to change regulatory interactions are not always available or known. Rewiring multicellular circuits can be even more challenging since one may need to rewire circuit components across multiple cell types.

Alternatively, one could engineer or reconstitute multicellular circuits by putting together components one by one and gradually increasing the complexity of the system. This bottom-up approach offers a complementary way to understand circuit design and can tackle the above challenges,

1. Bottom-up approaches allow one to study a particular multicellular circuit in a relatively isolated context. To achieve this, one can either reconstitute or build the circuit in selected cell types that express relevant circuit components and minimally express components from other interconnected circuits. For example, Li et al. reconstitute a Shh gradient formation circuit in NIH3T3 cells, which do not naturally express Shh ligands and can transduce Shh signals without differentiating. Using this system, the authors found that a double-negative regulatory logic and negative feedback confer speed and robustness to the gradient formation process (*49*). Alternatively, circuit components can be modified to break their interactions with other circuits, while still preserving the overall circuit architecture. For example, to uncover the minimal requirements for morphogen (long-range signaling molecules) gradient formation in drosophila wings, Stapornwongkul et al. replaced the morphogen Dpp, which has extensive interactions with many extracellular modulators, with an inert green fluorescent protein (GFP). The authors then neatly convert GFP into a morphogen by substituting the extracellular domains of Dpp receptors with anti-GFP nanobodies, thus allowing cell signaling through specific interactions between GFP and anti-GFP nanobodies. Using this system, they found that GFP can replace Dpp in patterning the tissue, and this patterning can be further improved with glycosylphosphatidylinositol-anchored nonsignaling receptor expression. (*50*)

2. Bottom-up approaches are suitable for performing quantitative analysis of a circuit. During bottom-up reconstitution of a circuit, it is very easy to add control handles through synthetic tools such as inducible promoters reviewed in (*51*) or

tunable degron (*52*). When reconstituting a multicellular circuit, one can first engineer multiple cell populations to express different circuit components or express the same components in different ways, and then mix these populations to reconstruct the circuit. For example, in the Shh gradient reconstitution system mentioned above, the authors were able to control ligand production rate through sender cell fractions, and control the strength of receptor negative feedback through a chemical inducer (*49*). To quantitatively analyze a macrophage-fibroblast two-cell circuit, Zhou et al. mixed two cell types at different ratios and quantified their numbers at different time points. This allowed them to map the phase diagram of this two-cell system and find that this system can evolve into two multiple states (*53*). Similar bottom-up approaches have been successful in gaining quantitative insights into other multicellular circuits (*54*, *55*).

3. Bottom-up approaches enable a more comprehensive exploration of the behaviors of a multicellular circuit, since the parameter regimes of a circuit are not constrained by certain natural contexts. For example, to understand how the BMP signaling pathway processes multi-ligand input, Antebi et al. added large combinations of recombinant BMP ligands on cells with BMP signaling fluorescent reporters. Through measurements of fluorescence level in response to different BMP ligand combinations, Antebi et al. found that a promiscuous ligand-receptor interaction architecture enables diverse signal-processing capabilities (*56*). This comprehensive picture of signal-processing capabilities by BMP signaling pathway can be hard to reveal using a top-down approach, since

usually only a few ligand and receptor combinations are present in a certain developmental context.

4. Additionally, we can compare a multicellular circuit design with alternative ones using bottom-up approaches. The engineering flexibility provided by the bottom-up reconstitution process allows one to easily construct alternatively wired multicellular circuits, and make side-by-side comparison of circuit behaviors. For example, Ma et al. used a synthetic cell-cell communication system to construct a paradoxical population control circuit, and compared this circuit with a simple negative feedback one (*57*). The simple negative feedback circuit is more susceptible to mutational escape, thus confirming the role of paradoxical circuit architecture in maintaining population homeostasis (*58*). This and other similar studies (*49*, *59*, *60*) help us understand the features and tradeoffs among different circuit designs, and allow us to identify the minimal sufficient circuit required for certain multicellular behaviors. These minimal circuits then can be used for building synthetic multicellular systems for therapeutic purposes.

This thesis shows how we use bottom-up approaches to explore the design principles of two multicellular circuits. The first circuit, called MultiFate, is a naturally-inspired synthetic multistable circuit that supports long-term, controllable, and expandable multistability in mammalian cells. In the second circuit, we utilize ligands and modulators found in the natural BMP signaling pathway to reconstitute gradient formation in vitro. This reconstitution system allows us to quantitatively study how BMP gradients are modulated, and shows the minimal circuit sufficient for a certain modulation such as gradient displacement.

1.3 Multistable circuits allow genetically identical cells to exist in multiple molecularly and functionally distinct cellular states.

In multicellular organisms, cells sharing the same genome can exhibit morphologically and functionally distinct phenotypes. This is realized by gene expression differences (*1*). For example, muscle cells around the body express myosins to enable muscle contraction (*61*), while beta cells in pancreatic islets express insulin to control the level of blood glucose (*62*). A central topic in biology is to understand how these gene expression differences are established and maintained during development.

Advances in technologies such as microarray (*63*) and RNA sequencing (*64*) allow us to measure gene expression differences among cell types at transcriptomic scales, and these measurements, especially those perform at single cells level (*65*), have led to three interesting observations of a cell's gene expression state: First, although genes in a multicellular organism (e.g. ~20000 genes for human cells) can give rise to a huge number of possible gene expression states, we only observe a limited number of stable gene expression states in a multicellular organism, corresponding to different cell types (*66*, *67*). Secondly, these cell type specific gene expression states are well-separated from each other, and any intermediate states, such as those observed during cell state transitions, are usually very transitory and unstable (*68*, *69*). Finally, these cell type specific gene expression states seem to be attractors, as people observe differentiating cells taking different routes on the gene expression space converge to a similar place (*70*).

Biologists have identified many key genes that control gene expression differences in different cell types (*20*, *21*, *71*, *72*). These genes, such as transcription factors, usually

regulate the expression of many downstream genes to activate or repress cell type specific programs. They are also highly interconnected, regulating the expression of each other to form gene regulatory networks (*28*). How can these gene regulatory networks explain the above observations? In dynamical systems theory, a gene regulatory network can behave as a multistable system (*73*), and properties of a multistable system match with the above observations: while the gene expression space can be high-dimensional and huge, a multistable system can define a limited number of stable steady states on this space, which could explain the limited number of cell types observed in our body. If a system is not at one of its stable steady states, it will approach one of the stable steady states until reaching that state. This could explain the attractiveness of cell type specific gene expression states and transitory nature of any intermediate states during development. This idea is most famously conceptualized by C. H. Waddington in 1957 (*74*). He proposed that gene regulatory networks shape an 'epigenetic landscape' with stable gene expression states represented as valleys on the landscape. Developmental processes are depicted as cells rolling down the hill to reach the bottom of valleys.

Several experimental evidence supports the theory that a gene regulatory network can behave as a multistable system. One of the earliest work is in the regulatory system of phage lambda, which shows that a mutual repression circuit between lambda repressor and Cro enables lambda to toggle between lytic and lysogenic states (*75*). Later, in multicellular organisms, work in Xenopus oocytes showed that a positive feedback circuit allows irreversible maturation in response to transient inductive stimulus, suggesting that this circuit can exist in two different states (basal and maturation) (*37*). These two

systems and others (*38*, *76–78*) showed that a genetic circuit can establish multiple gene expression states.

What are the design principles for these gene regulatory networks to exhibit multistability? To answer this question, synthetic biologists build synthetic multistable circuits, and use this bottom-up approach to explore different circuit architectures. The first synthetic bistable circuit was developed by Gardner et al. in bacteria, where two repressors mutually repress each other's expression, similar to the bistable switch circuit in phage lambda (*79*). Later, multistable circuits have been built in other organisms (*80–83*), using other molecular implementations (*82*, *84*) or using alternative circuit architectures (*81–83*, *85*). Most of these efforts have been limited to two-state systems. One exception is a four-state system in bacteria by Wu et al. in 2017, realized by a circuit with a combination of positive autoregulation and mutual inhibition (*85*). However, the use of transcriptional mutual repression still limits the circuit's expandability, since with N repressors, each repressor needs to be repressed by N-1 other respressors, which becomes impractical given limited programmable promoter space. For many of these circuits, especially those built in mammalian cells, certain gene expression states can gradually lose their stability, presumably due to limited robustness against biological noise. Thus, we have lacked an expandable circuit architecture for realizing robust multistability in mammalian cells.

Natural mammalian multistable circuits provide inspiration for such a synthetic architecture. In many natural fate control systems, transcription factors positively autoregulate their own expression, and competitively interact with one another to form a variety of homodimers, heterodimers, and higher order multimeric forms (18–24). For

example, during myogenesis, muscle regulatory factors (MRF) such as MyoD heterodimerize with E proteins to activate their own expression and the broader myogenesis program, while Id family proteins disrupt this process through competitive dimerization (23, 24). Similarly, during embryogenesis, Sox2 and Sox17 competitively interact with Oct4 to control fate decisions between pluripotency and endodermal differentiation (21, 22). Related combinations of positive autoregulation and cross-inhibition could extend multistability behaviors beyond bistability and generate bifurcation dynamics that explain the partial irreversibility of cell differentiation (9, 12). Nevertheless, it remains unclear whether these natural architectures could be adapted to enable synthetic multistability. In Chapter 2 of this thesis, we show how a synthetic multistable system based on principles derived from natural cell fate control systems can generate robust, controllable, expandable multistability in mammalian cells.

1.4 BMPs form concentration gradients to pattern the developing tissues, and their gradients are extensively modulated by extracellular factors.

To develop into reproducible forms, embryos not only need multistable circuits to establish multiple cell fates, but also circuits for cell-cell communications to instruct cell fate decisions in the developing tissues. One class of signaling molecules is called morphogens, an idea first coined by Alan Turning (*86*) and its molecular basis first discovered in drosophila embryos (*5*, *24*, *25*). They are secreted by source cells in a tissue, and can form a concentration gradient and act in long range (*87*).

Bone morphogenetic proteins (BMPs) belong to one family of morphogens. They are first discovered in bone morphogenesis (thus the name) (*88*), but later have been shown to

play important roles in early embryonic development as well as the development of multiple organs such as eyes and kidney (*89*). Similar to other morphogens, BMPs form concentration gradients to pattern the developing tissue (*90*).

One interesting feature of BMPs is their extensive interactions with diverse extracellular modulators (*91–93*). These modulators, like BMPs, are also secreted by the cells, and can interact with BMP ligands in a variety of different ways. These interactions can change the shape of BMP concentration gradients, thus controlling tissue patterning (*92*).

One interaction that is particularly interesting is interactions among BMP, Chordin, Twisted gastrulation (Twsg), and BMP-1. Chordin can bind to BMP and prevent it from binding to receptors for signaling (*94*), and Twsg can form ternary complexes with BMP-Chordin to further strengthen this interaction (*95–98*). Thus, Chordin is mostly known as a BMP antagonist. However, this inhibition may serve another purpose: it prevents BMP from receptor mediated internalization and degradation (*99*). Thus it is postulated that Chordin can "shuttle" BMP ligands on the cell surface and increase their diffusibility. Furthermore, BMP-1 (homolog of Tolloid in Drosophila) protease can cleave Chordin, which releases BMP from the ternary complex for signaling (*100*). Ligand shuttling enabled by a circuit with BMP, Chordin, Tsg, and BMP-1 has been suggested to produce non-intuitive gradient shapes during the embryonic development of Drosophila (*40*, *101*), Xenopus (*41*), and mouse (*102*). However, two papers refute the role of BMP ligand shuttling by Chordin in zebrafish embryos, despite the presence of all circuit components (*42*, *43*).

This controversy may come from the limitation of top-down approaches: with the same circuit, ligand shuttling and certain gradient shapes may happen in one parameter regime, but not the other. One developmental context may only allow us to explore a limited set of parameter regimes, within which some circuit behaviors may not appear. Furthermore, top-down approaches can only reveal necessary components and interactions for ligand shuttling and certain gradient shapes, but cannot identify the minimal sufficient circuit. In Chapter 3 of this thesis, we are trying to use a bottom-up reconstitution method to quantitatively understand how mammalian BMP gradients can be shaped by various extracellular modulators, and identify the minimal circuit for certain gradient modulation such as gradient displacement.

1.5 References

1. P. Weiss, THE SO-CALLED ORGANIZER AND THE PROBLEM OF ORGANIZATION IN AMPHIBIAN DEVELOPMENT. *Physiological Reviews*. **15**, 639–674 (1935).

2. K. Tumaneng, R. C. Russell, K.-L. Guan, Organ size control by Hippo and TOR pathways. *Curr. Biol.* **22**, R368–79 (2012).

3. N. Perrimon, C. Pitsouli, B.-Z. Shilo, Signaling Mechanisms Controlling Cell Fate and Embryonic Patterning. *Cold Spring Harbor Perspectives in Biology*. **4**, a005975–a005975 (2012).

4. A. Yanagida, E. Corujo-Simon, C. K. Revell, P. Sahu, G. G. Stirparo, I. M. Aspalter, A. K. Winkel, R. Peters, H. De Belly, D. A. D. Cassani, S. Achouri, R. Blumenfeld, K. Franze, E. Hannezo, E. K. Paluch, J. Nichols, K. J. Chalut, Cell surface fluctuations regulate early embryonic lineage sorting. *Cell*. **185**, 1258 (2022).

5. W. Driever, C. Nüsslein-Volhard, The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*. **54**, 95–104 (1988).

6. C. Wang, R. Lehmann, Nanos is the localized posterior determinant in Drosophila. *Cell*. **66**, 637–647 (1991).

7. D. Tautz, R. Lehmann, H. Schnürch, R. Schuh, E. Seifert, A. Kienlin, K. Jones, H. Jäckle, Finger protein of novel structure encoded by hunchback, a second member of the gap class of Drosophila segmentation genes. *Nature*. **327**, 383–389 (1987).

8. C. Nüsslein-Volhard, E. Wieschaus, Mutations affecting segment number and polarity in

Drosophila. *Nature*. **287**, 795–801 (1980).

9. J. Jaeger, The gap gene network. *Cellular and Molecular Life Sciences*. **68**, 243–274 (2011).

10. L. H. Hartwell, J. Culotti, J. R. Pringle, B. J. Reid, Genetic Control of the Cell Division Cycle in Yeast. *Science*. **183**, 46–51 (1974).

11. K. Vermeulen, D. R. Van Bockstaele, Z. N. Berneman, The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Proliferation*. **36**, 131–149 (2003).

12. C. Norbury, P. Nurse, ANIMAL CELL CYCLES AND THEIR CONTROL. *Annual Review of Biochemistry*. **61**, 441–468 (1992).

13. S. Elmore, Apoptosis: A Review of Programmed Cell Death. *Toxicologic Pathology*. **35**, 495–516 (2007).

14. Y. Fuchs, H. Steller, Programmed Cell Death in Animal Development and Disease. *Cell*. **147**, 1640 (2011).

15. M. D. Jacobson, M. Weil, M. C. Raff, Programmed Cell Death in Animal Development. *Cell*. **88**, 347–354 (1997).

16. H. Ellis, H. R. Horvitz, Genetic control of programmed cell death in the nematode C. elegans. *Cell*. **44**, 817–829 (1986).

17. L. Li, M. Leid, E. V. Rothenberg, An Early T Cell Lineage Commitment Checkpoint Dependent on the Transcription Factor *Bcl11b*. *Science*. **329**, 89–93 (2010).

18. E. W. Scott, M. C. Simon, J. Anastasi, H. Singh, Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science*. **265**, 1573–1577 (1994).

19. L. Pevny, M. C. Simon, E. Robertson, W. H. Klein, S. F. Tsai, V. D'Agati, S. H. Orkin, F. Costantini, Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*. **349**, 257–260 (1991).

20. S. H. Orkin, Transcription Factors and Hematopoietic Development. *Journal of Biological Chemistry*. **270**, 4955–4958 (1995).

21. M. A. Rudnicki, R. Jaenisch, The MyoD family of transcription factors and skeletal myogenesis. *BioEssays*. **17**, 203–209 (1995).

22. A. Barrallo-Gimeno, M. A. Nieto, The Snail genes as inducers of cell movement and survival: implications in development and cancer. *Development*. **132**, 3151–3161 (2005).

23. K. Vleminckx, R. Kemler, Cadherins and tissue formation: integrating adhesion and signaling. *Bioessays*. **21**, 211–220 (1999).

24. W. Driever, C. Nüsslein-Volhard, A gradient of bicoid protein in Drosophila embryos. *Cell*. **54**, 83–93 (1988).

25. G. Struhl, K. Struhl, P. M. Macdonald, The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*. **57**, 1259–1273 (1989).

26. M. Katoh, M. Katoh, WNT Signaling Pathway and Stem Cell Signaling Network. *Clinical Cancer Research*. **13**, 4042–4045 (2007).

27. J. Gerhart, 1998 warkany lecture: Signaling pathways in development. *Teratology*. **60**, 226–239 (1999).

28. M. Levine, E. H. Davidson, Gene Regulatory Networks for Development. *PNAS*, 125–185 (2005).

29. Y. Hu, Q. An, K. Sheu, B. Trejo, S. Fan, Y. Guo, Single Cell Multi-Omics Technology: Methodology and Application. *Front Cell Dev Biol*. **6**, 28 (2018).

30. S. K. Longo, M. G. Guo, A. L. Ji, P. A. Khavari, Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* **22**, 627–644 (2021).

31. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. **566**, 496–502 (2019).

32. Y. Lazebnik, Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer Cell*. **2**, 179–182 (2002).

33. G. Diss, D. Ascencio, A. DeLuna, C. R. Landry, Molecular mechanisms of paralogous compensation and the robustness of cellular networks. *J. Exp. Zool. B Mol. Dev. Evol.* **322**, 488–499 (2014).

34. I. Averbukh, D. Ben-Zvi, S. Mishra, N. Barkai, Scaling morphogen gradients during tissue growth by a cell division rule. *Development*. **141**, 2150–2156 (2014).

35. O. Wartlick, P. Mumcu, A. Kicheva, T. Bittig, C. Seum, F. Jülicher, M. González-Gaitán, Dynamics of Dpp signaling and proliferation control. *Science*. **331**, 1154–1159 (2011).

36. C. Sokolik, Y. Liu, D. Bauer, J. McPherson, M. Broeker, G. Heimberg, L. S. Qi, D. A. Sivak, M. Thomson, Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst*. **1**, 117–129 (2015).

37. W. Xiong, J. E. Ferrell, A positive-feedback-based bistable "memory module" that governs a cell fate decision. *Nature*. **426**, 460–465 (2003).

38. A. Gunne-Braden, A. Sullivan, B. Gharibi, R. S. M. Sheriff, A. Maity, Y.-F. Wang, A. Edwards, M. Jiang, M. Howell, R. Goldstone, R. Wollman, P. East, S. D. M. Santos, GATA3 Mediates a Fast, Irreversible Commitment to BMP4-Driven Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell*. **26**, 693–706.e9 (2020).

39. Y. Hart, S. Reich-Zeliger, Y. E. Antebi, I. Zaretsky, A. E. Mayo, U. Alon, N. Friedman, Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels. *Cell*. **158**, 1022–1032 (2014).

40. A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, N. Barkai, Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature*. **419**, 304–308 (2002).

41. D. Ben-Zvi, B.-Z. Shilo, A. Fainsod, N. Barkai, Scaling of the BMP activation gradient in

Xenopus embryos. *Nature*. **453**, 1205–1211 (2008).

42. A. P. Pomreinke, G. H. Soh, K. W. Rogers, J. K. Bergmann, A. J. Bläßle, P. Müller, Dynamics of BMP signaling and distribution during zebrafish dorsal-ventral patterning. *Elife*. **6** (2017), doi:10.7554/eLife.25861.

43. J. Zinski, Y. Bu, X. Wang, W. Dou, D. Umulis, M. C. Mullins, Systems biology derived source-sink mechanism of BMP gradient formation. *Elife*. **6** (2017), doi:10.7554/eLife.22199.

44. B. B. Tuch, H. Li, A. D. Johnson, Evolution of eukaryotic transcription circuits. *Science*. **319**, 1797–1799 (2008).

45. D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, S. Mundlos, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. **161**, 1012–1025 (2015).

46. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science*. **361**, 866–869 (2018).

47. D. A. Wright, T. Li, B. Yang, M. H. Spalding, TALEN-mediated genome editing: prospects and perspectives. *Biochem. J.* **462**, 15–24 (2014).

48. F. D. Urnov, E. J. Rebar, M. C. Holmes, H. Steve Zhang, P. D. Gregory, Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*. **11**, 636–646 (2010).

49. P. Li, J. S. Markson, S. Wang, S. Chen, V. Vachharajani, M. B. Elowitz, Morphogen gradient reconstitution reveals Hedgehog pathway design principles. *Science*. **360**, 543–548 (2018).

50. K. S. Stapornwongkul, M. de Gennes, L. Cocconi, G. Salbreux, J.-P. Vincent, Patterning and growth control in vivo by an engineered GFP gradient. *Science*. **370**, 321–327 (2020).

51. T. Kallunki, M. Barisic, M. Jäättelä, B. Liu, How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells*. **8** (2019), doi:10.3390/cells8080796.

52. M. Iwamoto, T. Björklund, C. Lundberg, D. Kirik, T. J. Wandless, A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17**, 981–988 (2010).

53. X. Zhou, R. A. Franklin, M. Adler, J. B. Jacox, W. Bailis, J. A. Shyer, R. A. Flavell, A. Mayo, U. Alon, R. Medzhitov, Circuit Design Features of a Stable Two-Cell System. *Cell*. **172**, 744–757.e17 (2018).

54. M. Matsuda, M. Koga, K. Woltjen, E. Nishida, M. Ebisuya, Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat. Commun.* **6**, 6195 (2015).

55. R. Sekine, T. Shibata, M. Ebisuya, Synthetic mammalian pattern formation driven by differential diffusivity of Nodal and Lefty. *Nat. Commun.* **9**, 5456 (2018).

56. Y. E. Antebi, J. M. Linton, H. Klumpe, B. Bintu, M. Gong, C. Su, R. McCardell, M. B. Elowitz, Combinatorial Signal Perception in the BMP Pathway. *Cell*. **170**, 1184–1196.e24

(2017).

57. Y. Ma, M. W. Budde, M. N. Mayalu, J. Zhu, A. C. Lu, R. M. Murray, M. B. Elowitz, Synthetic mammalian signaling circuits for robust cell population control. *Cell*. **185**, 967–979.e12 (2022).

58. O. Karin, U. Alon, Biphasic response as a mechanism against mutant takeover in tissue homeostasis circuits. *Mol. Syst. Biol.* **13**, 933 (2017).

59. S. Toda, L. R. Blauch, S. K. Y. Tang, L. Morsut, W. A. Lim, Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science*. **361**, 156–162 (2018).

60. S. Toda, W. L. McKeithan, T. J. Hakkinen, P. Lopez, O. D. Klein, W. A. Lim, Engineering synthetic morphogen systems that can program multicellular patterning. *Science*. **370**, 327–331 (2020).

61. L. Silberstein, S. G. Webster, M. Travis, H. M. Blau, Developmental progression of myosin gene expression in cultured muscle cells. *Cell*. **46**, 1075–1081 (1986).

62. L. Orci, K. H. Gabbay, W. J. Malaisse, Pancreatic beta-cell web: its possible role in insulin secretion. *Science*. **175**, 1128–1130 (1972).

63. D. D. L. Bowtell, Options available — from start to finish — for obtaining expression data by microarray. *Nature Genetics*. **21**, 25–32 (1999).

64. F. Ozsolak, P. M. Milos, RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*. **12**, 87–98 (2011).

65. A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, S. A. Teichmann, The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*. **58**, 610–620 (2015).

66. X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S. H. Orkin, G.-C. Yuan, M. Chen, G. Guo, Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*. **173**, 1307 (2018).

67. J. Cao, D. R. O'Day, H. A. Pliner, P. D. Kingsley, M. Deng, R. M. Daza, M. A. Zager, K. A. Aldinger, R. Blecher-Gonen, F. Zhang, M. Spielmann, J. Palis, D. Doherty, F. J. Steemers, I. A. Glass, C. Trapnell, J. Shendure, A human cell atlas of fetal gene expression. *Science*. **370** (2020), doi:10.1126/science.aba7721.

68. R. Soldatov, M. Kaucka, M. E. Kastriti, J. Petersen, T. Chontorotzea, L. Englmaier, N. Akkuratova, Y. Yang, M. Häring, V. Dyachuk, C. Bock, M. Farlik, M. L. Piacentino, F. Boismoreau, M. M. Hilscher, C. Yokota, X. Qian, M. Nilsson, M. E. Bronner, L. Croci, W.-Y. Hsiao, D. A. Guertin, J.-F. Brunet, G. G. Consalez, P. Ernfors, K. Fried, P. V. Kharchenko, I. Adameyko, Spatiotemporal structure of cell fate decisions in murine neural crest. *Science*. **364** (2019), doi:10.1126/science.aas9536.

69. A. Furlan, V. Dyachuk, M. E. Kastriti, L. Calvo-Enrique, H. Abdo, S. Hadjab, T. Chontorotzea, N. Akkuratova, D. Usoskin, D. Kamenev, J. Petersen, K. Sunadome, F. Memic, U. Marklund, K. Fried, P. Topilko, F. Lallemend, P. V. Kharchenko, P. Ernfors, I.

Adameyko, Multipotent peripheral glial cells generate neuroendocrine cells of the adrenal medulla. *Science*. **357** (2017), doi:10.1126/science.aal3753.

70. S. Huang, G. Eichler, Y. Bar-Yam, D. E. Ingber, Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701 (2005).

71. S. G. Almalki, D. K. Agrawal, Key transcription factors in the differentiation of mesenchymal stem cells. *Differentiation*. **92**, 41–51 (2016).

72. G. J. Pan, Z. Y. Chang, H. R. Schöler, P. E. I. Duanqing, Stem cell pluripotency and transcription factor Oct4. *Cell Research*. **12**, 321–329 (2002).

73. S. Huang, Multistability and Multicellularity: Cell Fates as High-Dimensional Attractors of Gene Regulatory Networks. *Computational Systems Biology*, 293–326 (2006).

74. C. H. Waddington, The Strategy of Genes. *American Journal of Botany*. **66**, 313–320 (1957).

75. M. Ptashne, *A Genetic Switch: Gene Control and Phage [lambda]* (Wiley-Blackwell, 1986).

76. H. Y. Kueh, A. Champhekar, S. L. Nutt, M. B. Elowitz, E. V. Rothenberg, Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science*. **341**, 670–673 (2013).

77. G. Yao, T. J. Lee, S. Mori, J. R. Nevins, L. You, A bistable Rb-E2F switch underlies the restriction point. *Nat. Cell Biol.* **10**, 476–482 (2008).

78. S. Huang, Y.-P. Guo, G. May, T. Enver, Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).

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

80. B. P. Kramer, A. U. Viretta, M. Daoud-El-Baba, D. Aubel, W. Weber, M. Fussenegger, An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004).

81. C. M. Ajo-Franklin, D. A. Drubin, J. A. Eskin, E. P. S. Gee, D. Landgraf, I. Phillips, P. A. Silver, Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007).

82. T. Lebar, U. Bezeljak, A. Golob, M. Jerala, L. Kadunc, B. Pirš, M. Stražar, D. Vučko, U. Zupančič, M. Benčina, V. Forstnerič, R. Gaber, J. Lonzarić, A. Majerle, A. Oblak, A. Smole, R. Jerala, A bistable genetic switch based on designable DNA-binding domains. *Nat. Commun.* **5**, 5007 (2014).

83. D. R. Burrill, M. C. Inniss, P. M. Boyle, P. A. Silver, Synthetic memory circuits for tracking human cell fate. *Genes Dev.* **26**, 1486–1497 (2012).

84. J. Santos-Moreno, E. Tasiudi, J. Stelling, Y. Schaerli, Author Correction: Multistable and dynamic CRISPRi-based synthetic circuits. *Nat. Commun.* **12**, 3119 (2021).

85. F. Wu, R.-Q. Su, Y.-C. Lai, X. Wang, Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. *Elife*. **6** (2017), doi:10.7554/eLife.23702.

86. A. Turning, The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. **237**, 37–72 (1952).

87. K. W. Rogers, A. F. Schier, Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407 (2011).

88. J. M. Wozney, V. Rosen, A. J. Celeste, L. M. Mitsock, M. J. Whitters, R. W. Kriz, R. M. Hewick, E. A. Wang, Novel regulators of bone formation: molecular clones and activities. *Science*. **242**, 1528–1534 (1988).

89. R. N. Wang, J. Green, Z. Wang, Y. Deng, M. Qiao, M. Peabody, Q. Zhang, J. Ye, Z. Yan, S. Denduluri, O. Idowu, M. Li, C. Shen, A. Hu, R. C. Haydon, R. Kang, J. Mok, M. J. Lee, H. L. Luu, L. L. Shi, Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis*. **1**, 87–105 (2014).

90. E. Bier, E. M. De Robertis, EMBRYO DEVELOPMENT. BMP gradients: A paradigm for morphogen-mediated developmental patterning. *Science*. **348**, aaa5838 (2015).

91. D. Umulis, M. B. O'Connor, S. S. Blair, The extracellular regulation of bone morphogenetic protein signaling. *Development*. **136**, 3715–3728 (2009).

92. L. Zakin, E. M. De Robertis, Extracellular regulation of BMP signaling. *Curr. Biol.* **20**, R89–92 (2010).

93. W. Balemans, W. Van Hul, Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev. Biol.* **250**, 231–250 (2002).

94. S. Piccolo, Y. Sasai, B. Lu, E. M. De Robertis, Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell*. **86**, 589–598 (1996).

95. J. J. Ross, O. Shimmi, P. Vilmos, A. Petryk, H. Kim, K. Gaudenz, S. Hermanson, S. C. Ekker, M. B. O'Connor, J. L. Marsh, Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature*. **410**, 479–483 (2001).

96. C. Chang, D. A. Holtzman, S. Chau, T. Chickering, E. A. Woolf, L. M. Holmgren, J. Bodorova, D. P. Gearing, W. E. Holmes, A. H. Brivanlou, Twisted gastrulation can function as a BMP antagonist. *Nature*. **410**, 483–487 (2001).

97. I. C. Scott, I. L. Blitz, W. N. Pappano, S. A. Maas, K. W. Y. Cho, D. S. Greenspan, Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling. *Nature*. **410**, 475–478 (2001).

98. O. Shimmi, M. B. O'Connor, Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the Drosophila embryo. *Development*. **130**, 4673–4682 (2003).

99. M. B. O'Connor, D. Umulis, H. G. Othmer, S. S. Blair, Shaping BMP morphogen gradients in the Drosophila embryo and pupal wing. *Development*. **133**, 183–193 (2006).

100. I. C. Scott, I. L. Blitz, W. N. Pappano, Y. Imamura, T. G. Clark, B. M. Steiglitz, C. L. Thomas, S. A. Maas, K. Takahara, K. W. Cho, D. S. Greenspan, Mammalian BMP-1/Tolloid-related metalloproteinases, including novel family member mammalian

Tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. *Dev. Biol.* **213**, 283–300 (1999).

101. C. M. Mizutani, Q. Nie, F. Y. M. Wan, Y.-T. Zhang, P. Vilmos, R. Sousa-Neves, E. Bier, J. L. Marsh, A. D. Lander, Formation of the BMP activity gradient in the Drosophila embryo. *Dev. Cell*. **8**, 915–924 (2005).

102. L. Zakin, C. A. Metzinger, E. Y. Chang, C. Coffinier, E. M. De Robertis, Development of the vertebral morphogenetic field in the mouse: interactions between Crossveinless-2 and Twisted Gastrulation. *Dev. Biol.* **323**, 6–18 (2008).

Chapter 2. Synthetic multistability in mammalian cells

Abstract

In multicellular organisms, gene regulatory circuits generate thousands of molecularly distinct, mitotically heritable states, through the property of multistability. Designing synthetic multistable circuits would provide insight into natural cell fate control circuit architectures and allow engineering of multicellular programs that require interactions among cells in distinct states. Here we introduce MultiFate, a naturally-inspired, synthetic circuit that supports long-term, controllable, and expandable multistability in mammalian cells. MultiFate uses engineered zinc finger transcription factors that transcriptionally self-activate as homodimers and mutually inhibit one another through heterodimerization. Using a model-based design, we engineered MultiFate circuits that generate up to seven states, each stable for at least 18 days. MultiFate permits controlled state-switching and modulation of state stability through external inputs, and can be easily expanded with additional transcription factors. Together, these results provide a foundation for engineering multicellular behaviors in mammalian cells.

2.1 Introduction

Multistability allows genetically identical cells to exist in thousands of molecularly distinct and mitotically stable cell types or states (*1*, *2*). Understanding natural multistable circuits and engineering synthetic ones have been long-standing challenges in developmental and synthetic biology (*3–14*). Building synthetic multistable circuits could provide insight into the minimal circuitry sufficient for multistability, and establish a

foundation for exploiting multicellularity in engineered cell therapies. However, efforts in mammalian cells have been limited to two-state systems and used architectures that cannot be easily expanded to larger numbers of states (*5*–*7*). An ideal synthetic multistable system would allow cells to remain in any of a set of distinct expression states over many cell cycles, despite biological noise. In addition, it would provide three key capabilities exhibited by its natural counterparts (Fig. 2.1A): First, it would permit transient external inputs to switch cells between states, similar to the way signaling pathways direct fate decisions (*15*, *16*). Second, it would support control over the stability of different states, and enable irreversible transitions, similar to those that occur during natural differentiation (*13*, *14*). Third, it would be expandable by introducing additional components without re-engineering an existing functional circuit, analogous to expansion of cell types during evolution (*17*).

Natural mammalian multistable circuits provide inspiration for such a synthetic architecture. In many natural fate control systems, transcription factors positively autoregulate their own expression, and competitively interact with one another to form a variety of homodimers, heterodimers, and higher order multimeric forms (Fig. 2.1B) (*18*–*24*). For example, during myogenesis, muscle regulatory factors (MRF) such as MyoD heterodimerize with E proteins to activate their own expression and the broader myogenesis program, while Id family proteins disrupt this process through competitive dimerization (*23*, *24*). Similarly, during embryogenesis, Sox2 and Sox17 competitively interact with Oct4 to control fate decisions between pluripotency and endodermal differentiation (*21*, *22*). Related combinations of positive autoregulation and cross-inhibition could extend multistability behaviors beyond bistability and generate

bifurcation dynamics that explain the partial irreversibility of cell differentiation (*9*, *12*). Nevertheless, it remains unclear whether these natural architectures could be adapted to enable synthetic multistability. Here, we show how a synthetic multistable system based on principles derived from natural cell fate control systems can generate robust, controllable, expandable multistability in mammalian cells.

Fig. 2.1

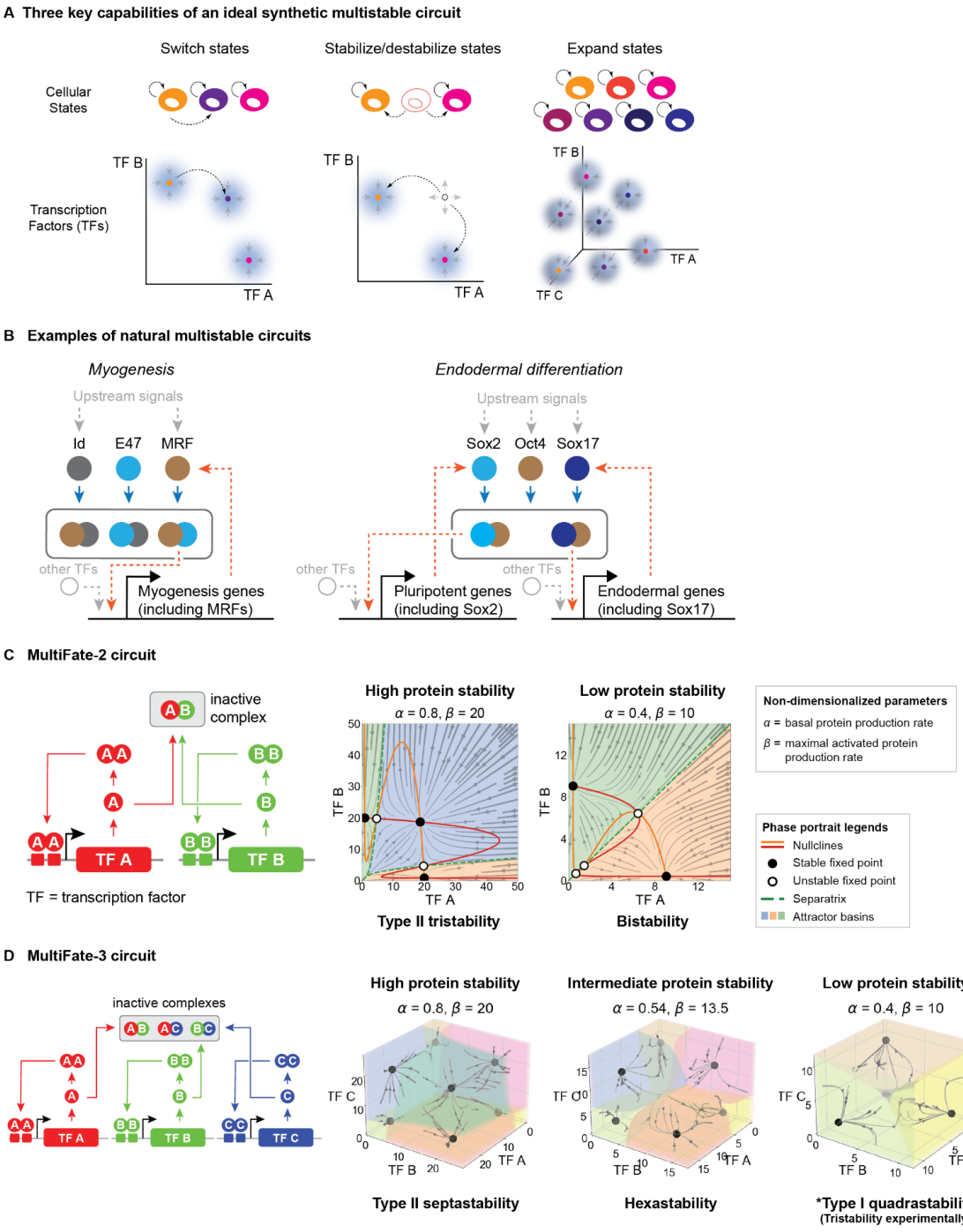

Fig. 2.1 The naturally-inspired MultiFate architecture generates diverse types of multistability in the model.

(A) A hypothetical synthetic multistable circuit is represented by colored cell cartoons (upper level) and attractors in a transcription factor phase space (lower level schematic, TF A-C on coordinate axes represent transcription factor concentrations). An ideal synthetic multistable circuit should generate multiple stable states, support control of

state-switching (left) and state stability (middle), and allow easy expansion of states by addition of more transcription factors (right). (B) Competitive protein-protein interactions and autoregulatory feedback are prevalent in natural multistable circuits that control myogenesis (left) and endodermal differentiation (right) as shown by these simplified and abridged diagrams. Blue arrows indicate competitive protein-protein interactions, which can involve higher order multimerization. Orange dashed arrows indicate direct or indirect positive transcriptional feedback. (C, D) Models of the MultiFate-2 circuit and MultiFate-3 circuit (Box 2.1) (*25*) generate diverse types of multistability in different parameter regimes (indicated above plots). In the model of the MultiFate-3 circuit, low protein stability generates 4 stable states (type I quadrastability), but the state in which all transcription factors are lowly expressed is unstable in the presence of biological noise (Fig. S2.23), consistent with experimental results in Fig. 2.5B, Low TMP columns. Complete lists of multistability regimes are shown in Fig. S2.1 and S2.2. All models used here are symmetric and non-dimensionalized, with rescaled dimerization dissociation constant $K_d = 1$ and Hill coefficient $n = 1.5$ (Box 2.1). For both panels, each axis represents the dimensionless total concentration of each transcription factor. Note that in the non-dimensionalized model, changing protein stability is equivalent to multiplying α and β with the same factor (Box 2.1).

2.2 MultiFate generates diverse types of multistability through a set of promiscuously dimerizing, autoregulatory transcription factors

Inspired by natural fate control circuits, we designed a new synthetic multistable system called MultiFate (Fig. 2.1C). In MultiFate system, transcription factors share a common dimerization domain, allowing them to competitively form both homodimers and heterodimers. The promoter of each transcription factor gene contains binding sites that can be strongly bound only by its own homodimers, allowing homodimer-dependent self-activation. By contrast, heterodimers do not efficiently bind to any promoter in this design. Heterodimerization thus acts to mutually inhibit the activity of both constituent transcription factors.

Mathematical modeling shows how the MultiFate architecture provides each of the desired capabilities described above (Fig. 2.1A) in physiologically reasonable parameter regimes (Box 2.1 and Table S2.1) (*25*). A MultiFate circuit with just two transcription

factors, designated MultiFate-2, can produce diverse types of multistability containing 2, 3, or 4 stable fixed points depending on protein stability and other parameter values (Figs. 2.1C and S2.1A). In particular, a regime designated type II tristability is analogous to multilineage priming in uncommitted progenitor cells, with the double positive state playing the role of a multipotent progenitor (*26–28*). Transient expression of one transcription factor can switch cells between states (Fig. S2.3). Reducing the protein stability of transcription factors can cause bifurcations that selectively destabilize certain states (Fig. 2.1C and S2.1A). Finally, the model is expandable: addition of a new transcription factor to the MultiFate-2 model generates a MultiFate-3 circuit that supports additional stable states with the same parameter values (Figs. 2.1D and S2.2A). Together, these modeling results suggest that the MultiFate architecture can support a rich array of multistable behaviors.

Fig. 2.2

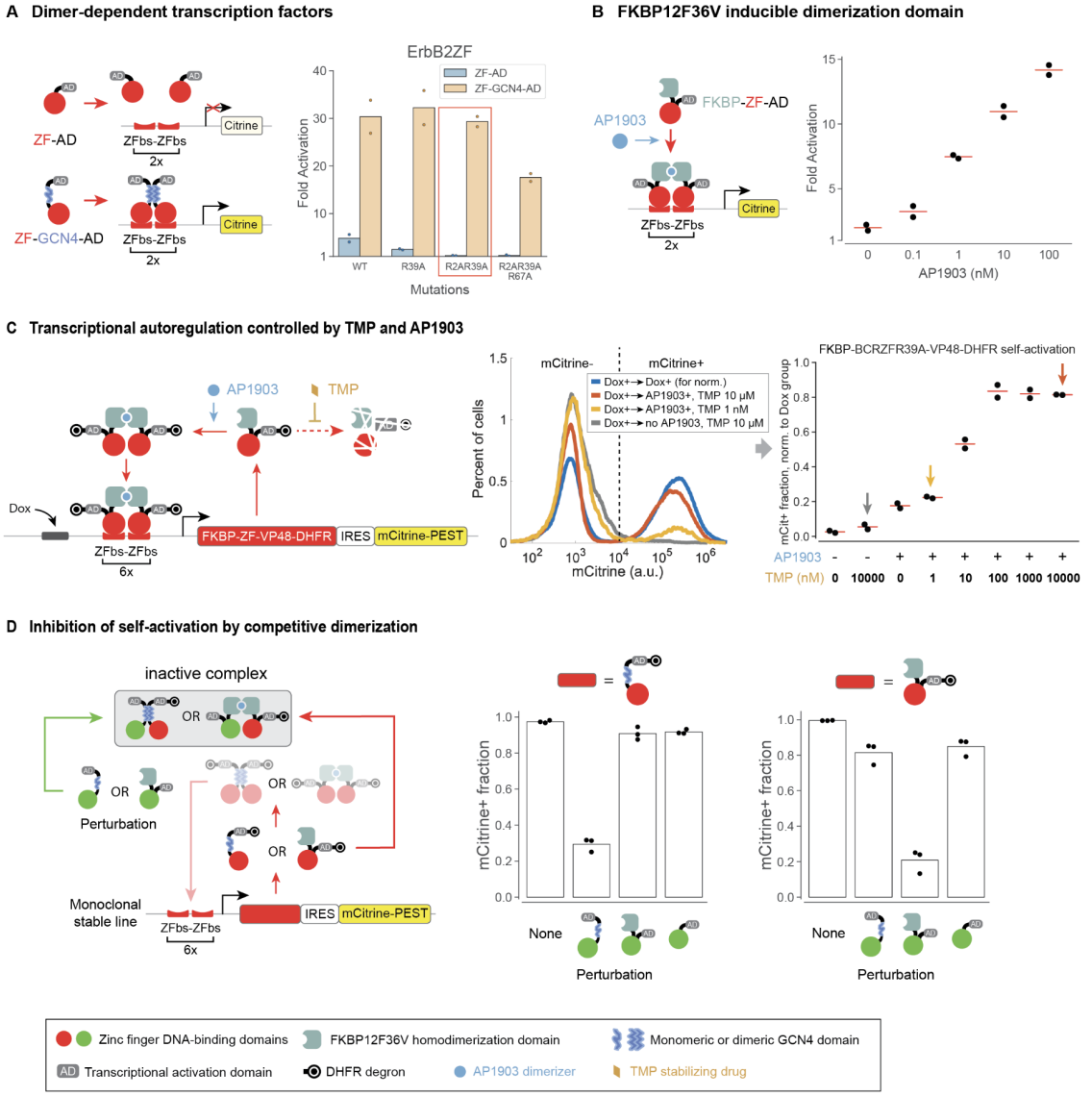

Fig. 2.2 Engineered transcription factors enable homodimer-dependent autoregulation and heterodimerization-based inhibition.

(A) ZF transcription factors enable homodimer-dependent activation. (Left) Design of test constructs, in which ErbB2ZF (*29*) (red circle) fused to VP48 (AD) and in some cases GCN4 (blue squiggle) domains bind to target sites (red pads) to activate Citrine expression. Activators were expressed from a constitutive CAG promoter (*30*). (Right) R-to-A mutations in ZF modulated reporter activation by ZF-GCN4-AD and ZF-AD. The R2AR39A variant was selected due to high ZF-GCN4-AD activation and minimal ZF-AD activation. Fold activation is defined in Fig. S2.4A. WT = wild-type variant. (B) FKBP12F36V (FKBP) (light cyan partial box) allows dose-dependent control of activation by AP1903 (cyan circle). Red circle = BCRZFR39A. (C) Transcription factor self-activation can be controlled by TMP and AP1903. (Left) Design of the controllable self-activation circuit. IRES = internal ribosome entry site; PEST = constitutive

degradation tag (*31*); (Middle) Stable polyclonal cells showed bimodal mCitrine distribution upon circuit activation. An empirical threshold at mCitrine = 10^4 separates the distribution into two fractions, and the normalized mCitrine+ fraction was used to quantify the self-activation strength (*25*). (Right) Colored arrows indicate data from the middle panel. In AP1903+ samples, AP1903 concentration was 100 nM. (D) Self-activation was inhibited by proteins with a different ZF and matching dimerization domains. Two monoclonal stable lines could spontaneously self-activate in media containing AP1903 and TMP (Fig. S2.5B). Each perturbation construct is introduced by stable integration (*25*). The integrated plasmid in the "None" group did not express any perturbation protein. Red circle = 42ZFR2AR39AR67A; Green circle = BCRZFR39A. In all panels, each dot represents one biological replicate, and each red line or bar indicates the mean of replicates. Lists of constructs and cell lines are in Table S2.2 and S2.3.

2.3 Engineered zinc fingers enable homodimer-dependent self-activation and heterodimer-dependent inhibition

Synthetic zinc finger (ZF) transcription factors provide an ideal platform to implement the MultiFate circuit. They can recognize and activate a promoter containing target DNA binding sites with high specificity (*32*, *33*). Further, engineered ZF DNA-binding domains containing three fingers bind weakly as monomers to 9bp target sites, but can bind much more strongly as homodimers to 18bp tandem binding site pairs (*29*, *34*). This property allows homodimer-dependent transcriptional activity and potentially allows inhibition through heterodimerization.

To engineer ZF transcription factors, we fused the ErbB2 ZF DNA-binding domain to a GCN4 homodimerization domain and a VP48 transcriptional activation domain to create the synthetic transcription factor, termed ZF-GCN4-AD (Fig. 2.2A) (*29*). A transcription factor (ZF-AD) lacking GCN4 was used as a monomeric control. To assay their transcriptional activity, we constructed a reporter containing 18bp homodimer binding sites driving the expression of Citrine (*29*). We then co-transfected each transcription factor together with the reporter and an mTagBFP2 (*35*) co-transfection marker into

Chinese hamster ovary K1 (CHO-K1) cells, and analyzed Citrine expression by flow cytometry 36 hours later (Figs. 2.2A and S2.4A) (*25*). The wild-type (WT) ZF-GCN4-AD factors strongly activated the reporter, as desired, whereas ZF-AD exhibited weaker, but still undesirable, basal activity (Figs. 2.2A and S2.4B). Following previous work (*32*, *36*, *37*), we incorporated arginine-to-alanine mutations at key positions in the ZF known to weaken DNA binding, which decreased monomeric activity without reducing homodimer activity (bars within red square in Fig. 2.2A). Replacing GCN4 with the FKBP12F36V (FKBP) homodimerization domain (*38*) allowed us to achieve dose-dependent control of dimerization with the small molecule AP1903 (Fig. 2.2B). Finally, we repeated this general design to engineer a set of additional homodimer-dependent ZF transcription factors with orthogonal DNA-binding specificities (Fig. S2.4B and S2.4C).

The MultiFate circuit design requires that each transcription factor positively autoregulates its own expression in a homodimer-dependent manner. To validate this capability, we designed a self-activation construct (Fig. 2.2C, left), in which a transcription factor with FKBP dimerization domain is expressed from a promoter containing its own 18bp homodimer binding sites (Table S2.2). This construct allowed independent Dox-inducible activation through upstream Tet3G (Takara Bio) binding sites. It also incorporated a dihydrofolate reductase (DHFR) degron (*39*), which can be inhibited by trimethoprim (TMP), permitting control of protein stability. Finally, we incorporated a destabilized mCitrine for dynamic readout of construct expression. We integrated this construct into Tet3G-expressing CHO-K1 cells, generating a stable polyclonal population for further analysis (Table S2.3) (*25*).

To test for self-activation, we transiently induced transcription factor expression for 24 hours with Dox, and then withdrew Dox and checked whether cells could sustain circuit activation when dimerization strength and protein stability were varied by AP1903 and TMP, respectively. In the presence, but not the absence, of AP1903, cells exhibited a bimodal distribution of mCitrine fluorescence, with well-separated peaks (Fig. 2.2C, middle), consistent with homodimer-dependent self-activation in a subset of cells. TMP, by stabilizing transcription factors, also promoted self-activation in a dose-dependent manner (Figs. 2.2C and S2.5A). Thus a single dimer-dependent transcription factor can self-activate and sustain its own expression in a controllable manner.

MultiFate's final requirement is the ability of one transcription factor to effectively inhibit another through heterodimerization. To test this, we selected monoclonal cell lines with the self-activating circuits, and then stably integrated constructs expressing proteins with a different ZF DNA-binding domain and a matching or mismatching dimerization domain to generate a polyclonal cell population for each perturbation construct (Table S2.2 and Table S2.3) (*25*). Consistent with inhibition through heterodimerization, the proteins with matching dimerization domains strongly inhibited the self-activating transcription factor, while similar proteins with non-matching dimerization domains exhibited much weaker inhibition, possibly through non-specific mechanisms (Figs. 2.2D and S2.5B). Taken together, these results provided a set of engineered ZF transcription factors that exhibit controllable homodimer-dependent activation and heterodimer-dependent inhibition.

Fig. 2.3

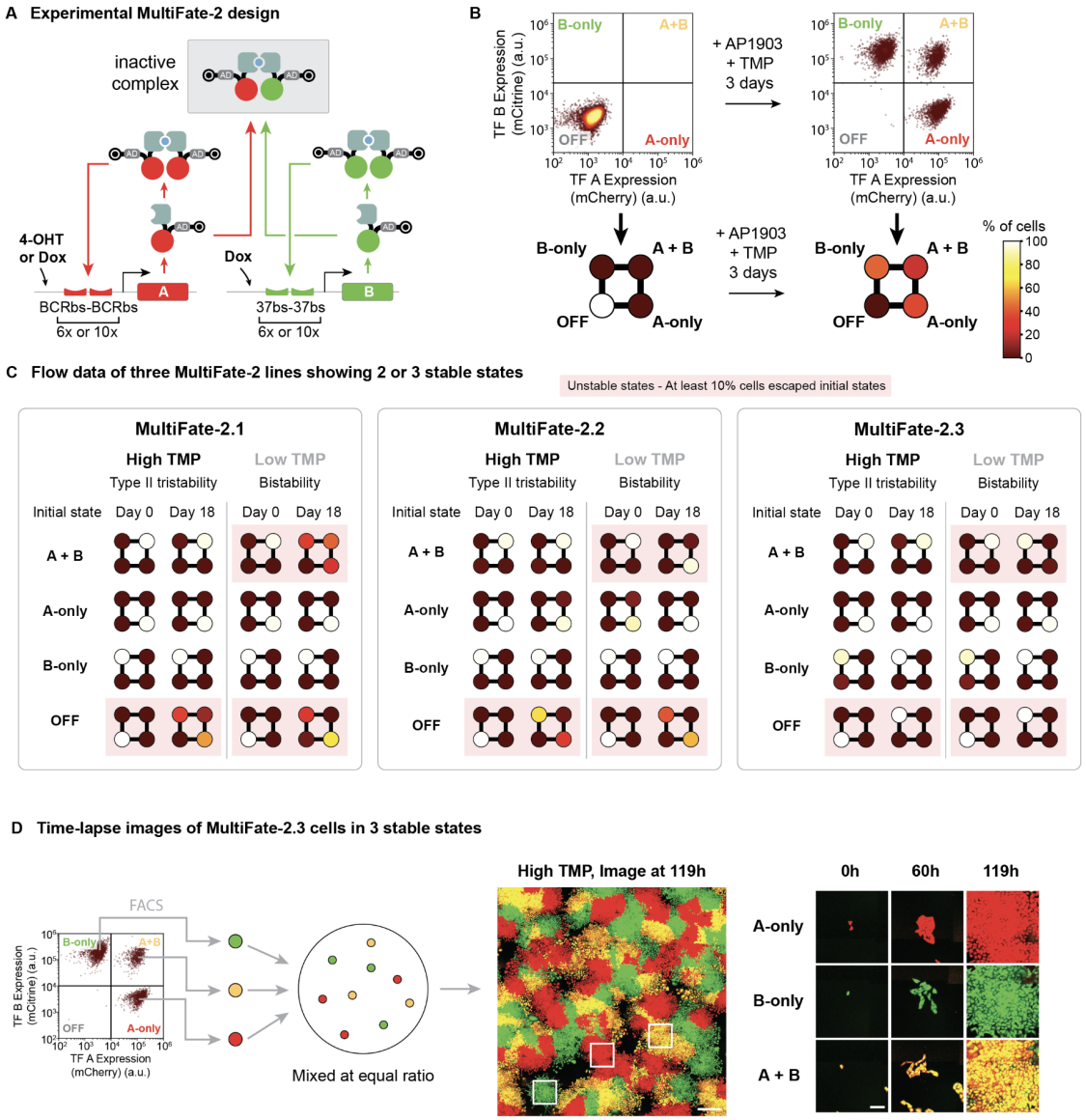

Fig. 2.3 MultiFate-2 generates multiple stable states.
(A) The experimental MultiFate-2 design uses two self-activation cassettes differing only in their ZF DNA-binding domains and binding sites, and fluorescent proteins. Each cassette expresses FKBP-ZF-VP16-DHFR-IRES-FP-PEST, where ZF represents either BCRZFR39A or 37ZFR2AR11AR39AR67A and FP represents either mCherry or mCitrine, for A and B, respectively. Detailed construct maps and differences among MultiFate-2 lines are shown in Table S2.2 and Table S2.3. (B) MultiFate-2.1 cells spontaneously activate A, B or both cassettes upon addition of 100 nM AP1903 and 10 μM TMP. Cell percentages in OFF, A-only, B-only and A+B states were quantified and plotted as a square with four colored circles (*25*). (C) Three MultiFate-2 lines all exhibited type II tristability in the High TMP condition, and bistability in the Low TMP condition. In all conditions, we added 100 nM AP1903. Exact concentrations of TMP are shown in Fig. S2.9-2.11. Unstable states, defined by states having more than 10% cells

escaping their initial states after 18 days, were marked in pink rectangles. Each square represents the mean fractions of three biological replicates. Initial A-only, B-only and A+B cells were sorted from a population of cells in different states, while initial OFF cells came from cells in regular CHO media without any inducers. (D) A-only, B-only and A+B states were each stable during growth from single MultiFate-2.3 cells into colonies over 5 days under a time-lapse microscope. (Left) We first sorted a mixed MultiFate-2.3 cell population to separate cells in 3 different states. Then we seeded cells in these three states at equal ratio in the same well and performed time-lapse imaging (*25*). (Right) Scale bar: 500 μm for the wide field image, 100 μm for zoomed in images. "High TMP" = 100 nM AP1903 + 10 μM TMP.

2.4 The MultiFate-2 circuit generates tristability

To construct a complete MultiFate circuit, we selected two dimer-dependent transcription factors, henceforth designated A and B, with distinct DNA binding specificities but the same FKBP homodimerization domain. Their expressions are driven by promoters containing multiple repeats of their corresponding 18bp homodimeric binding sites (Fig. 2.3A and Table S2.2). The promoters also incorporated Tet3G or ERT2-Gal4 response elements (*40*) to allow independent external activation of transcription. A and B were transcriptionally co-expressed with destabilized mCherry or mCitrine fluorescent proteins, respectively, each placed after an internal ribosome entry site (IRES), allowing fluorescent readout of transcription rates in individual cells (Fig. S2.6). We stably integrated both genes simultaneously in CHO-K1 cells expressing Tet3G and ERT2-Gal4 proteins, selected and further characterized three stable monoclonal cell lines, designated MultiFate-2.1, MultiFate-2.2 and MultiFate-2.3 with different promoter configurations (Fig. S2.7A and Table S2.3) (*25*).

To test whether MultiFate circuits support multistability, we activated the circuit by transferring MultiFate-2.1 cells to media containing AP1903 and TMP to allow dimerization and stabilizing the transcription factors. As expected in the regime of type II

tristability (Fig. 2.1C), cells went from low expression of both transcription factors (OFF state) to one of three distinct states, with either A, B, or both transcription factors highly expressed (Fig. 2.3B). We designate these states A-only, B-only and A+B, respectively. The three states were well-separated by ~25-to-50-fold differences in either mCherry or mCitrine expression, and cells grew at similar rates among states (Fig. S2.8). To assess their stability, we sorted cells from each of these states and cultured them continuously for 18 days (*25*). Strikingly, nearly all cells remained in the sorted state for this extended period (Fig. 2.3C, MultiFate-2.1 High TMP columns, and Fig. S2.9), despite gene expression noise (observable from the spread of cellular fluorescence on flow cytometry plots). This showed that cells were attracted to these states. Stability required the positive autoregulation, as withdrawal of AP1903 and TMP collapsed expression of both factors within 2 days (Fig. S2.9). Similar overall behavior was also observed in MultiFate-2.2 and MultiFate-2.3 (Figs. 2.3C, S2.10 and S2.11). All three MultiFate-2 cell lines thus exhibited dynamics consistent with type II tristability (Fig. 2.1C).

Time-lapse imaging provided a more direct view of multistability. We cultured an equal ratio of single cells sorted from three different initial states in the same well and imaged them as they developed into colonies (Fig. 2.3D) (*25*). In almost all colonies (132 of 134), all cells maintained their initial states for the full duration of the movie, at least 5 days or 7 to 8 cell cycles (Figs. 2.3D, S2.12A, S2.13A). Together with the flow cytometry analysis, these results demonstrate that all three MultiFate-2 lines can sustain long-term tristability.

Fig. 2.4

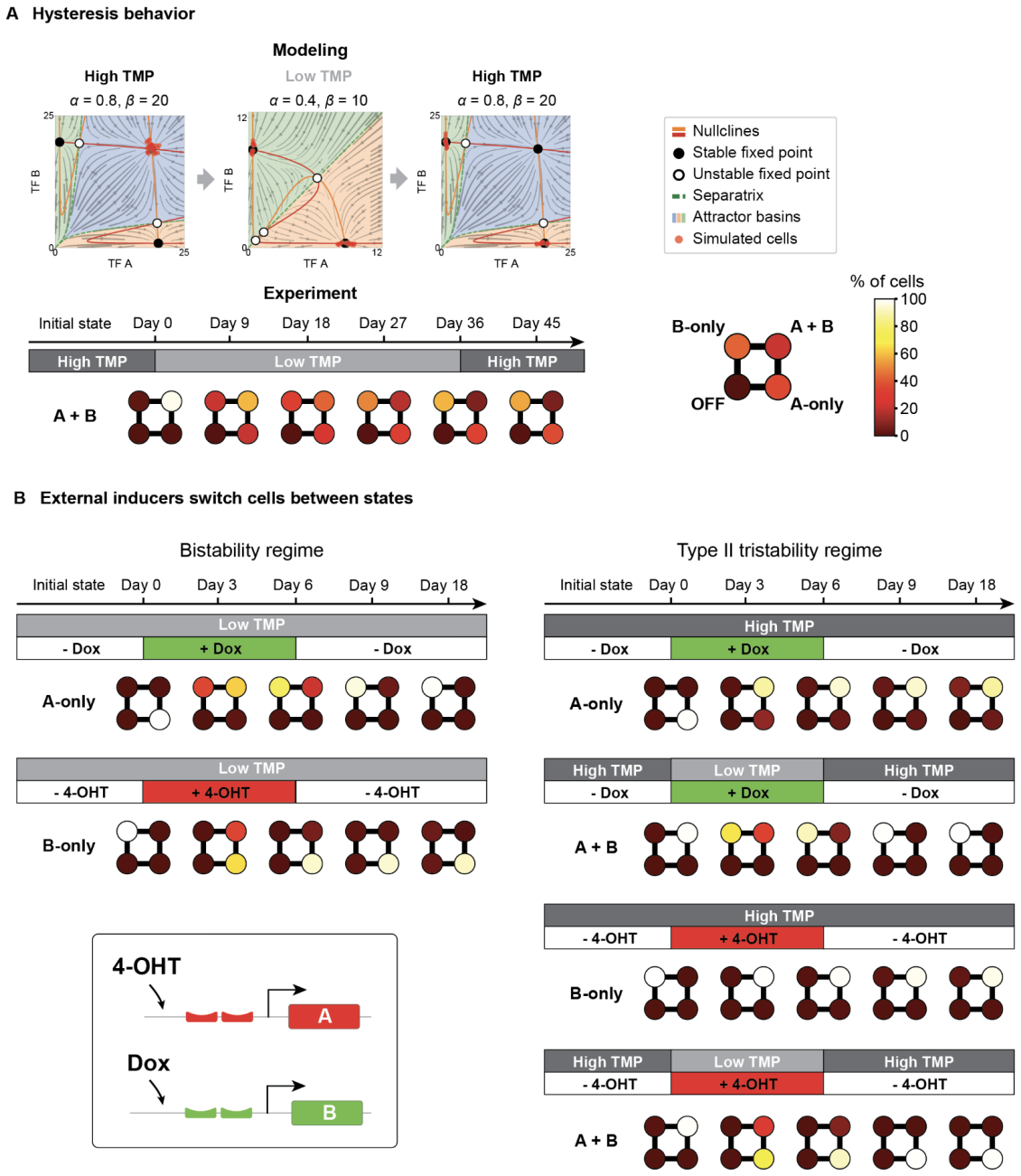

Fig. 2.4 MultiFate-2 supports modulation of state stability and allows state-switching.

(A) Escape from the destabilized A+B state was irreversible, as shown by both modeling, and experiment using MultiFate-2.1 cells. (Top) The model used here is symmetric and non-dimensionalized, with rescaled dimerization dissociation constant $K_d = 1$ and Hill coefficient $n = 1.5$ (Box 2.1). The x and y axes are total dimensionless concentrations of TF A and TF B, respectively. Simulated cells on phase portraits were calculated using the Gillespie algorithm (*41*). Note that in the non-dimensionalized model, changing protein stability is equivalent to multiplying α and β with the same factor (Box 2.1).

(Bottom) Throughout the experiment, we added 100 nM AP1903. Exact concentrations of TMP are shown in Fig. S2.9. (B) MultiFate-2.3 cells can be switched between states by transient 4-OHT or Dox treatment. In all conditions, we added 100 nM AP1903. Exact concentrations of TMP are shown in Fig. S2.16. 4-OHT = 25 nM, Dox = 500 ng/ml. In all panels, initial A-only, B-only and A+B cells were sorted from a population of cells in different states. Each square represents the mean fractions of three biological replicates.

2.5 MultiFate-2 supports modulation of state stability and allows controlled state-switching

The ability of a transient stimulus to destabilize multipotent states and trigger an irreversible fate change is a hallmark of many cell fate control systems (*12–14*). In the model, reducing protein stability can eliminate the A+B state but preserve A-only and B-only states (Fig. 2.1C). As a result, cells initially occupying the A+B state transit to A-only or B-only states (Fig. 2.4A, top). When protein stability is restored to its initial value, the A+B attractor reappears. However, for the parameter sets analyzed here, cells remain within the attractor basins of A-only and B-only states, and therefore do not return to the A+B state (Fig. 2.4A, top). Stochastic simulations of single cell dynamics confirmed this irreversible (hysteretic) behavior (Fig. 2.4A, top).

To test whether similar bifurcation and hysteretic dynamics occur in the experimental system, we transferred A-only, B-only and A+B cells from media containing high TMP concentrations (“High TMP”) to similar media with reduced TMP concentrations (“Low TMP”), which decreased protein stability by permitting degron function. As predicted, reducing protein stability selectively destabilized the A+B state, but not the A-only and B-only states, shifting cells from A+B state to the A-only or B-only states (Fig. 2.3C, Low TMP columns, Fig. 4A, bottom). Different MultiFate-2 cell lines exhibited different transition biases, reflecting clone-specific asymmetries in the experimental MultiFate-2

systems (Figs. 2.3C, S2.9 to S2.11), in a manner consistent with an asymmetric MultiFate model (Figs. S2.14 and S2.15) (*25*). Escape from the destabilized A+B state was irreversible, as cells remained in the A-only or B-only state even after they were transferred back to the High TMP media (Fig. 2.4A, bottom, and Fig. S2.9). Thus, MultiFate's ability to support irreversible transitions allows it to produce behaviors resembling stem cell differentiation.

Finally, we asked to what extent we could deliberately switch cells from one state to another through transient perturbations. We used MultiFate-2.3, in which the A and B genes can be independently activated by 4-hydroxy-tamoxifen (4-OHT) and Dox, respectively, to address this question. In this line, the response elements for the inducers are adjacent to the homodimer binding sites. Therefore, the addition of inducers increases A or B expression up to, but not substantially beyond, the level produced by self-activation (Figs. 2.2C and S2.16). In the bistable regime, transient induction of either transcription factor switched cells into the corresponding state, where they remained in the absence of further induction (Fig. S2.3A, Fig. 2.4B, left, and Fig. S2.16A). In the tristable regime, the model predicted, and experiments confirmed, that transient induction of B by Dox could switch A-only cells to the A+B state, but not beyond it to the B-only state (Fig. S2.3B first row, Fig. 2.4B, top right, and Fig S2.16B). Combining transient Dox addition to induce B expression with TMP reduction to destabilize the A+B state successfully transitioned cells from A+B to the B-only state (Fig. S2.3B, second row, and Fig. 2.4B, right second row). The reciprocal experiments, in which we induced A expression with 4-OHT with or without reduced TMP, produced equivalent results (Fig. 2.4B, right column, lower two rows). Taken together, these results demonstrate that

MultiFate-2 circuits allow modulation of state stability, irreversible cell state transitions, and direct control of state-switching with transient external inducers.

Fig. 2.5

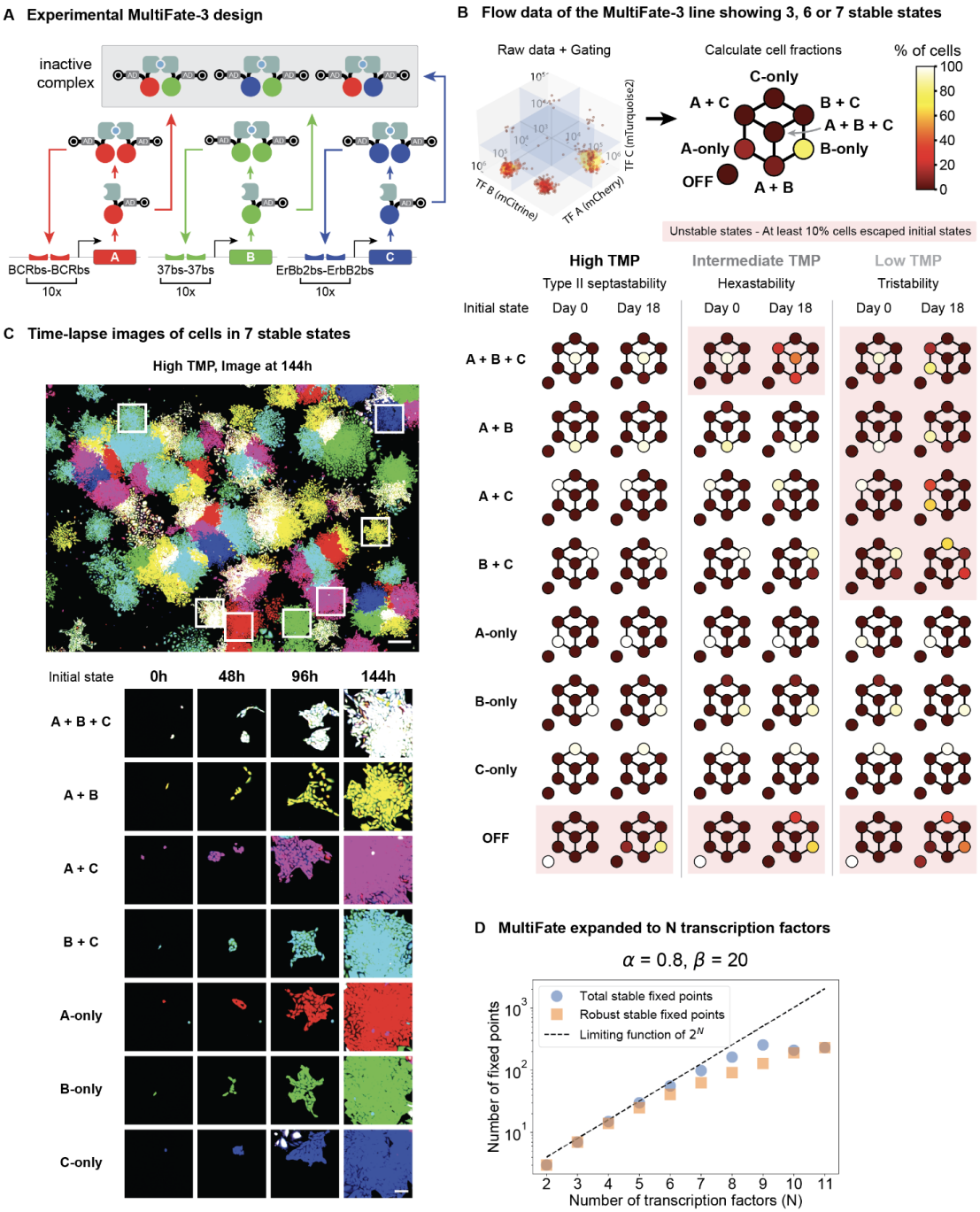

Fig. 2.5. MultiFate architecture is expandable to include three and potentially even more transcription factors.

(A) The experimental MultiFate-3 design uses three self-activation cassettes differing only in their ZF DNA-binding domains and binding sites, and fluorescent proteins. Each cassette expresses FKBP-ZF-VP16-DHFR-IRES-FP-PEST, where ZF represents either BCRZFR39A, 37ZFR2AR11AR39AR67A or ErbB2ZFR2AR39A, and FP represents either mCherry, mCitrine or mTurquoise2, for A, B and C, respectively. Detailed

construct maps are available in Table S2. (B) The MultiFate-3 line exhibited type II septastability, hexastability and tristability in three different TMP conditions. (Top) State percentages in each octant were quantified and plotted as eight colored circles (*25*). (Bottom) High TMP condition = 100 nM AP1903 + 100 nM TMP; Intermediate TMP condition = 100 nM AP1903 + 40 nM TMP; Low TMP condition = 100 nM AP1903 + 10 nM TMP. Except for OFF state cells, cells in different initial states were sorted from a mixed population of cells in the High TMP condition. Initial OFF cells came from cells in regular CHO media without any inducers. Each plot represents the mean percentages of three biological replicates. (C) Cells in each of the seven states were stable during growth from single cells into colonies over 6 days under a time-lapse microscope. We sorted cells and seeded an equal ratio of cells in 7 states using the same method for Fig. 3D. Scale bar: 500 μm for the wide field image (left), 100 μm for zoomed in images (right). (D) MultiFate is expandable (model). The number of robust stable fixed points grows monotonically with the number of transcription factors species (N) in the model. A robust stable fixed point is defined as a stable fixed point that has fewer than 10% cells escaping at the end of stochastic simulations (*25*). The parameter set provided above the plot (with $K_d = 1$ and Hill coefficient $n = 1.5$) is the same non-dimensionalized parameter set used in MultiFate-2 and MultiFate-3 models under high protein stability.

2.6 MultiFate is expandable

Because the MultiFate system implements mutual inhibition among transcription factors through heterodimerization, it can be expanded by adding additional transcription factors, without re-engineering existing components. In the model, adding a third transcription factor to a MultiFate-2 circuit produces a range of new stability regimes containing 3, 4, 6, 7, or 8 stable fixed points, depending on parameter values (Figs. 2.1D, S2.2) (*25*). To test whether experimental MultiFate-2 circuits be similarly expanded, we stably integrated a third ZF transcription factor, denoted C, containing the same FKBP dimerization domain as A and B, co-expressed with a third fluorescent protein, mTurqoise2, into the MultiFate-2.2 cell line to obtain the MultiFate-3 cell line (Figs. 2.5A, S2.7B and Table S2.3) (*25*).

After the addition of AP1903 and TMP, MultiFate-3 cells went from low expression of all genes (OFF state) to one of seven distinct expression states, termed A-only, B-only,

C-only, A+B, A+C, B+C, and A+B+C states (Fig. 2.5B), consistent with a type II septastability regime (Figs 2.1D and S2.2A). Most cells occupied the B-only state (79.5%±0.3%), reflecting asymmetries within the circuit (Figs. S2.14 and S2.15). To assess the stability of these states, we sorted cells from each of the seven states, and continuously cultured them in media containing AP1903 and TMP, analyzing the culture every 3 days by flow cytometry (*25*). Remarkably, each of the seven states was stable for the full 18-day duration of the experiment (Fig. 2.5B, High TMP columns, and Fig. S2.17). Long-term stability required AP1903 and TMP, as expected (Fig. S2.18). Finally, cells from each state could be reset by withdrawal of AP1903 and TMP and then re-differentiated into all 7 states when AP1903 and TMP were added back (Fig. S2.18). This indicates that the observed stability is not the result of a mixture of clones permanently locked into distinct expression states.

To directly visualize the septastable dynamics of MultiFate-3, we co-cultured single cells sorted from each of the seven states and performed live imaging as they grew into colonies (*25*). Consistent with the flow cytometry results, cells retained their initial states for the full 6-day duration of the experiment in almost every colony (153 of 157) (Figs. 2.5C, S2.12B, S2.13B).

Like MultiFate-2, the number and stability of different states in MultiFate-3 can be modulated. In the model, reducing protein stability repeatedly bifurcates the system from type II septastability (7 stable states) through hexastability (6 stable states) to tristability (3 stable states) (Fig. 2.1D). This process resembles the progressive loss of cell fate potential during stem cell differentiation (*42*). To experimentally test this prediction, we transferred cells in each of the 7 states cultured under the High (100 nM) TMP condition

(high protein stability) to similar media with Intermediate (40 nM) or Low (10 nM) TMP conditions. As predicted by the model, the Intermediate TMP condition destabilized only the A+B+C state, but not the other 6 states (Fig. 2.5B, Intermediate TMP columns, and Fig. S2.19), whereas the Low TMP condition destabilized all multi-protein states, preserving only A-only, B-only and C-only states (Fig. 2.5B, Low TMP, and Fig. S2.20). Consistent with the model, these transitions were also irreversible: restoring High TMP concentrations did not cause cells to repopulate previously destabilized states (Fig. S2.21). Taken together, these results demonstrate that the MultiFate-3 circuit supports septastability, and allows controlled bifurcations to produce irreversible cell state transitions.

Can the MultiFate architecture be expanded beyond three transcription factors? To understand higher order systems, we modeled MultiFate circuits containing up to N=11 transcription factors (*25*). Using the same parameter values established for MultiFate-2 and MultiFate-3, the number of attractors reached a maximum of 256 at N=9. Analysis of attractor escape rates in stochastic simulations revealed that most of these attractors were robust to gene expression noise (Figs. 2.5D and S2.22) (*25*, *43*). The number of attractors grew more slowly than the theoretical limit of $\sim 2^N$ because stable attractors could only sustain high levels of up to four transcription factors at a time (Fig. S2.23, middle row). This limitation reflects the diminishing share of the active homodimers relative to all dimers. Similarly, the combined basal expression of all transcription factors suppressed homodimer formation, resulting in a decline in the number of attractors for systems containing more than 9 transcription factors (Fig. 2.5D and Fig. S2.23, middle row). Finally, we note that the precise values of the maximum number of stable attractors can

be modulated up or down by parameters that impact overall gene expression (Fig. S2.23). Together, these results indicate that the MultiFate architecture can be expanded to generate large numbers of robust stable states.

2.7 Discussion

The astonishing diversity of cell types in our own bodies underscores the critical importance of multistable circuits and provokes the fundamental question of how to engineer a robust, controllable, and expandable synthetic multistable system. We took inspiration from two ubiquitous features of natural multistable systems, namely competitive protein-protein interactions and transcriptional autoregulation, to design a synthetic multistable architecture that operates in mammalian cells. The MultiFate circuits exhibit many of the hallmarks of natural cell fate control systems. They generate as many as seven molecularly distinct, mitotically heritable cell states (Figs. 2.3 and 2.5). They allow controlled switching of cells between states with transient transcription factor expression (Fig. 2.4B), similar to fate reprogramming (*16*). They support modulation of state stability (Figs. 2.3 and 2.5) and permit irreversible cellular transitions through externally controllable parameters such as protein stability (Figs. 2.4A and S2.21), similar to the irreversible loss of cell fate potential during stem cell differentiation (*12*). Finally, implementing cross-inhibition at the protein level makes MultiFate expandable by ‘plugging in’ additional transcription factors, without re-engineering the existing circuit, a useful feature for synthetic biology. The same design principle may play a related role in natural systems, allowing the emergence of new cell states through transcription factor duplication and subfunctionalization in a manner analogous to the stepwise expansion of MultiFate circuits demonstrated here (*21*, *22*, *44*, *45*).

A remarkable feature of this circuit is its close agreement with predictions from a dynamical systems model (Box 2.1). Despite a lack of precise quantitative parameter values for many molecular interactions, the qualitative behaviors possible with this circuit design can be enumerated and explained from simple properties of the components and their interactions. More precise measurements of effective biochemical parameters and stochastic fluctuations could help to explain, eliminate, or exploit asymmetries and provide a better understanding of the timescales of state transitions.

MultiFate has a relatively simple structure, requiring a small number of genes, all of the same type, yet exhibits robust memory behaviors, scalability, and predictive design. Future work should extend MultiFate into a full-fledged synthetic cell fate control system. Coupling MultiFate to synthetic cell-cell communication systems such as synNotch (*46*, *47*), MESA (*48*), synthekines (*49*), engineered GFP (*50*) and auxin (*51*) should enable navigation of cells through a series of fate choices, recapitulating cell behaviors associated with normal development. MultiFate could also allow engineering of multicellular cell therapeutic programs. For example, one could engineer a stem-like state that can either self-renew or “differentiate” into other states that recognize and remember different input signals and communicate with one another to coordinate complex response programs. Such strategies will benefit from the ability of MultiFate to allow probabilistic differentiation into multiple different states in the same condition (Fig. S2.14). In this way, we anticipate that the MultiFate architecture will provide a scalable foundation for exploring the circuit-level principles of cell fate control and enable new multicellular applications in synthetic biology.

Box 2.1 Design of the MultiFate circuit

Here we introduce the mathematical model of the MultiFate circuit and show how it can be used to design the experimental system and predict its behavior. For simplicity, we focus on a symmetric MultiFate-2 circuit whose two transcription factors share identical biochemical parameters and differ only in their DNA binding site specificity. A similar analysis of systems with more transcription factors and asymmetric parameters is presented in (*25*).

We represent the dynamics of protein production and degradation using ordinary differential equations (ODEs) for the total concentrations of the transcription factors A and B, denoted $[A_{tot}]$ and $[B_{tot}]$, respectively. We assume that the rate of production of each protein follows a Hill function of the corresponding homodimer concentration, $[A_2]$ or $[B_2]$, with maximal rate β, Hill coefficient n, and half-maximal activation at a homodimer concentration of K_M. A low basal protein production rate, denoted α, is included to allow self-activation from low initial expression states. Finally, each protein can degrade and be diluted (due to cell division) at a total rate δ, regardless of its dimerization state. To simplify analysis, we non-dimensionalize the model by rescaling time in units of δ^{-1}, concentrations in units of K_M (*25*), and obtain:

$$\frac{d[A_{tot}]}{dt} = \alpha + \frac{\beta[A_2]^n}{1 + [A_2]^n} - [A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = \alpha + \frac{\beta[B_2]^n}{1 + [B_2]^n} - [B_{tot}]$$

Here, the Hill coefficient *n* only represents ultrasensitivity introduced by transcriptional activation. We include a more detailed discussion on additional ultrasensitivity provided by homodimerization and molecular titration in (*25*).

Since dimerization dynamics occur on a faster timescale than protein production and degradation (*52*), we assume that the distribution of monomer and dimer states remains close to their equilibrium values. This generates the following relationships between the concentrations of monomers, $[A]$and $[B]$, and dimers, $[A_2]$, $[B_2]$, and$[AB]$:

$$[A]^2 = K_d[A_2]$$

$$[B]^2 = K_d[B_2]$$

$$2[A][B] = K_d[AB]$$

Since the two transcription factors share the same dimerization domain, homo- and hetero-dimerization are assumed to occur with equal dissociation constants, K_d. Additionally, conservation of mass implies that $[A_{tot}] = [A] + [AB] + 2[A_2]$, with a similar relationship for B. Introducing the equilibrium equations given above into this conservation law produces expressions for the concentrations of the activating homodimers in terms of the total concentrations of A and B:

$$[A_2] = \frac{2[A_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d^2 + 8([A_{tot}] + [B_{tot}])K_d}}$$

$$[B_2] = \frac{2[B_{tot}]^2}{K_d + 4\,([A_{tot}] + [B_{tot}]) + \sqrt{K_d{}^2 + 8([A_{tot}] + [B_{tot}])\,K_d}}$$

Inserting these expressions into the differential equations for $[A_{tot}]$ and $[B_{tot}]$ above, we obtain a pair of coupled ordinary differential equations with only $[A_{tot}]$ and $[B_{tot}]$ as variables.

To understand the behavior of this system in physiologically reasonable parameter regimes (Table S2.1) (*25*), we used standard approaches from dynamical systems analysis (*53*). Based on ODEs, we first generated a phase portrait of variables $[A_{tot}]$ and $[B_{tot}]$ (labeled 'TF A' and 'TF B', which are dimensionless total TF A or B concentrations), where the linewidth of a vector (Fig. 2.1C, gray arrows) at any point is proportional to the speed of that point. On the phase portrait, we plotted the nullclines (Fig. 2.1C, solid lines), defined by setting each of the ODEs above to zero. We then identified fixed points at nullcline intersections, and determined their linear stability (Fig. 2.1C, black and white dots) (*53*). Finally, we delineated the basins of attraction for each stable fixed point (Fig. 2.1C, shaded regions).

Using this analysis, we identified parameter values that support type II tristability, a regime that minimally embodies the developmental concept of multilineage priming (*26–28*) (Figs. 2.1C and S2.1B). Stronger self-activation (higher values of β) was more likely to produce type II tristability (Fig. S1B, β row and column). Too much leaky production (high α) allowed both transcription factors to self-activate, reducing the degree of multistability, whereas too little (low α) stabilized the undesired OFF state (Fig.

S2.1B, α column). Strong dimerization (low K_d) was essential for type II tristability (Fig. S2.1B, K_d row and column). Finally, a broad range of Hill coefficients $n \geq 1$ were compatible with type II tristability. Although higher values of n led to a reduced sensitivity to other parameters and allowed the system to tolerate higher values of α, they also stabilized the OFF state (Fig. S2.1B, n row and column). Together, these results suggested that an ideal design would maximize β, minimize K_d, and use intermediate values of α and n.

Based on these conclusions, we incorporated multiple repeats of the homodimeric binding sites to maximize β, used strongly associating FKBP12F36V homodimerization domains (*38*) to minimize K_d, and modified the promoter sequences to allow some leaky expression to optimize α (Fig. S2.24) (*25*). Finally, although we did not directly control the Hill coefficient n, we expected that the repeated homodimeric binding sites should lead to some ultrasensitivity (*54*). These design choices produced the selected type II tristability in the experimental system (Fig. 2.3C).

A key feature of the MultiFate design is its ability to qualitatively change its multistability properties through bifurcations in response to parameter changes. In particular, the mathematical model predicts that protein stability can control the number of stable fixed points in phase space. In the non-dimensionalized model, the protein degradation rate, δ, does not appear explicitly but enters through the rescaling of α and β by $(\delta K_M)^{-1}$(*25*). Thus, tuning protein stability is equivalent to multiplying bothα and β by a common factor, which we term the "protein stability factor". Reducing protein

stability shifts the nullclines closer to the origin, causing the two unstable fixed points to collide with the stable A+B fixed point in a subcritical pitchfork bifurcation (Fig. 2.1C) (*53*). The result is a bistable system with A-only and B-only stable fixed points at somewhat lower concentrations (Fig. 2.1C). To experimentally realize this bifurcation, we designed the circuit to allow external control of transcription factor protein stability using the drug-inducible DHFR degron (Fig. 2.2C) (*39*). As predicted, reducing protein stability destabilized the A+B state, but preserved the A-only and B-only stable states (Fig. 2.3C). In this way, model-based design enabled us to rationally engineer tristability as well as externally controllable transitions to bistability in the experimental system.

2.8 Supplementary Figures

Fig. S2.1

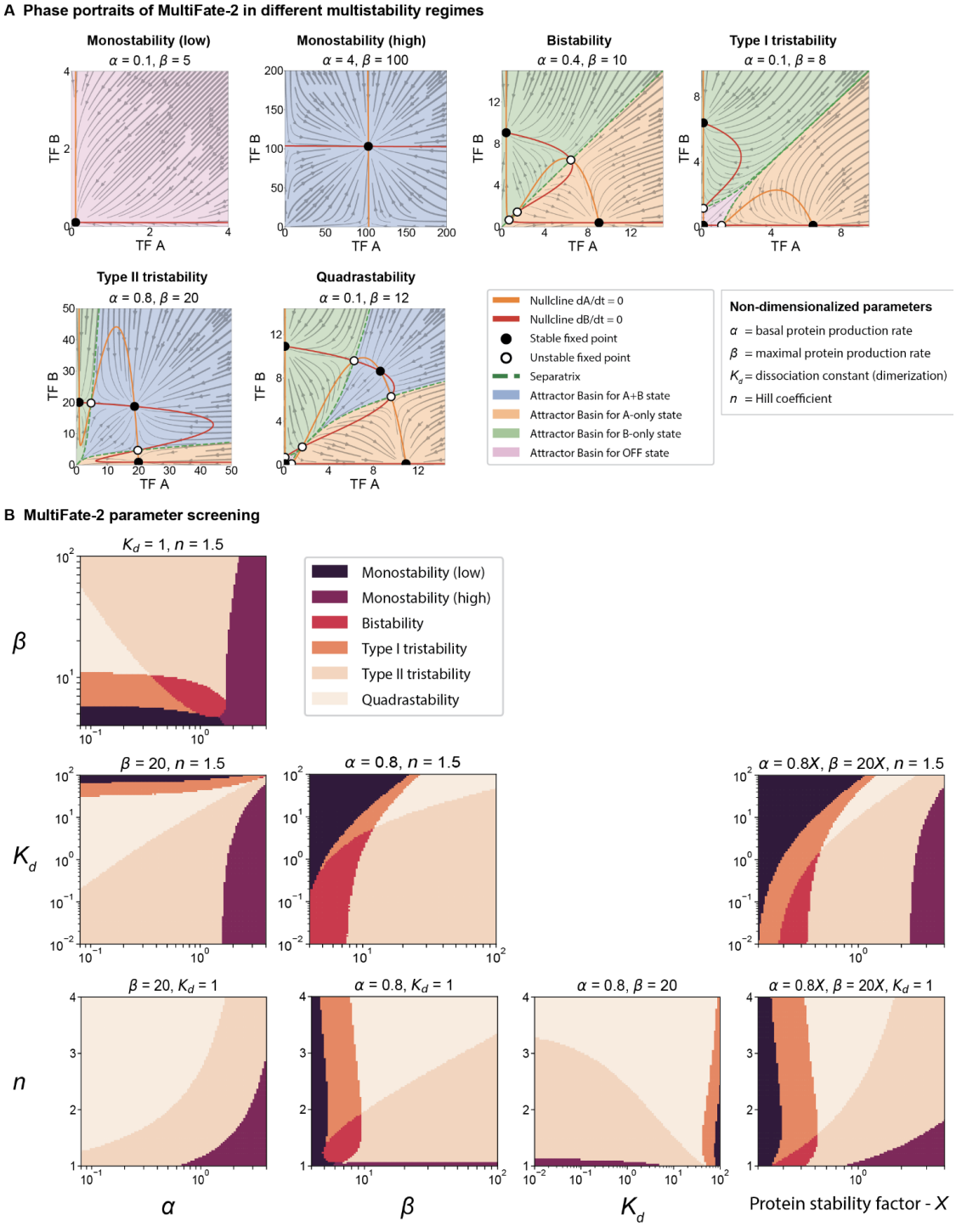

Fig. S2.1 MultiFate-2 model generates diverse types of multistability.
(A) In different symmetric parameter regimes (in which parameters for the two transcription factors are identical), MultiFate-2 can generate two types of monostability, bistability, two types of tristability, and quadrastability. For each regime, non-

dimensionalized parameters α and β are provided above the plot, and $K_d = 1$ and $n = 1.5$. The x and y axes are dimensionless TF A and TF B, respectively. (B) Parameter screen reveals how each of the non-dimensionalized parameters individually affects the global structure of the system. Each row and column in this grid of plots represents a titration of one parameter value, indicated at left and bottom. Within each plot, different colors represent different stability regimes, as in (A), determined by numerically solving for steady state values and their linear stability at each point in each parameter space (*25*). In the non-dimensionalized model, changing protein stability is equivalent to multiplying α and β with the same factor (Box 2.1), and is shown in the fourth column ("protein stability factor - X"), with higher values representing greater protein stability. Higher leaky transcription (high α) allows transcription factors to self-activate, destabilizing the OFF state (α column). Very high αvalues push the system towards monostability where only the state in which both transcription factors are highly expressed is stable. Stronger self-activation (higher values of β) is more likely to produce type II tristability and quadrastability (β column). Strong dimerization (low K_d) is essential for type II tristability (K_d row). A broad range of Hill coefficients $n \geq 1$ are compatible with different types of multistability (n row). While higher values of n reduce sensitivity to other parameters and allow the system generate type II tristability even with higher values of α, they also stabilize the OFF state to favor quadrastability.

Fig. S2.2

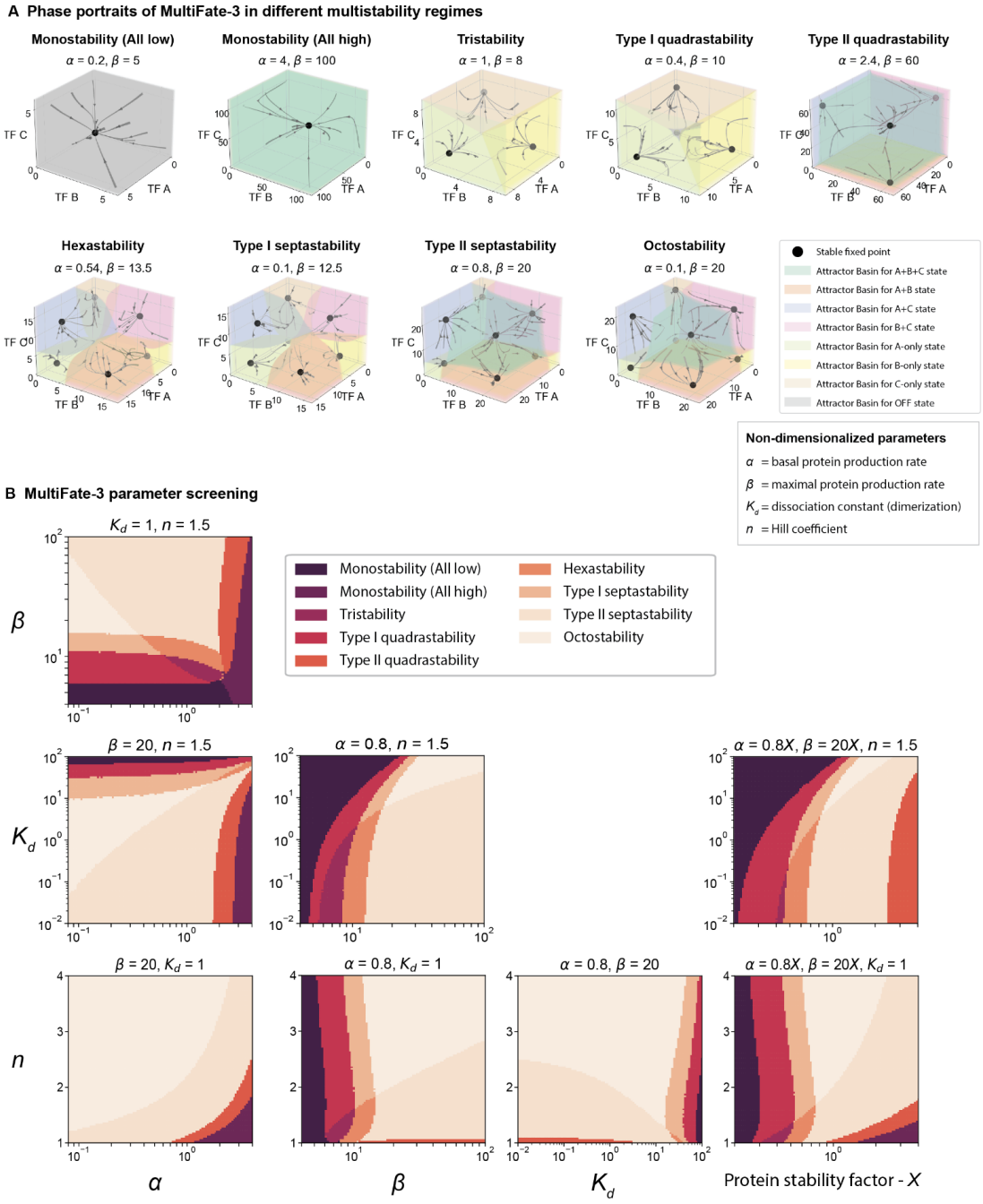

Fig. S2.2 MultiFate-3 model generates more diverse types of multistability.
(A) In different parameter regimes, symmetric MultiFate-3 can generate a diverse repertoire of different types of stability. The x, y and z axes are total dimensionless concentrations TF A, B and C, respectively. For each regime, non-dimensionalized parameters α and β are provided above the plot (Box 2.1), and $K_d = 1$ and $n = 1.5$. (B) Parameter screen reveals how each of the non-dimensionalized parameters individually affects the global structure of the system. As in Fig. S2.1, changing protein stability in the non-dimensionalized model is equivalent to multiplying α and β with protein stability

factor (X). Each plot in the grid shows a titration of two parameter values (left and bottom). Higher leaky transcription (high α) allows transcription factors to self-activate, destabilizing the OFF state. Very high α pushes the system towards monostability where only the state which all transcription factors are highly expressed is stable (α column). Stronger self-activation (higher values of β) generally favors higher levels of multistability, such as septastability and octostability (β column). Strong dimerization (low K_d) is essential for type II septastability (K_d column). A broad range of Hill coefficients $n \geq 1$ are compatible with different types of multistability. Higher values of n reduce sensitivity to other parameters, and allow the system to generate type II septastability even at higher values of α. However, they also stabilize the OFF state to favor octostability (n column). Note that the MultiFate-3 parameter screen graph structure resembles that of MultiFate-2 (Fig. S2.1B), in which monostability, octostability and type II septastability of MultiFate-3 appear at similar positions as monostability, quadrastability and type II tristability of MultiFate-2. As in Fig. S2.1B, each plot is calculated by numerical solution of the MultiFate-3 model for steady state values and their linear stability (*25*).

Fig. S2.3

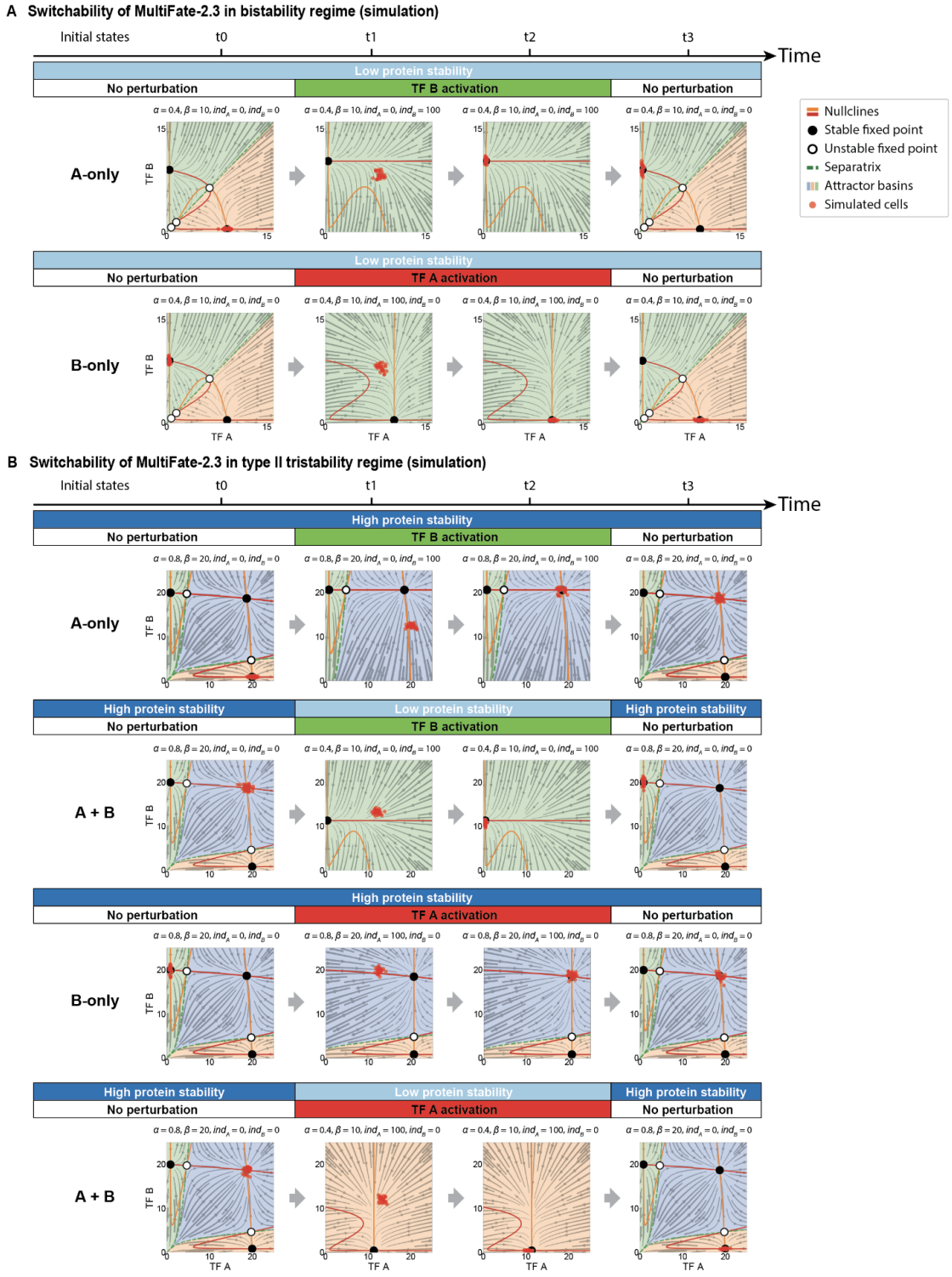

Fig. S2.3 Modeling state-switching dynamics.
We used Gillespie simulations to simulate the effects of transient perturbations on switching of cells among different states in both bistable regime (A) and type II tristable regime (B). We used a modified MultiFate-2 model incorporating external inducers

shown in (*25*). In this model, the strength of inducers is represented by the parameter ind_A or ind_B. Initially (left plot), cells (red dots) are in A-only, B-only or A+B state. Increasing ind_A or ind_B destabilizes the initial state (central two plots), allowing cells to transition to the target state. Terminating inducer treatment restores $ind_A = ind_B = 0$. However, at this point, cells are already stabilized in the target state. The α, β, K_d, n used for bistable regime and type II tristable regime are the same as Fig. 2.1C.

Fig. S2.4

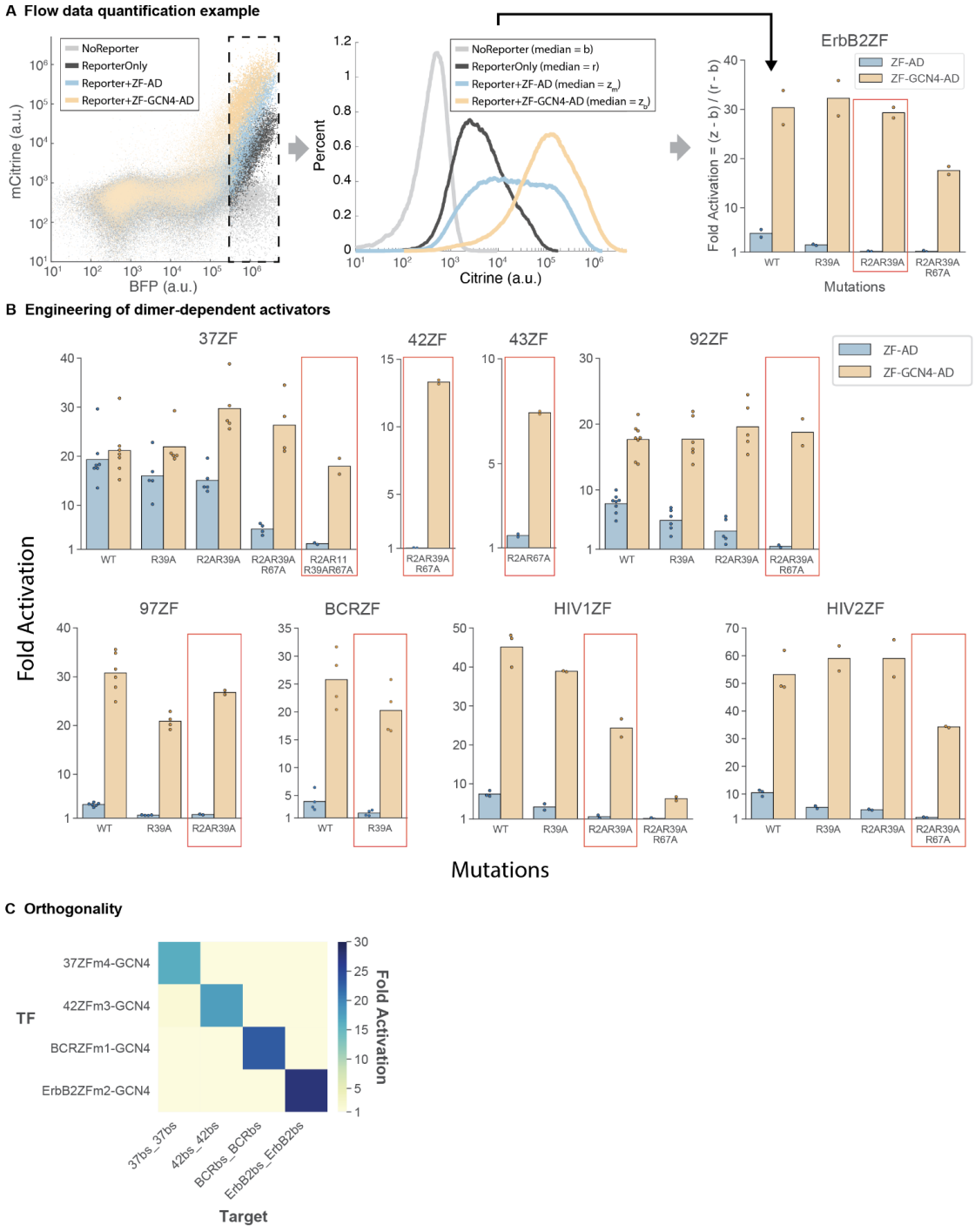

Fig. S2.4 Engineering dimer-dependent transcriptional regulation.
(A) To characterize the activation strength of different zinc finger transcription factor variants, we co-transfected each of them with a reporter construct (Citrine) and a co-transfection marker (mTagBFP2). Highly transfected cells were gated with co-transfected BFP > $3x10^5$ (dashed box) to extract their individual histograms (middle). From each histogram, median Citrine fluorescence intensities of gated cells were used to calculate fold activation (right) (*25*). Black arrow indicates the original zinc finger

sequence from the middle panel (WT). The R2AR39A variant (red box) was selected for its high ZF-GCN4-AD activation and minimal ZF-AD activation (AD denotes the VP48 activation domain). (B) Zinc finger mutation variants with minimal activation by ZF-AD and strong activation by ZF-GCN4-AD (red boxes) were selected for use in MultiFate circuits. Each dot represents one biological replicate, and each bar indicates the mean of all replicates. 37ZF, 42ZF, 43ZF, 92ZF, 97ZF were taken from (*32*). These zinc fingers were previously (*32*) denoted, respectively, as 37-12 array, 42-10 array, 43-8 array, 92-1 array, and 97-4 array. BCRZF, HIV1ZF, HIV2ZF, and ErbB2ZF are from (*29*), and BCR denotes the BCR_ABL domain. (C) Four selected zinc finger transcription factors used later in the paper exhibit orthogonal trans-activation. Each row represents a transcription factor with abbreviated labels for figure layout. Full transcription factor descriptions are, from top to bottom, 37ZFR2AR11AR39AR67A-GCN4-VP48; 42ZFR2AR39AR67A-GCN4-VP48; BCRZFR39A-GCN4-VP48; and ErbB2ZFR2AR39A-GCN4-VP48. Each target (column) is the same Citrine fluorescent reporter used in panel (A), with 2 repeats of 18bp tandem binding site pairs, denoted “ZFbs_ZFbs” for each type, at the promoter. Each square in the matrix is the mean of two biological replicates.

Fig. S2.5

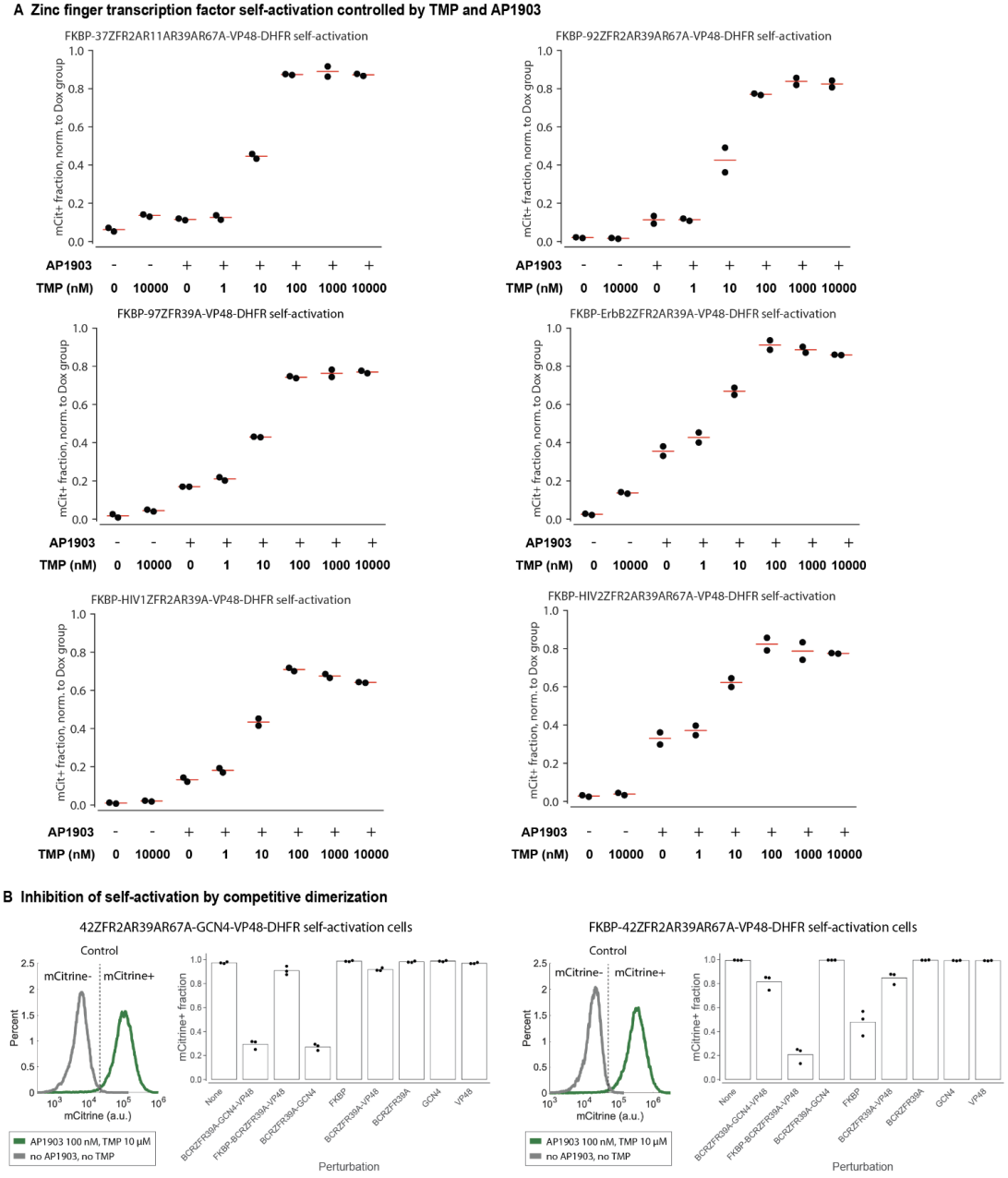

Fig. S2.5 Engineered dimer-dependent transcription factors enable transcriptional positive autoregulation and mutual inhibition through competitive dimerization.
(A) Positive autoregulation can be controlled by TMP and AP1903. Here, normalized mCitrine+ fractions are used to quantify self-activation strength, as in Fig. 2.2C (see (*25*) for quantification methods) for 6 additional ZFs. (B) Positive autoregulation was inhibited by competing proteins with matching dimerization domains. Two monoclonal stable lines (plot subtitles) could spontaneously self-activate in media containing 100 nM AP1903 and 10 µM TMP (histograms). Potentially competing proteins were expressed from plasmids and stably integrated into each monoclonal line (*25*). Note that in 42ZFR2AR39AR67A-GCN4-VP48-DFHR self-activation cells (left), the GCN4 domain

by itself did not have inhibitory effects, but could efficiently inhibit self-activation when fused with BCRZFR39A. In FKBP-42ZFR2AR39AR67A-VP48-DHFR self-activation cells (right), the FKBP domain by itself can partially inhibit self-activation. The mCitrine threshold is $2x10^4$ for 42ZFR2AR39AR67A-GCN4-VP48-DHFR self-activating cells, and $5x10^4$ for FKBP-42ZFR2AR39AR67A-VP48-DHFR cells. The integrated plasmid in the "None" group did not express any transcription factor variants. In all panels, each dot represents one biological replicate, and each red line or bar indicates the mean of replicates.

Fig. S2.6

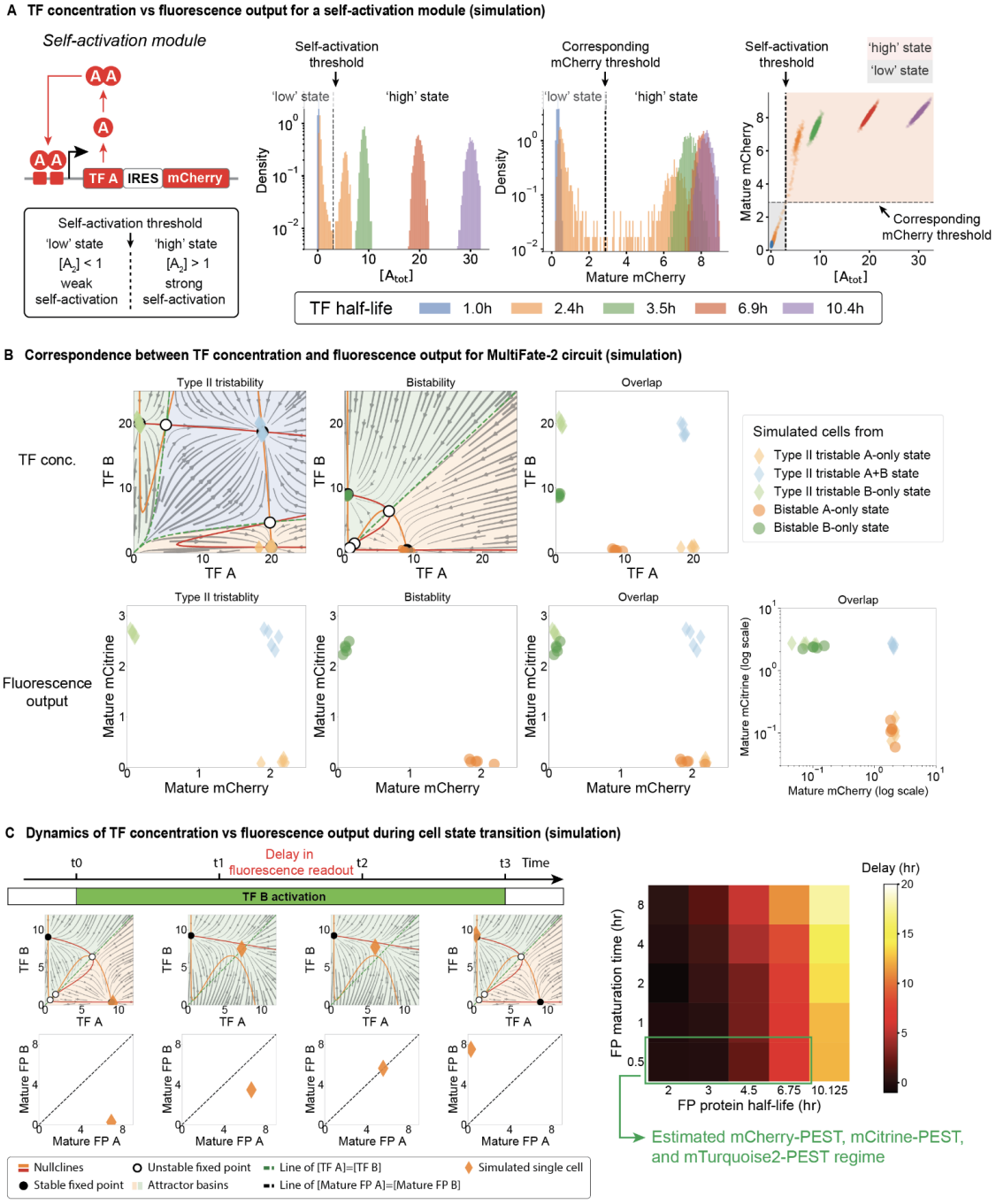

Fig. S2.6 Modeling the relationship between TF concentrations and fluorescence readout.

(A) Self-activation module creates a threshold-like behavior: when transcription factor concentration is higher than the threshold, the module would be highly active and express a high level of fluorescent proteins, resulting in a 'high' state; transcription factor concentration is lower than the threshold, the module is inactive and express a minimal level of fluorescent proteins, resulting in a 'low' state. (Left) The threshold is the TF concentration that produces a homodimer concentration of 1. (Middle) In the 'high' state,

transcription factor concentrations are sensitive to protein half-life, while fluorescence readouts (right) are not sensitive and almost overlap with each other. (B) Modeling shows that MultiFate-2 fluorescence readouts are well separated into distinct clusters, and each cluster can be unambiguously assigned to its corresponding state defined by transcription factor concentrations. (Top) Transcription factor concentrations of simulated cells cluster around stable fixed points. For A-only state or B-only state, transcription factor concentrations differ by more than 2 folds between type II tristable regime and bistable regime, as shown on the 'Overlap' plot. (Bottom) By contrast, fluorescence readouts for A-only state or B-only state almost overlap with each other between type II tristable regime and bistable regime, consistent with experimental observation in Fig. S2.9-2.11. In this panel, the MultiFate-2 parameters related to transcription factor dynamics (α, β, K_d, n) are the same as those used in Fig. 2.1C. Parameters related to fluorescent protein dynamics are in Table S2.1. (C) Fluorescence readouts have a time delay compared with transcription factor concentrations during cell state transition. We used the same modified MultiFate-2 model incorporating external input as that used in Fig. S3 to simulate transcription factor dynamics, with $\alpha = 0.4$, $\beta = 10$, $K_d = 1$, $n = 1.5$. (Left) Simulated single cell dynamics of transcription factor concentrations (top) and fluorescence readouts (bottom) during cell state transition from A-only state to B-only state. Selected timepoints are: t_1 is when the cell is in A-only state at the start of simulation; t_2 is when transcription factor concentrations cross the state boundary (i.e. [TF A] = [TF B]); t_3 is when fluorescence readouts cross the state boundary (i.e. [Mature FP A] = [Mature FP B]); t_4 is when the cell is in B-only state at the end of simulation. For this simulation, maturation time of both fluorescence proteins is 8 hours, and fluorescent protein half-life is 10.125 hours. (Right) Both longer maturation time and longer fluorescent protein half-life increase the time delay of fluorescence readouts. The color of each block represents the mean delay time of 200 cells.

Fig. S2.7

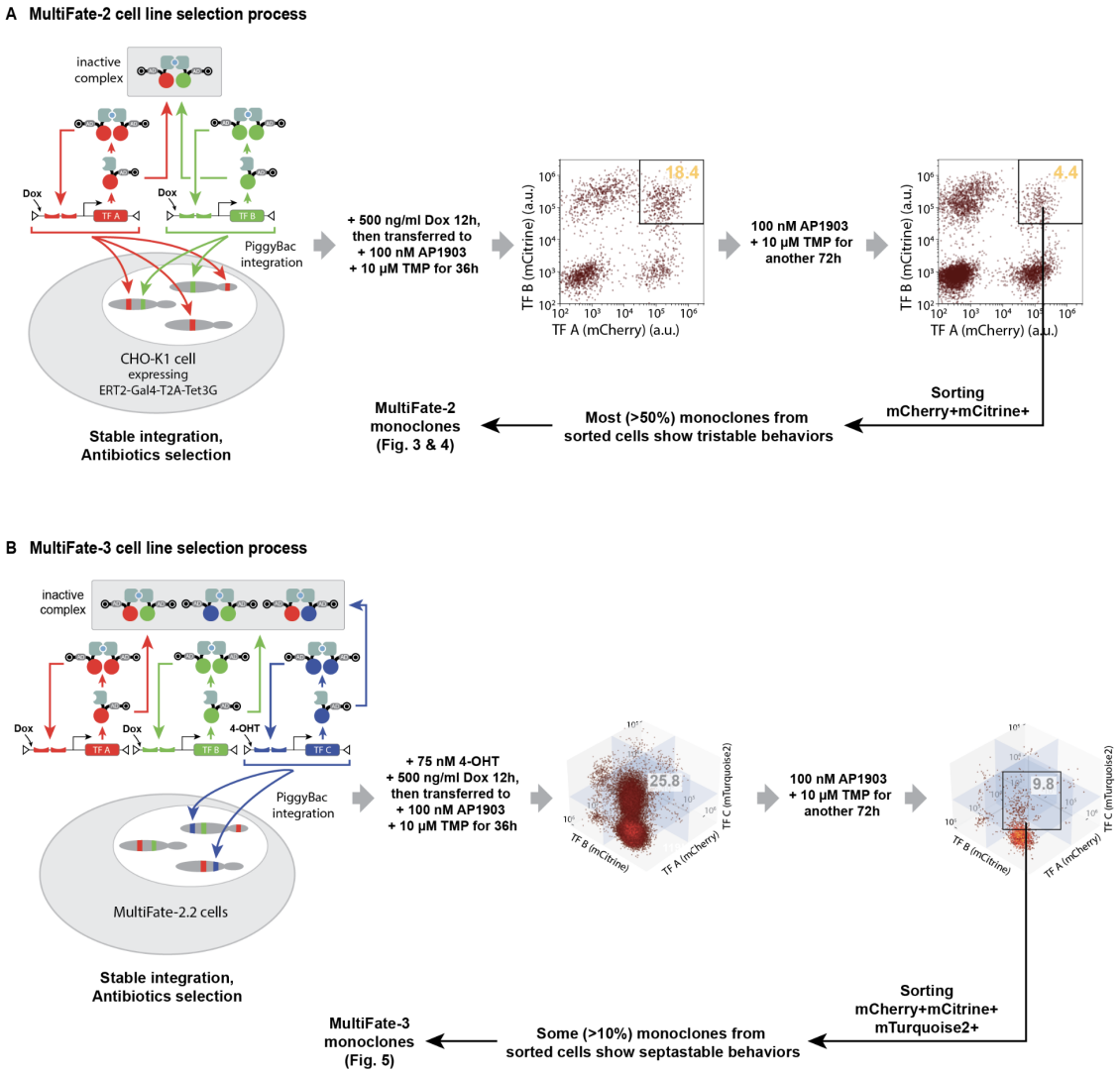

Fig. S2.7 Schematics of MultiFate-2 and MultiFate-3 clone selection processes.
(A) MultiFate-2 monoclones were selected from a population of cells that can maintain a stable double-positive state for at least 72 hours. (B) MultiFate-3 monoclones were selected from a population of cells that can maintain a stable triple-positive state for at least 72 hours. Also see (*25*) for a description of clone selection processes.

Fig. S2.8

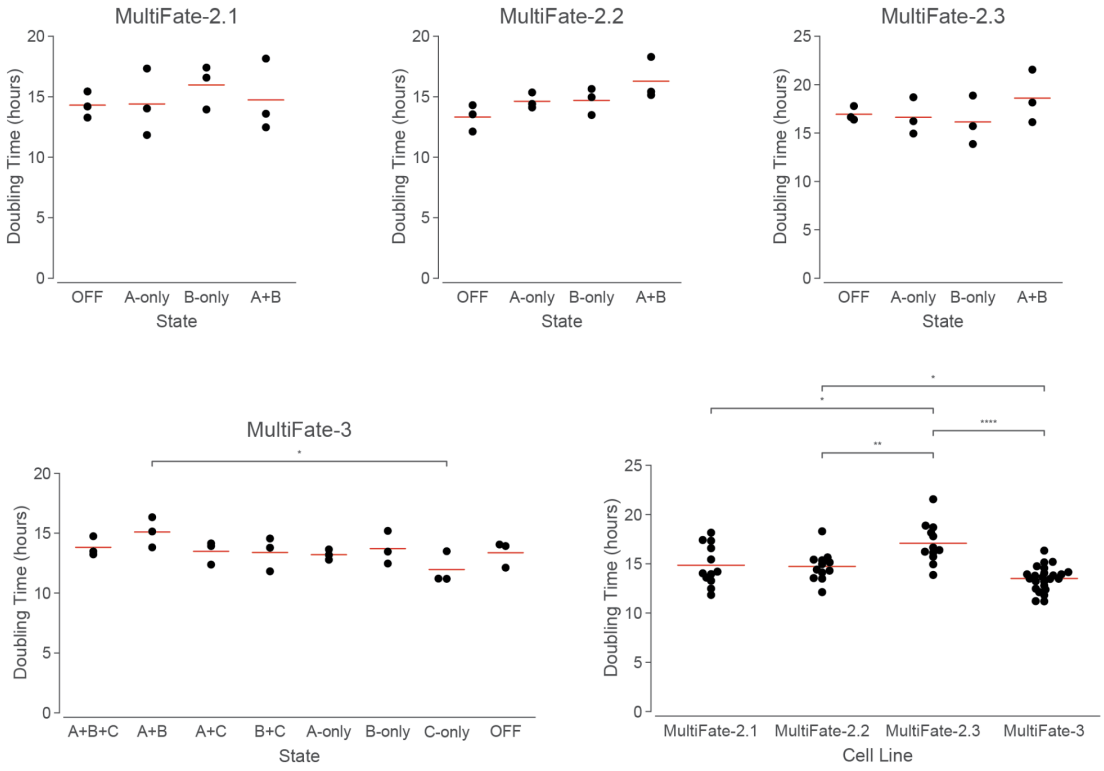

Fig. S2.8 Doubling time of MultiFate cells.
For each MultiFate line, most differences in doubling time among cells in different states are not significant. We used Welch's t-test (threshold p = 0.05) since it is suitable for pairwise comparison without assuming equal variance. Only significant difference is that MultiFate-3 cells in A+B state grow slower than cells in C-only state. The differences in doubling time among different MultiFate lines are significant, probably due to clonal differences. Each dot represents one biological replicate, and each red line indicates the mean of all replicates. Doubling time of the same cell line from different states (from first four plots) are combined to generate the last plot. *: 1e-2 < p ≤ 5e-2; **: 1e-3 < p ≤ 1e-2; ***: 1e-4 < p ≤ 1e-3; ****: p ≤ 1e-4.

Fig. S2.9

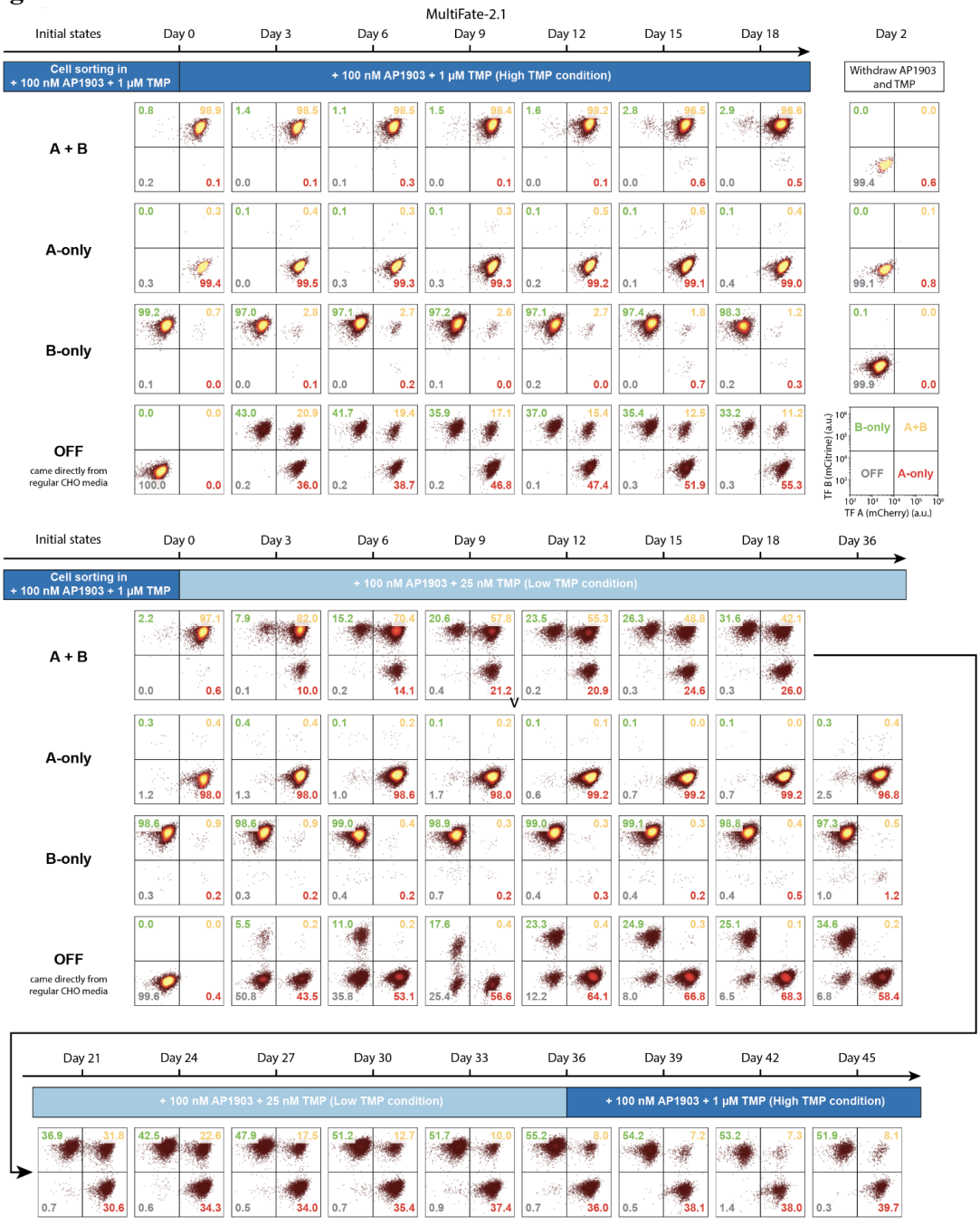

Fig. S2.9 Raw flow cytometry analysis of the MultiFate-2.1 line.
Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.3C, MultiFate-2.1 columns, and Fig. 2.4A). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top, and initial OFF cells came directly from cells in regular CHO media without any inducers. Each 2-dimensional flow cytometry plot was divided at mCherry = 10^4 and mCitrine = $2x10^4$ into four quadrants, representing four states. For each plot, the percentage of cells in each of the four states is labeled on the corresponding corner. Timelines above each set of plots represent the indicated inducer conditions.

Fig. S2.10

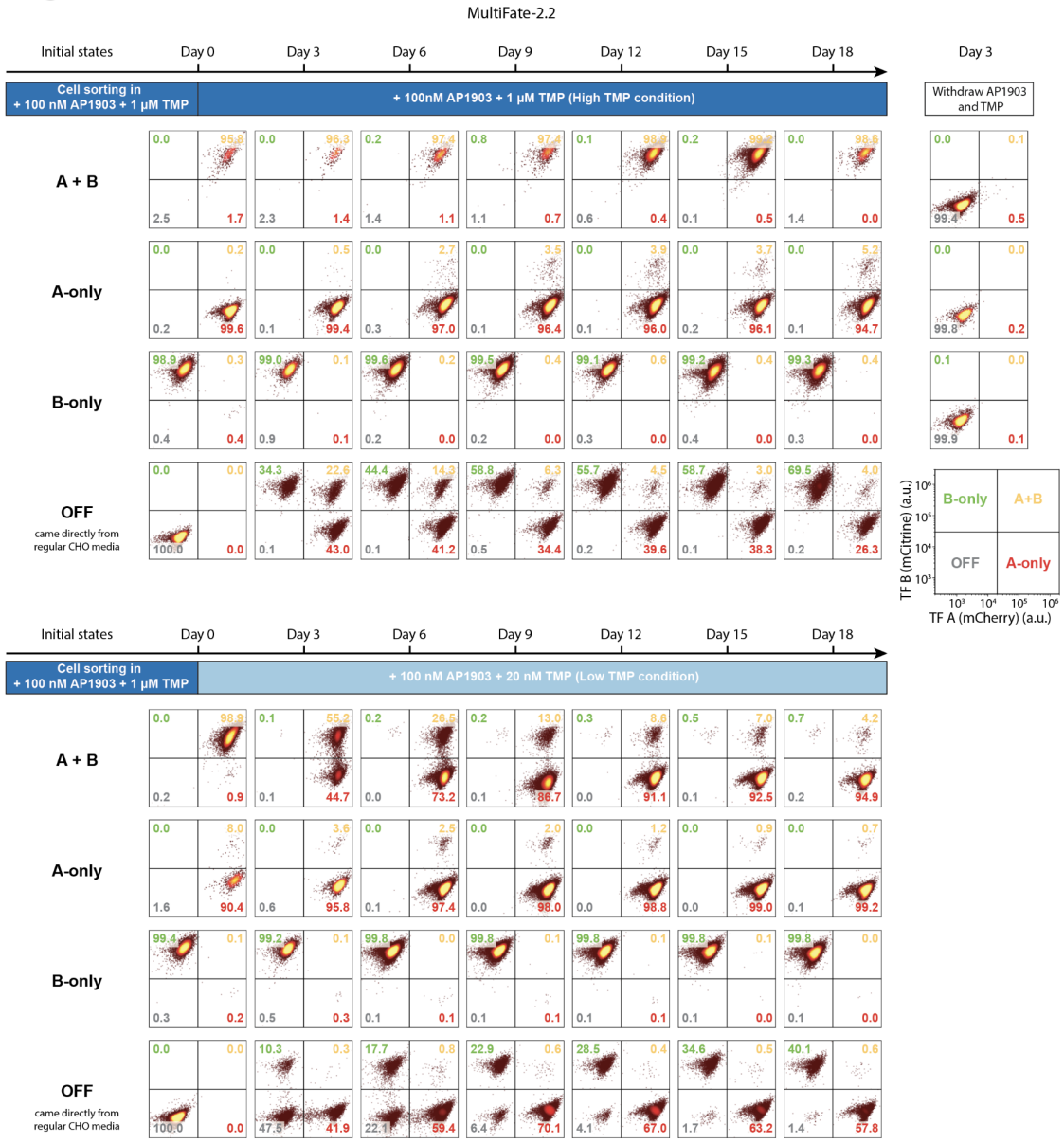

Fig. S2.10 Raw flow cytometry analysis of the MultiFate-2.2 line.
Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.3C, MultiFate-2.2 columns). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top, and initial OFF cells came directly from cells in regular CHO media without any inducers. Each 2-dimensional flow cytometry plot was divided at mCherry = $2x10^4$ and mCitrine = $3x10^4$ into four quadrants, representing four states. For each plot, the percentage of cells in each of the four states is labeled on the corresponding corner. Timelines above each set of plots represent the indicated inducer conditions.

Fig. S2.11

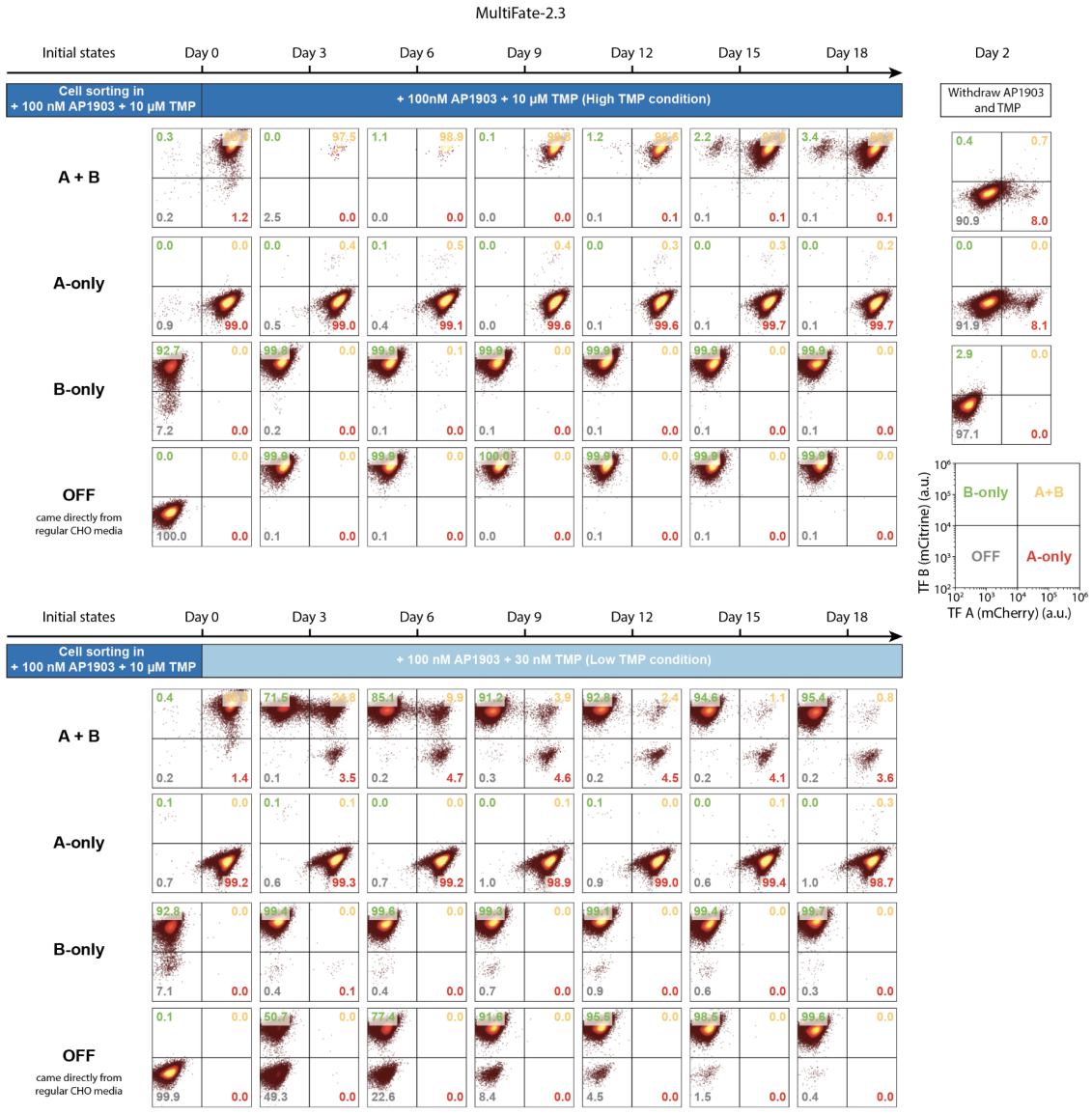

Fig. S2.11 Raw flow cytometry analysis of the MultiFate-2.3 line.
Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.3C, MultiFate-2.3 columns). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top, and initial OFF cells came directly from cells in regular CHO media without any inducers. Each 2-dimensional flow cytometry plot was divided at mCherry = 10^4 and mCitrine = 10^4 into four quadrants, representing four states. For each plot, the percentage of cells in each of the four states is labeled on the corresponding corner. Timelines above each set of plots represent the indicated inducer conditions.

Fig. S2.12

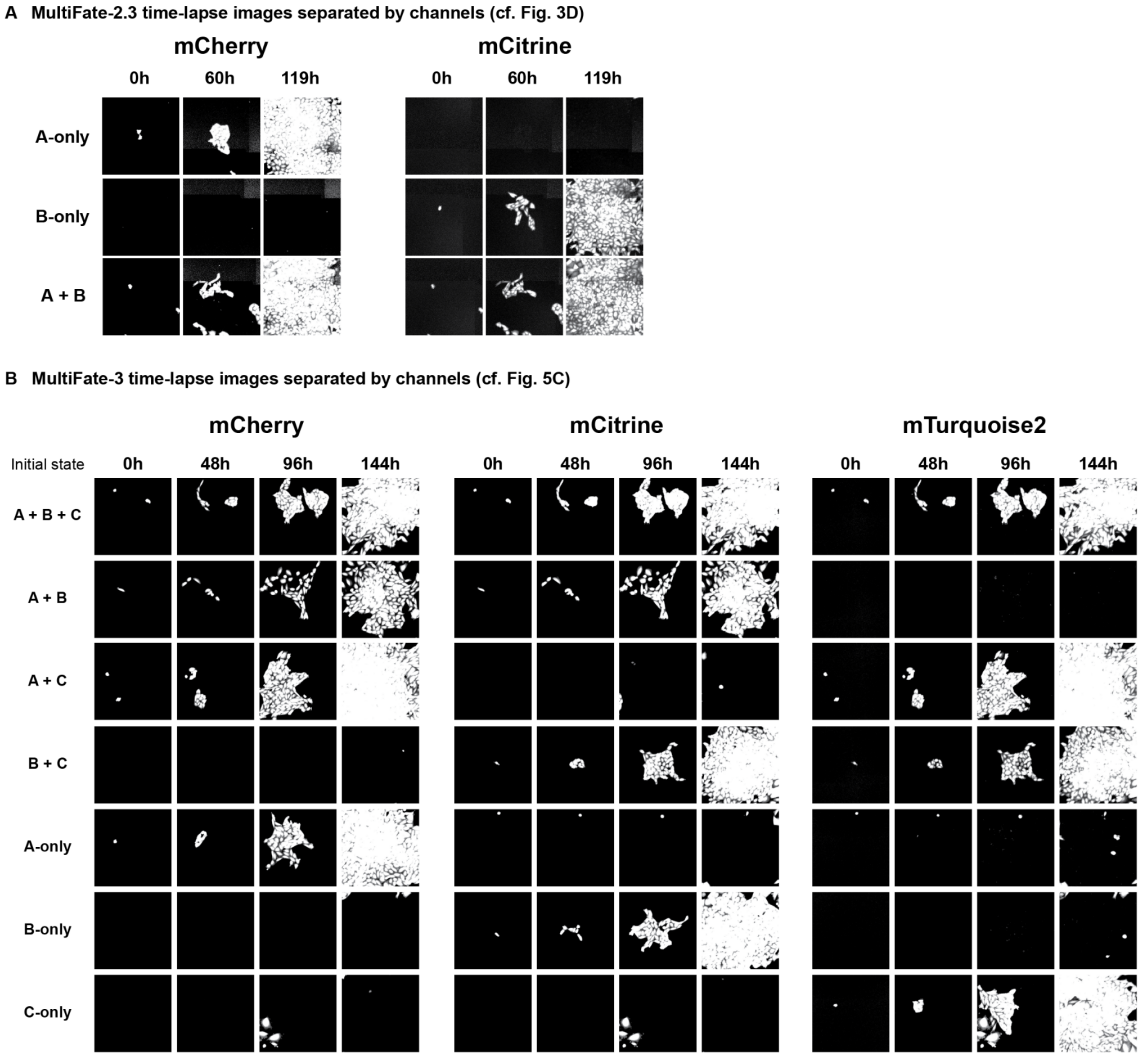

Fig. S2.12 Raw time-lapse images separated by channels.
Representative time-lapse images separated by channels. The brightness and contrast for images in the same movie were adjusted to be the same. (A) For MultiFate-2.3 time-lapse images, the intensity range of mCherry channel and mCitrine channel are [550, 800] and [300, 800], respectively. (B) For MultiFate-3 time-lapse images, the intensity range of mCherry channel, mCitrine channel and mTurquoise2 channel are [600, 1400], [350, 800] and [1000, 1500], respectively. Raw image files are available at data.caltech.edu/records/1882.

Fig. S2.13

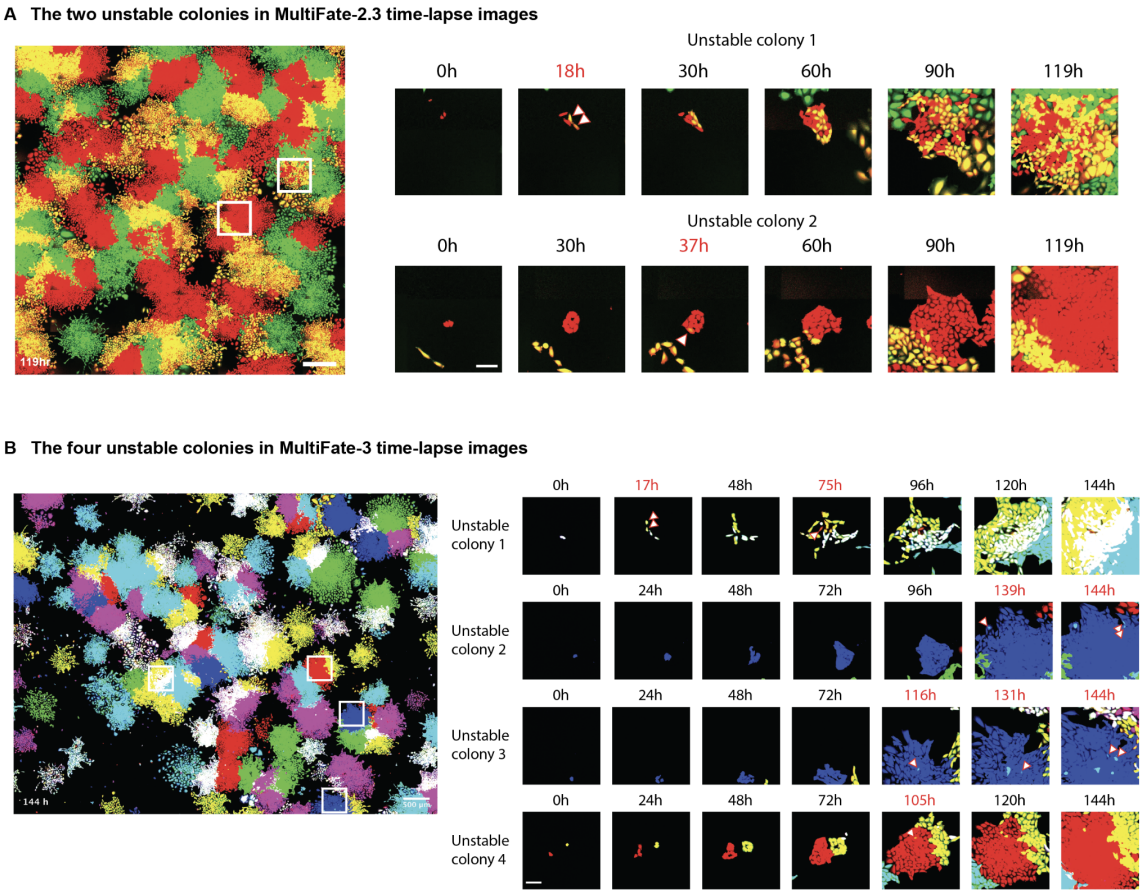

Fig. S2.13 Time-lapse movies allow direct visualization of rare spontaneous state-switching events.

(A) Two colonies (white boxes) in a MultiFate-2.3 time-lapse movie exhibited spontaneous state-switching events. Each example is shown as a filmstrip to the right. Arrowheads indicate the cell that switched states. In example 1 (top), a pair of A+B cells (yellow) appear from a colony started in the A-only (red) state (arrowheads). A similar transition was also identified in example 2 (bottom, white arrowhead). (B) We identified four state-switching events (highlighted in white rectangles) in a MultiFate-3 time-lapse movie. Filmstrips (right) show the events. White arrowheads indicate cells that have switched states. Example 1 (top row) shows a transition from A+B+C (white) to A+B (yellow) and another transition to A-only state (red). Examples 2 and 3 (second row) show a transition from C-only (blue) to B+C (cyan). Example 4 shows a transition from A-only (red) to A+B (yellow). Time points of spontaneous state-switching are labeled in red type. Scale bar is 500 μm for the wide field image (left), and 100 μm for zoomed in images (right).

Fig. S2.14

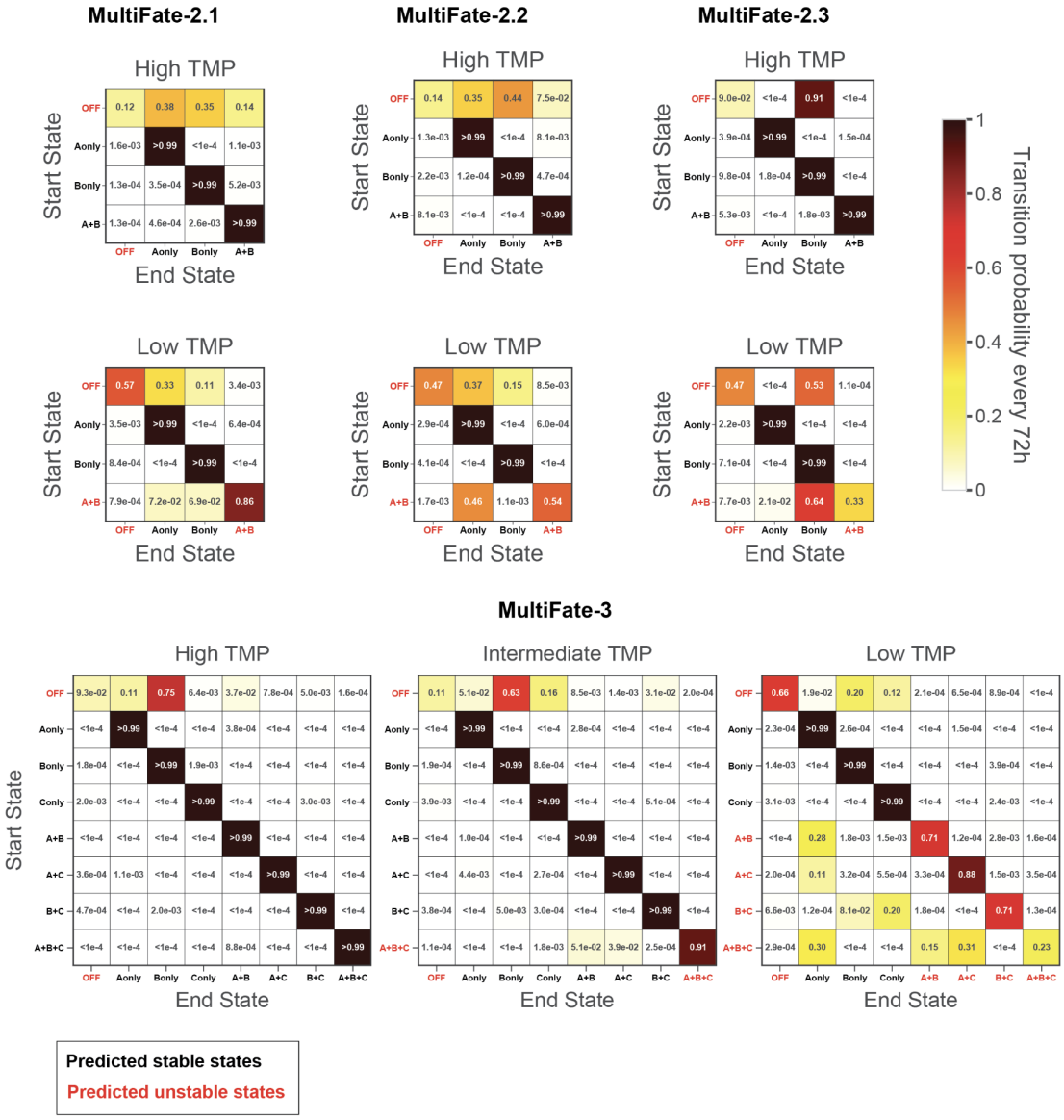

Fig. S2.14 MultiFate state transition matrix shows low transition rates out of stable states, and distinct transition preferences out of unstable states for different cell lines.

These transition matrices show that each predicted stable state (black bold text) has a very low level of cell transition out of the state (<1% every 3 days) as shown by weak off-diagonal blocks. The preferences of cells in each predicted unstable state (red bold text) to transition to different stable states are shown by the intensity of off diagonal blocks, and are consistent with results in Figs. 2.3C and 2.5B. We calculated the transition matrix among MultiFate states for different MultiFate lines from multi-day flow cytometry data (cf. Figs. S2.9 to S2.11 and other two biological replicates) using a method developed in (*55*).

Fig. S2.15

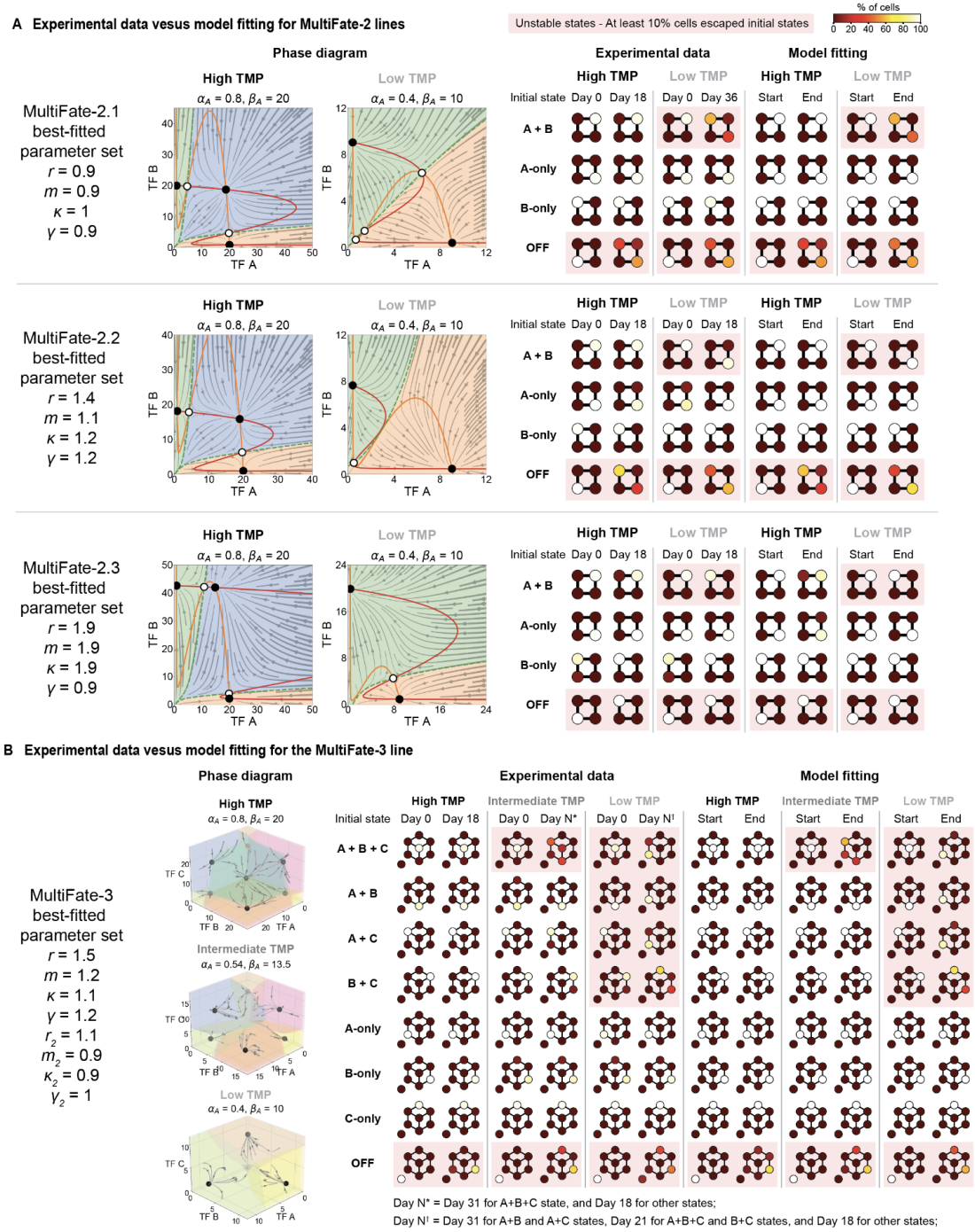

Fig. S2.15 Simulated cell fractions from best-fitted asymmetry parameter sets recapitulate experimental cell fractions of different MultiFate-2 lines in various conditions.

Using stochastic asymmetry MultiFate models, we obtained the best-fitted parameter set to recapitulate the experimental data of each MultiFate cell line (*25*). For both panels, each pair of start-end plots in 'Model fitting' columns were generated by stochastic

simulations of 400 cells starting from each initial state in each condition. The asymmetry parameters for each MultiFate line were shown on the left, and the symmetry parameters were shown on the phase diagrams. $K_d = 1$ and $n = 1.5$. Note that the 3D phase diagrams in panel B were tilted slightly differently compared with those in Figs. 2.1D and S2.2A, for better visualization of asymmetry.

Fig. S2.16

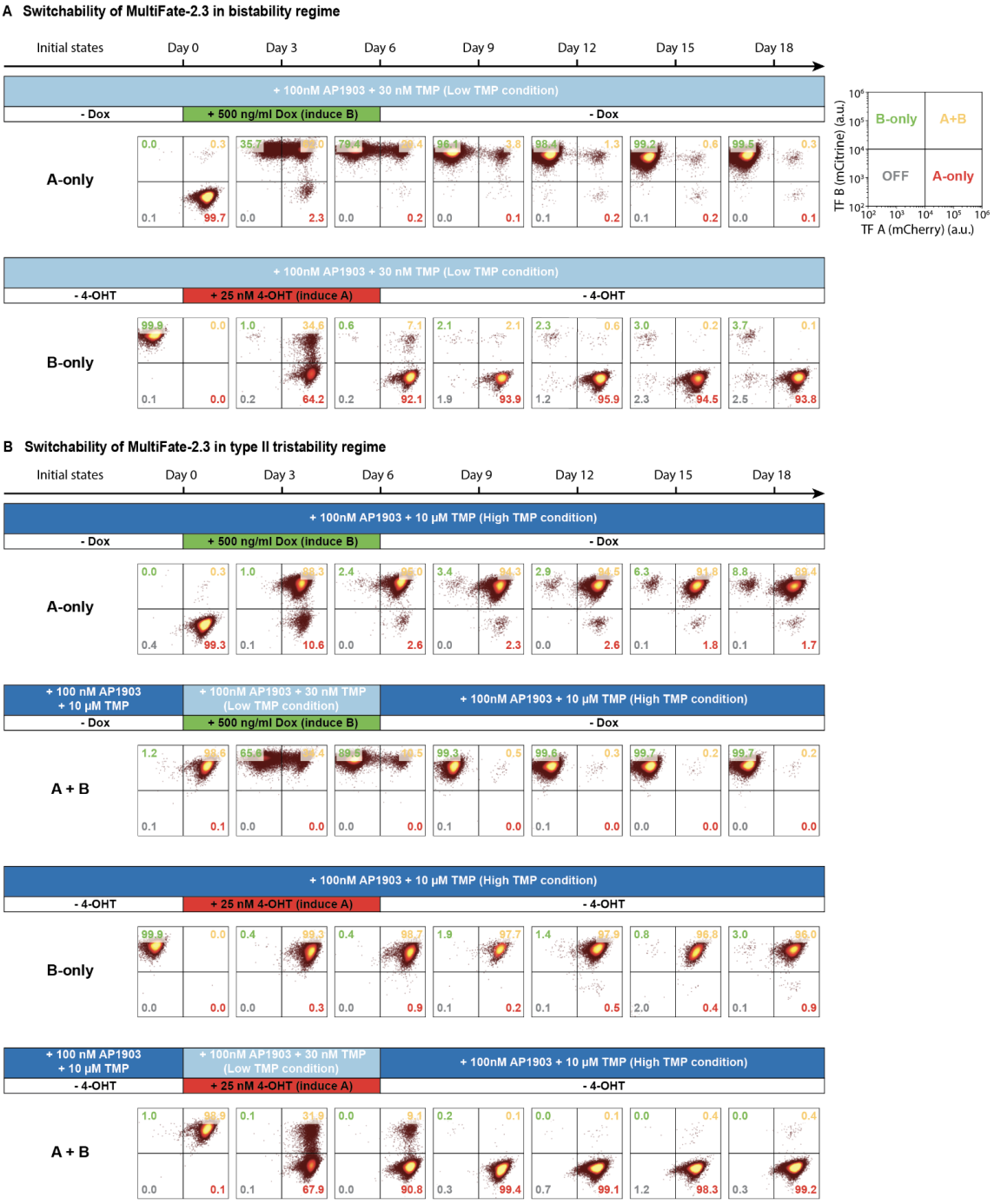

Fig. S2.16. Raw flow cytometry analysis of MultiFate-2.3 state-switching.
Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.4B). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top. Each 2-dimensional flow cytometry plot was divided at mCherry = 10^4 and mCitrine = 10^4 into four quadrants, representing four states. For each plot, the percentage of cells in each of the four states is labeled on the corresponding corner. Timelines above each set of plots represent the time and indicated inducer

conditions. Note that since the response elements for 4-OHT or Dox are adjacent to the transcription factor homodimer binding sites (Table S2.2), the addition of 4-OHT or Dox increases A or B expression up to, but not substantially beyond, the level produced by transcription factor self-activation. For example, in panel B third row, transcription factor A self-activation resulted in a 125-fold increase in expression (fold increase is calculated by dividing mCherry median in Day 18 by mCherry median in Day 0, mean of three replicates). Compared with transcription factor self-activation (Day 18), additional 4-OHT activation resulted in only another 2.8-fold increase in A expression (Day 6 versus Day 18). Similarly, in panel B first row, transcription factor B self-activation resulted in a 48-fold increase in expression (Day 18 versus Day 0). Additional Dox activation further increased the B expression only by another 1.4 fold (Day 6 versus Day 18).

Fig. S2.17

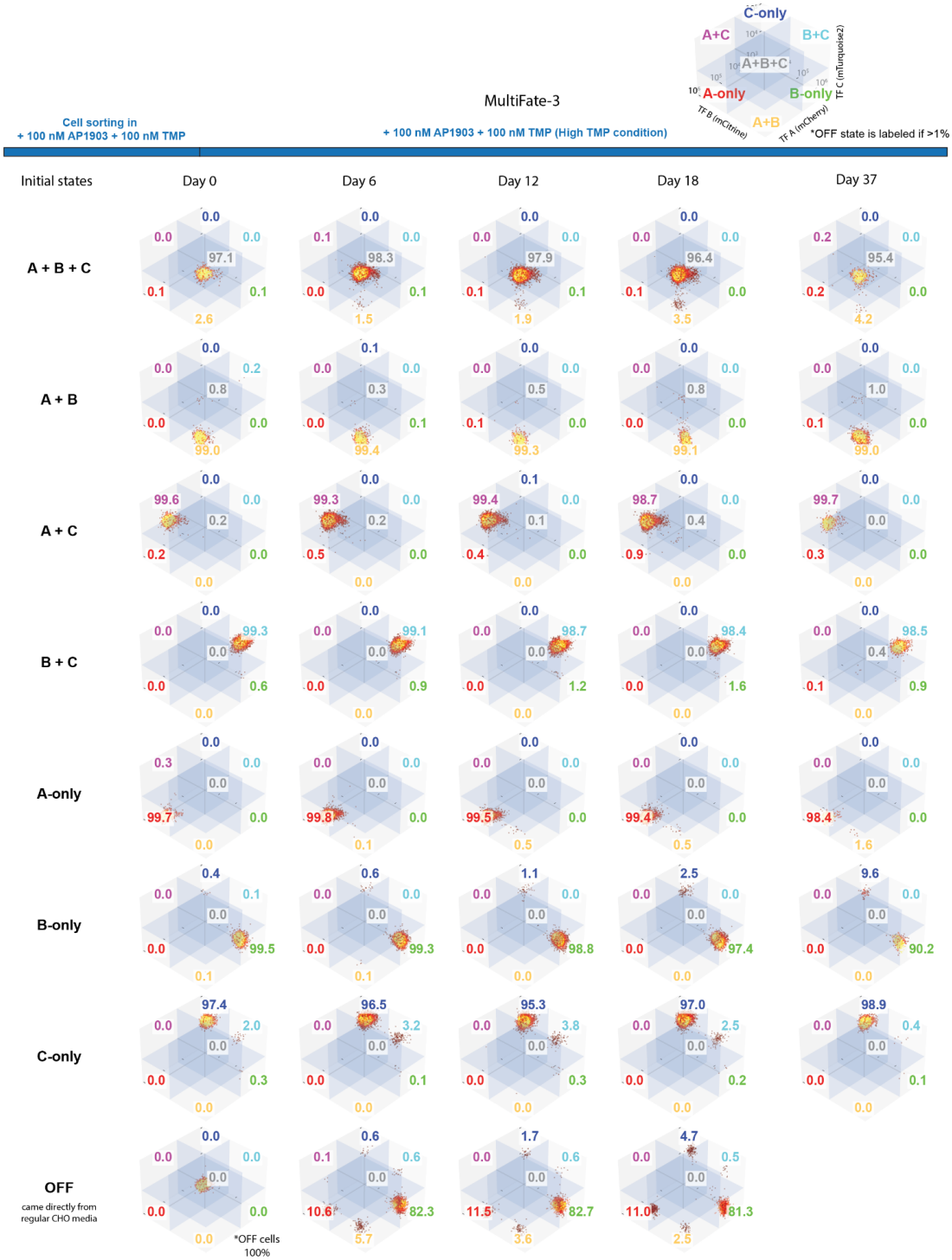

Fig. S2.17 Raw flow cytometry data of MultiFate-3 line under the High TMP condition.

Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.5B, High TMP). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top, and initial OFF cells came directly from cells in

regular CHO media without any inducers. Each 3-dimensional flow cytometry plot was divided at mCherry = $2x10^4$, mCitrine = $4x10^4$ and mTurquoise2 = $9x10^3$ into eight octants, representing eight states. For each plot, the percentage of cells in each of the 7 states (excluding the OFF state) is labeled on the corresponding octant, as shown in legend at top-right. The timeline (top) represents the indicated inducer conditions. OFF state percentages are usually very low (<1%) across all conditions, and are separately labeled if the percentage is greater than 1%. One of three replicates of cells from each of the 7 initial states (excluding the OFF state) were continuously cultured beyond 18 days. In all 7 states, >90% of cells remained in their original state at day 37 (see indicated percentages).

Fig. S2.18

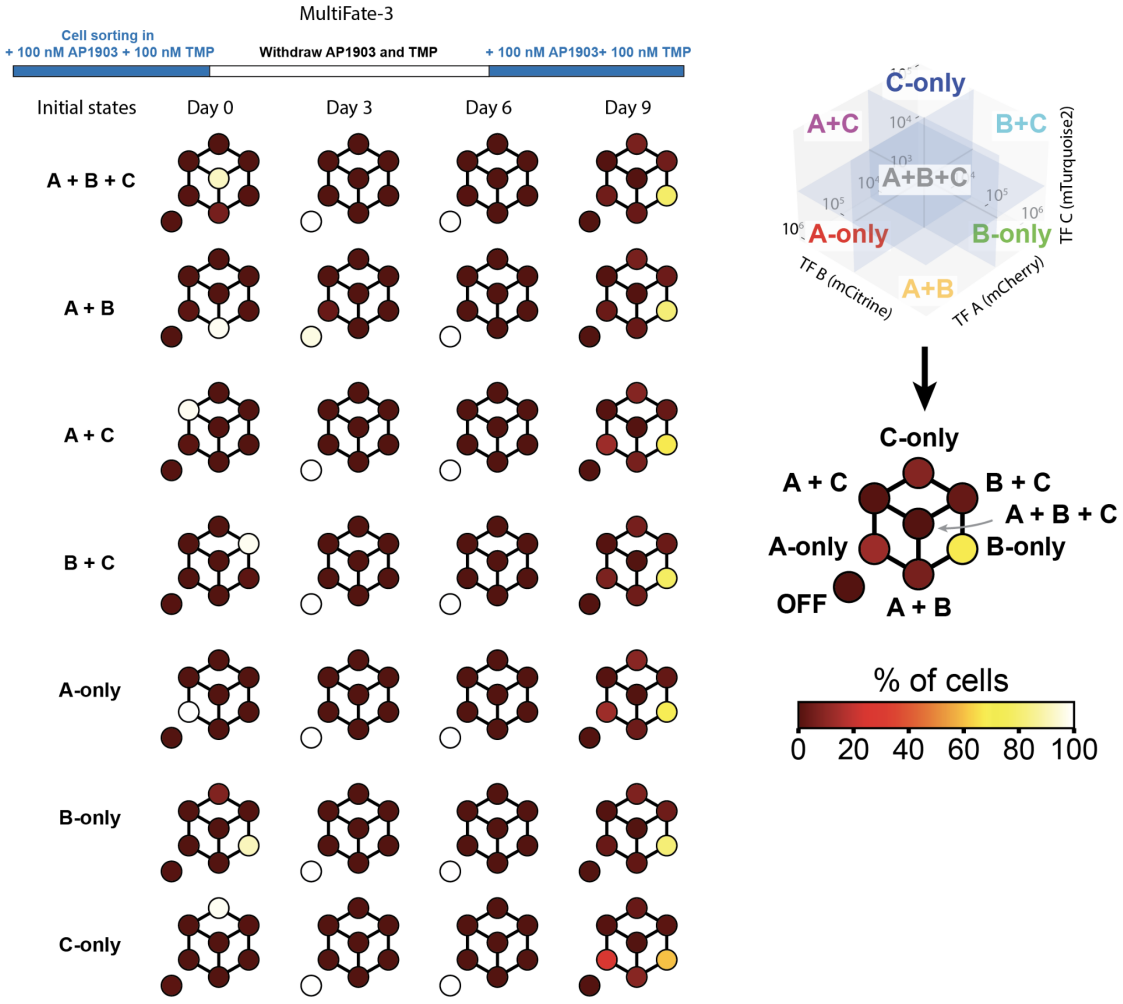

Fig. S2.18 Inducer withdrawal and reintroduction experiments showed MultiFate dependency on positive autoregulation and ruled out the possibility of mixed clones. Sorted cells in seven different states were transferred from AP1903+TMP media into regular media without any inducers. Most cells returned to the OFF state within 3 days (second column). After 6 days, AP1903+TMP was added back to the media, and cells were measured by flow cytometry after another 3 days. The resulting state distributions (fourth column) were similar to each other, suggesting that sorted cells in seven different states come from the same monoclonal MultiFate-3 line. Each plot represents the mean fractions of three biological replicates.

Fig. S2.19

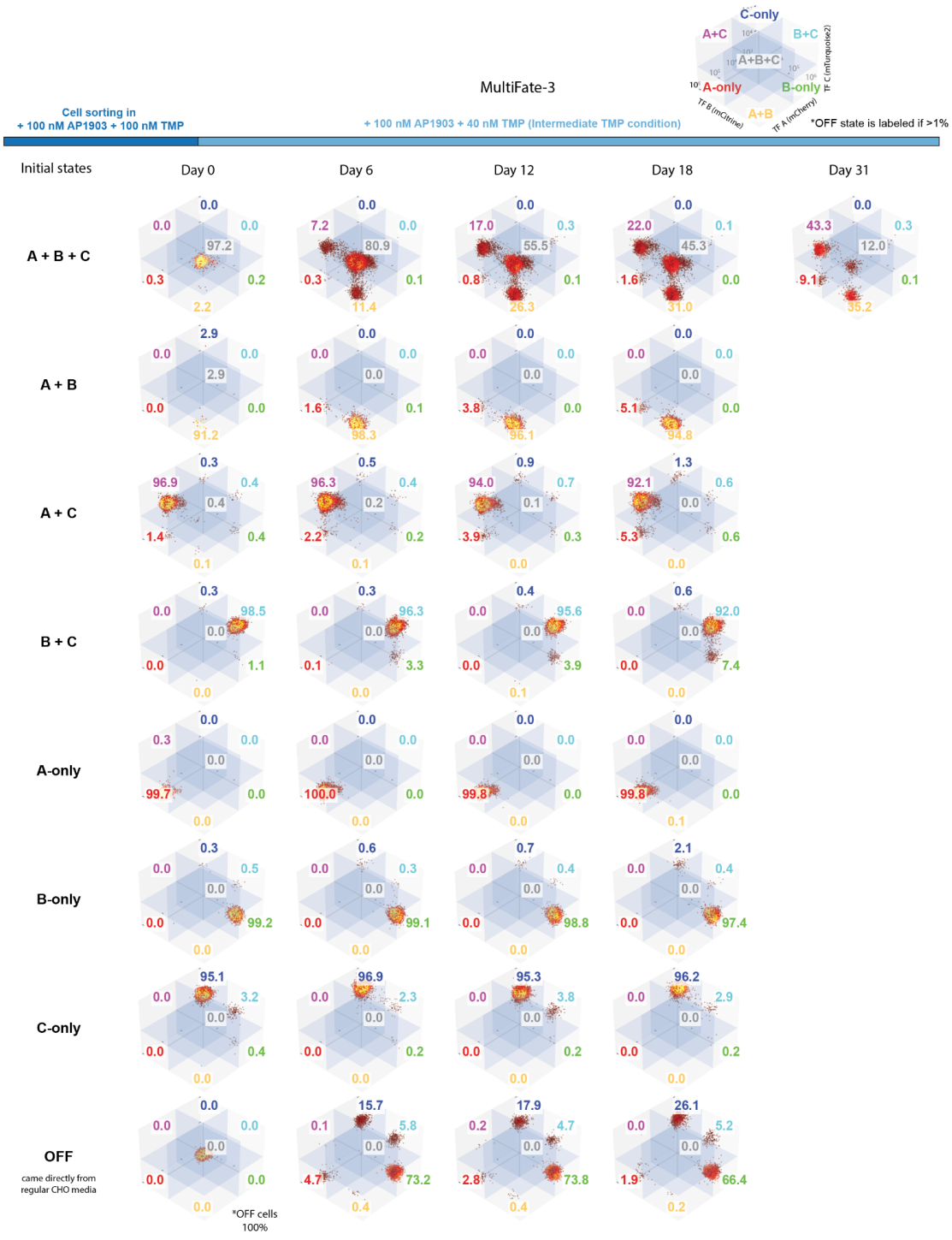

Fig. S2.19 Raw flow cytometry data of MultiFate-3 line under the Intermediate TMP condition.

Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.5B, Intermediate TMP). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top, and initial OFF cells came directly from cells

in regular CHO media without any inducers. For each plot, the percentage of cells in each of the 7 states (excluding the OFF state) is labeled on the corresponding octant. The timeline (top) represents the indicated inducer conditions. OFF state percentages are usually very low (<1%) across all conditions, and are separately labeled if the percentage is greater than 1%. Cells from A+B+C initial state were continuously cultured beyond 18 days and measured at day 31. This extended analysis revealed that cells continuously escaped from A+B+C state, as predicted, under the Intermediate TMP condition.

Fig. S2.20

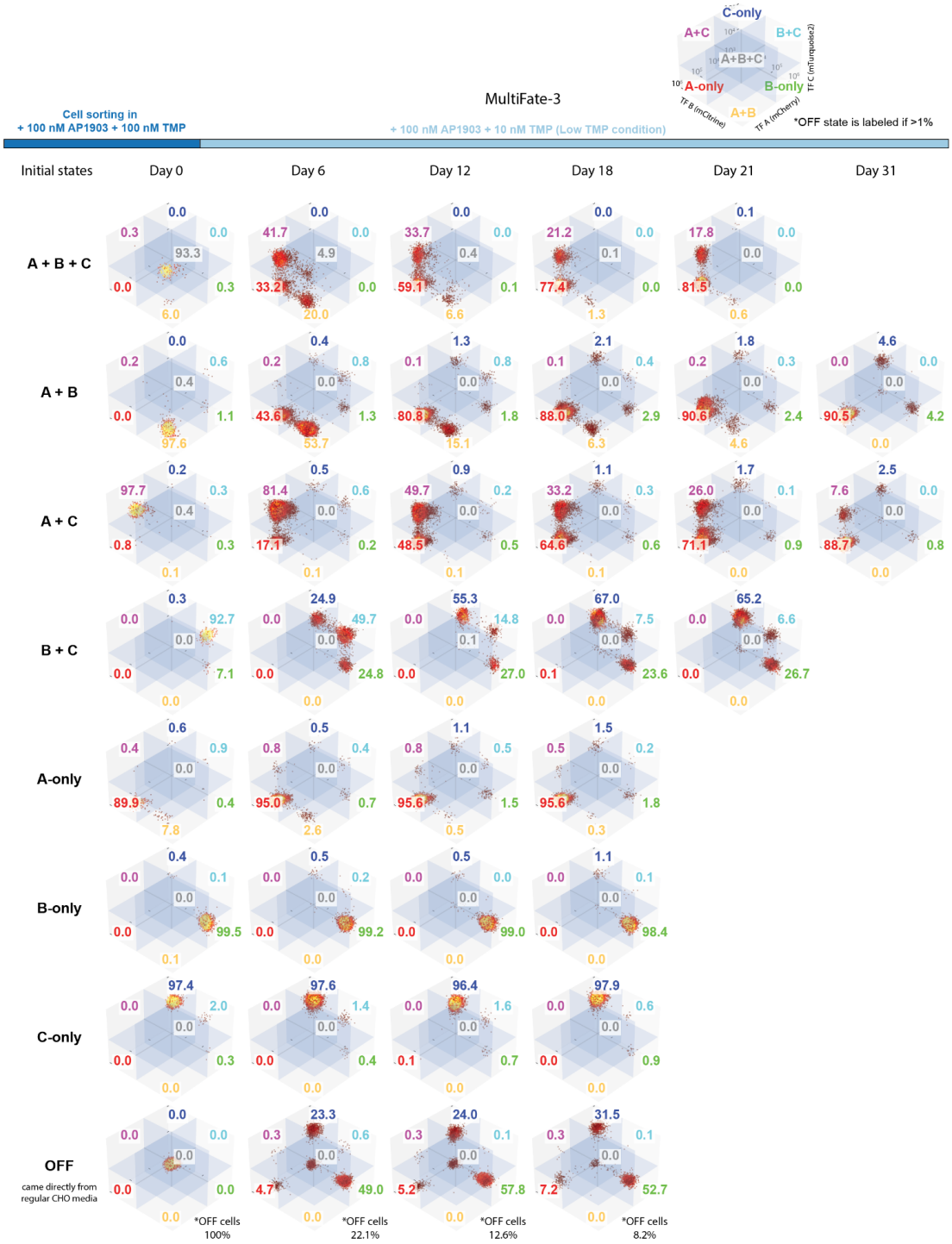

Fig. S2.20 Raw flow cytometry data of MultiFate-3 line under the Low TMP condition.

Each plot represents one of three biological replicates at the indicated time point (cf. Fig. 2.5B, Low TMP). Initial A-only, B-only and A+B cells (rows) were sorted under the media conditions indicated on the top, and initial OFF cells came directly from cells in

regular CHO media without any inducers. For each plot, the percentage of cells in each of the 7 states (excluding the OFF state) is labeled on the corresponding octant. The timeline (top) represents the indicated inducer conditions. OFF state percentages are usually very low (<1%) across all conditions, and are separately labeled if the percentage is greater than 1%. Cells from A+B+C, A+B, A+C and B+C initial states were continuously cultured beyond 18 days. This extended analysis revealed that cells continuously escaped from these unstable states, as predicted, under the Low TMP condition.

Fig. S2.21

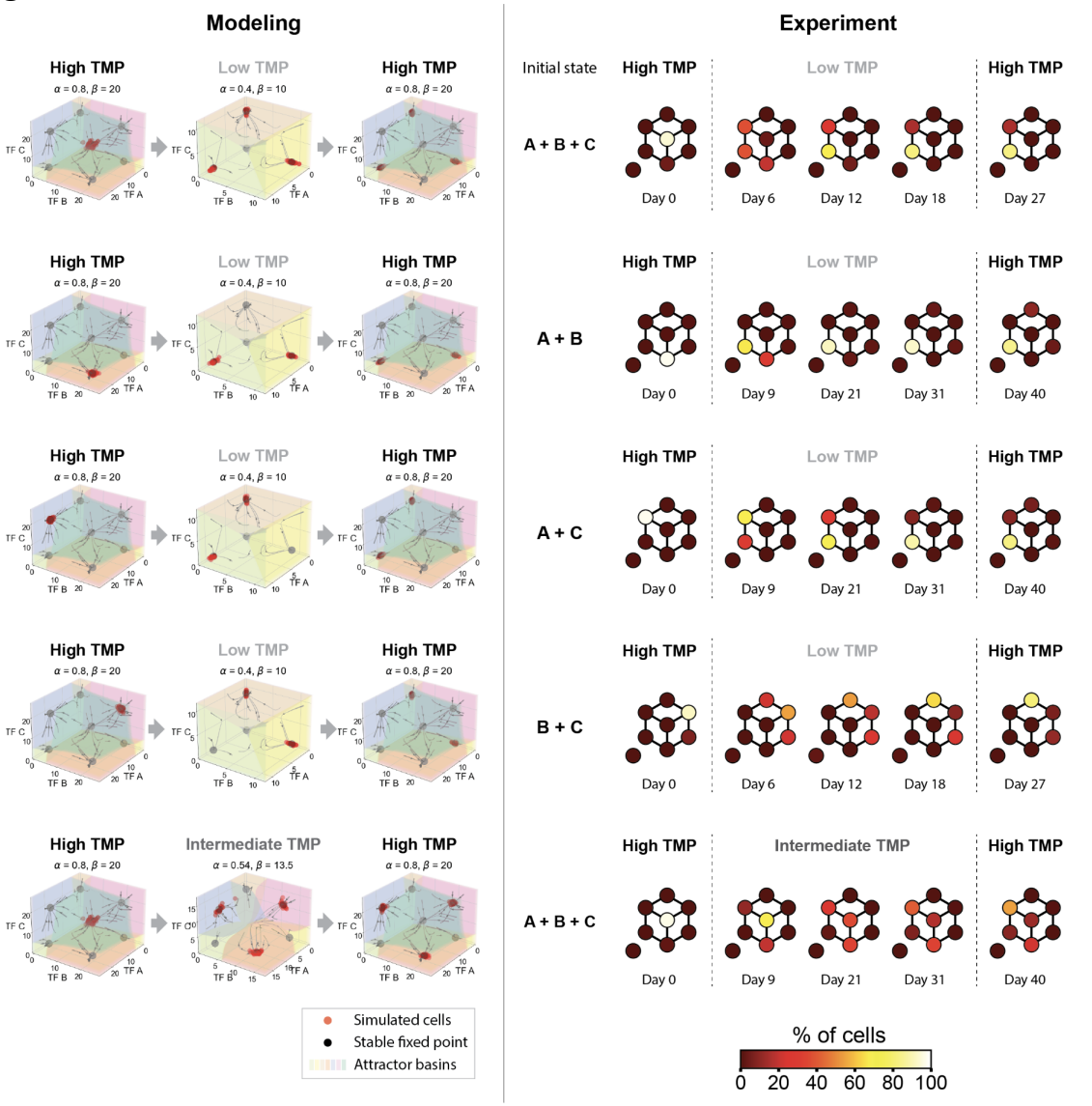

Fig. S2.21 MultiFate-3 exhibits predicted hysteresis.
When transferred from High to Intermediate or Low TMP conditions, cells transition out of destabilized states, as expected. These transitions were irreversible, as shown by both modeling (left) and experiments (right). (Left) The model used here is symmetric and non-dimensionalized, with $K_d = 1$ and $n = 1.5$. The x, y and z axes are total dimensionless concentrations of TF A, B and C, respectively. Simulated cells on phase diagrams were calculated using the Gillespie algorithm. The left column shows initial conditions in simulations, with all cells in a single state at High TMP. The middle column shows steady-state density of cells in different states under Low or Intermediate TMP conditions. The right column shows that cells remain in states in the middle column after switching back to the High TMP condition. (Right) Experiments showed similar hysteretic behaviors, largely consistent with modeling. In each row, initial cells for indicated states were sorted from the High TMP condition, where they were cultured for at least 3 days, and immediately transferred to Intermediate or Low TMP on day 0. They

were then maintained in that condition for 18 or 31 days, as indicated, and then transferred back to the High TMP condition. The color indicates density of cells in each of the indicated states, as in Fig. 2.5. Note that a difference between the simulations and experimental results is that actual cells escaping from destabilized states preferentially occupied the A-only state or states containing high A expression, instead of evenly distributing themselves across all states. This reflects some asymmetry of the experimental MultiFate-3 circuit. High TMP condition = 100 nM AP1903 + 100 nM TMP; Intermediate TMP condition = 100 nM AP1903 + 40 nM TMP; Low TMP condition = 100 nM AP1903 + 10 nM TMP. Each plot represents the mean fractions of three biological replicates.

Fig. S2.22

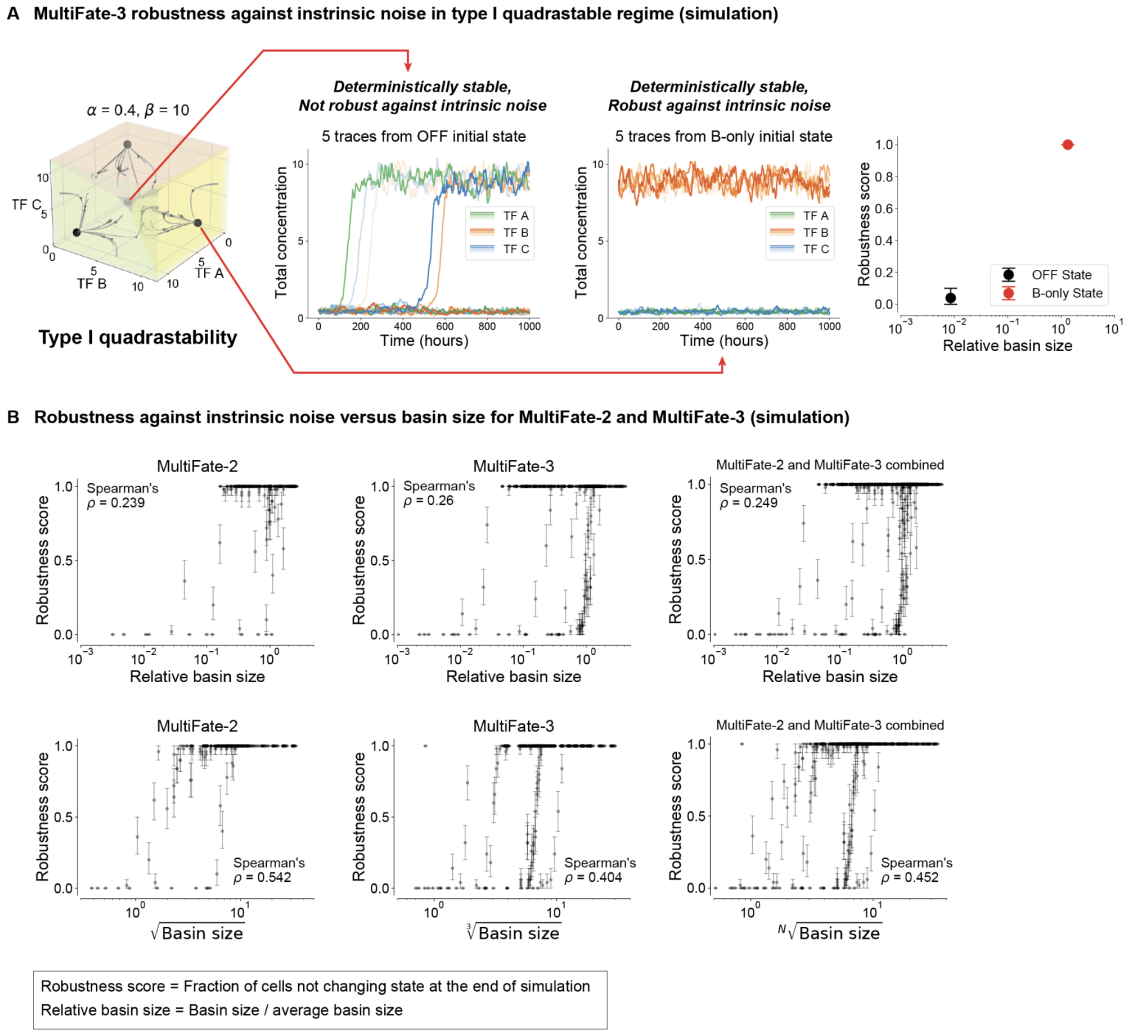

Fig. S2.22 Modeling the robustness of MultiFate against intrinsic biological noise. (A) OFF state is less robust against intrinsic noise compared with B-only state. (Left) The model used here is symmetric and non-dimensionalized, with $K_d = 1$ and $n = 1.5$. The x, y and z axes are total dimensionless concentrations of TF A, B and C, respectively. (Middle) Cells starting from OFF state spontaneously switch out of OFF state, while cells from B-only state stay in their original state at the end of simulation. The traces are generated by Gillespie algorithm. The concentration is dimensionless. (Right) For MultiFate-3 type I quadrastable regime, larger relative basin size corresponds to higher robustness. Robustness score is defined as the fraction of cells not changing state at the end of simulation. The filled dots are the mean robustness scores of 50 simulated cells, and the error bar is the 95% confidence interval generated by bootstrapping. (B) For both MultiFate-2 and MultiFate-3, robustness score is positively correlated with attractor basin size, as shown by the positive Spearman's ρ values. We used Spearman's ρ since it could assess non-linear monotonicity We each simulated 100 sets of parameters with different combinations of α and β for MultiFate-2 and MultiFate-3. For each parameter set, the basin size of each stable fixed point is calculated, and the robustness of each stable fixed

point is quantified by simulating 50 cells starting from that fixed point. For each fixed point, the relative basin size (top) is calculated by dividing its basin size by the average basin size of all fixed points from one parameter set. The characteristic length (bottom) is the Nth root (N is the dimension, either 2 or 3) of basin size, which makes comparison of basin sizes between MultiFate-2 and MultiFate-3 possible.

Fig. S2.23

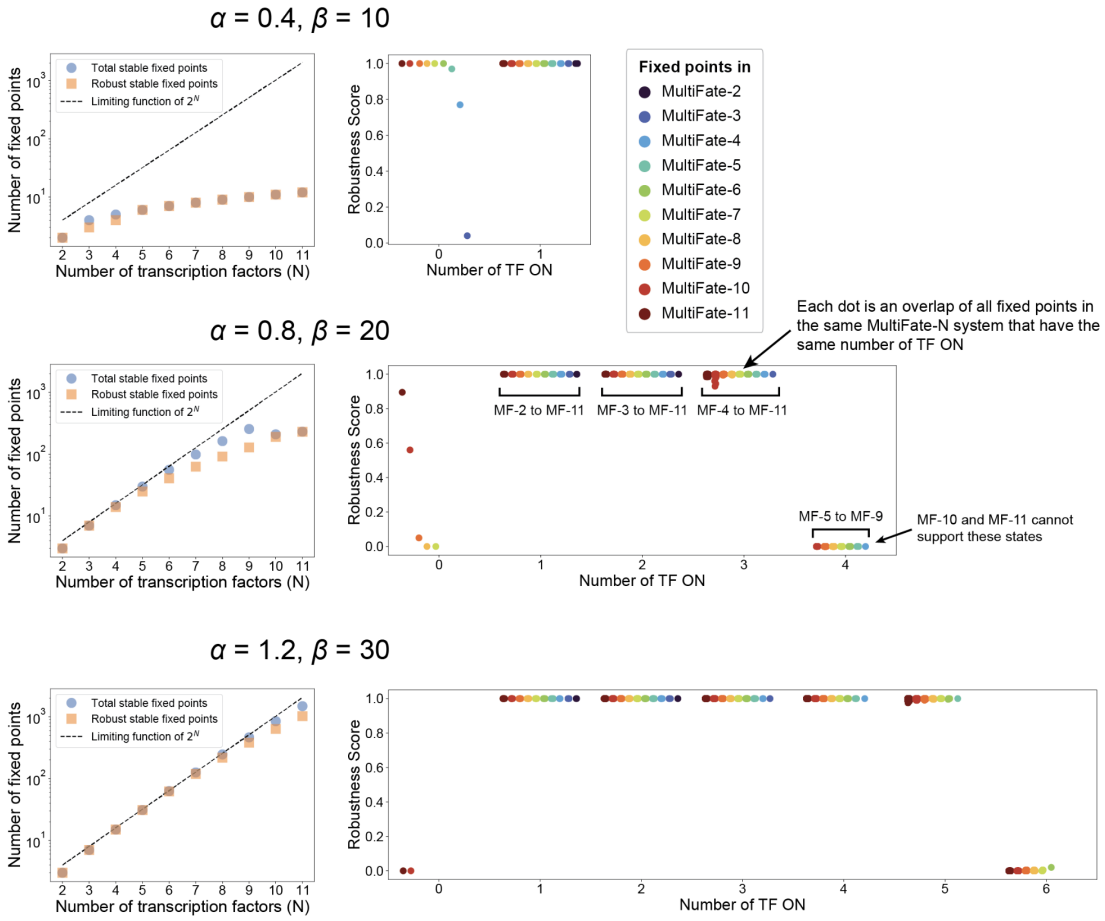

Fig. S2.23 The number of robust stable fixed points increased as MultiFate was expanded to include more transcription factors.
(Left) The number of stable fixed points (blue dots) mostly increased monotonically with the number of transcription factors (except for MultiFate-10 and 11 in the middle row), at a rate slower than the theoretical limit of 2^N (N is the total number of transcription factors). This increase rate (the slope of blue dots) can be modulated up or down by α and β values. Since the model is non-dimensionalized, α and β can be tuned by transcriptional activation strength, protein stability and zinc finger DNA-binding affinity. The theoretical limit is 2^N because each transcription factor can be in either highly expressed or basally expressed state for any fixed points, which resulted in 2^N points considering all combinations of binary states from N transcription factors. Among all stable fixed points, most (orange square) were robust to intrinsic biological noise (*43*), thus the number of robust stable fixed points followed the monotonic increasing trend of total stable fixed points. The robustness of a fixed point was quantified by a robustness score, which was the fraction of simulated cells not escaping from that fixed point at the end of stochastic simulation (*25*). (Right) The number of fixed points grew more slowly than the theoretical limit because each parameter set only supports fixed points with up to a certain number of transcription factors simultaneously expressed at high level (denoted as number of TF ON). For each parameter set, all stable fixed points were plotted in the

same plot, with number of TF ON on x axis and robustness score on y axis. Fixed points from different MultiFate systems were labeled with different colors. Since symmetric models were used, stable fixed points that have the same number of TF ON should have the same robustness score, thus each dot is an overlap of many fixed points. Small deviations resulted from stochasticity in the simulations. Low α and β values (first row) only supported fixed points up to one transcription factor highly expressed. High α and β values (second row) supported fixed points with up to four transcription factors highly expressed. Among them, fixed points with up to three transcription factors simultaneously ON were robust. A parameter set with even higher α and β values (third row) supported fixed points with more transcription factors simultaneously ON. In these regimes, OFF fixed points sometimes were not stable or not robust. Here we used symmetric, non-dimensionalized and expanded MultiFate models with $K_d = 1$ and $n = 1.5$.

Fig. S2.24

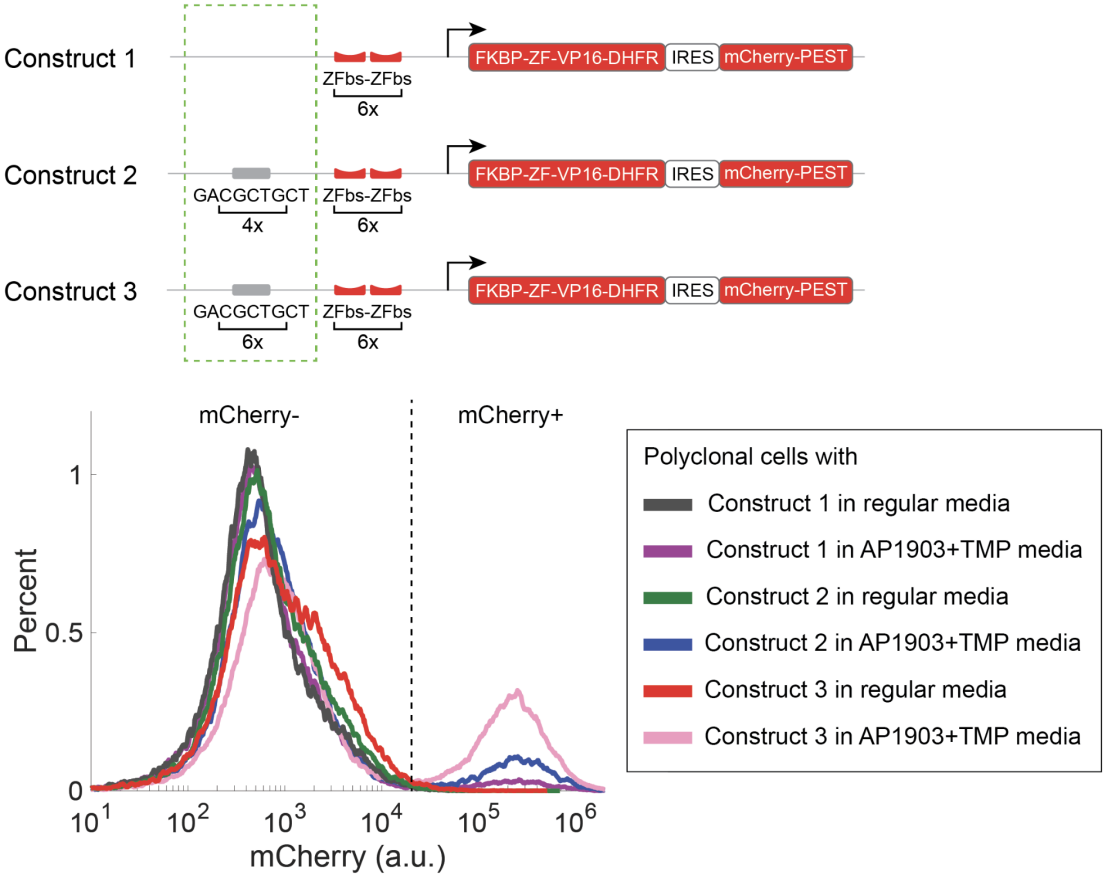

Fig. S2.24 Basal promoter expression can be modulated by modifying promoter sequences.

Basal promoter expression and spontaneous self-activation in CHO cells can be increased by inserting GACGCTGCT repeats in the promoter. Note that GACGCTGCT is also the sequence motif bound by 42ZF, but its effect in increasing basal promoter expression does not require 42ZF. (Top) Schematics of three self-activation constructs different only in the number of GACGCTGCT repeats in the promoter. ZF = BCRZFR39A. Detailed construct maps are in Table S2. (Bottom) Increasing the number of GACGCTGCT repeats in the promoter increased the basal promoter expression, which can be observed by increased right shifts of mCherry- cell populations in regular media (red > green > black). Higher basal promoter expression resulted in an increase in the fractions of cells (mCherry+ cells) that can spontaneously self-activate upon the addition of 100 nM AP1903 and 10 μM TMP (pink > blue > magenta). Each of the three polyclonal cell populations was generated by stably integrating each construct in the CHO-K1 cells (Table S2.3) (*25*). Different from Fig. 2.2C, polyclonal cell population was transferred directly from regular media to media containing AP1903 and TMP (instead of adding transient Dox treatment in between) to test spontaneous self-activation. Cells were harvested and measured by flow cytometry after 48 hours in regular media or AP1903+TMP media.

2.9 Supplementary Materials

Plasmid construction

Constructs used in this study are listed in Table S2.2. Some constructs were generated using standard cloning procedures. The inserts were generated using PCR or gBlock synthesis (IDT) and were annealed by Gibson assembly with backbones that are linearized using restriction digestion. The rest of the constructs were designed by the authors and synthesized by GenScript. All the construct maps are available at data.caltech.edu/records/1882, and selected constructs used to build MultiFate lines are deposited onto Addgene.

Tissue culture

Chinese hamster ovary K1 cells (CHO-K1, ATCC) were cultured at 37ºC in a humidity-controlled chamber with 5% CO_2. The growth media consisted of Alpha MEM Earle's Salts (FujiFilm Irvine Scientific) supplemented with 10% FBS, 1 U/ml penicillin, 1 µg/ml streptomycin and 1 mM L-glutamine. For experiments requiring a change of inducer conditions, cells were first washed 3 times using media with the new inducer condition. After the wash, cells were rinsed once with Dulbecco's Phosphate-Buffered Saline (DPBS, Life Technologies) and trypsinized with 0.25% Trypsin (Life Technologies) for 3 min at 37ºC. Trypsinized cells were then transferred into a new well with media added with the new inducer condition.

Transient transfection

24 hours before transfection, $0.05x10^6$ CHO-K1 cells were seeded per well of a 24-well plate using standard culture media. The next day, cells were transfected with plasmids using Lipofectamine LTX and PLUS Reagents (Thermo fisher) according to manufacturer's protocol.

Cell line construction

Stable cell lines used in this study are listed in Table S2.2**.** Stable cell lines were generated using the PiggyBac Transposon system (System Biosciences). CHO-K1 cells in a 24-well plate were co-transfected with transgene constructs in a PiggyBac expression backbone, an EF1α-PuroR plasmid and a Super PiggyBac Transposase plasmid. Cells were transferred into a 6-well plate and selected with 10 μg/ml puromycin for 3 days to obtain a stable polyclonal population.

To identify potential MultiFate-2 clones that could operate in several multistability regimes, we turned to our MultiFate mathematical model for intuition. Through parameter screening we found that when one progressively reduces protein stability starting from a value at which the state with all transcription factors expressed simultaneously (all-ON) was stable could generate a progressive reduction of multistability (Fig. S1B).

To achieve similar behavior experimentally, we sought to select MultiFate-2 monoclones that exhibit stable A+B state (Fig. S2.7A). We transiently induce the expression of all transcription factors in polyclonal MultiFate-2 cells by Dox (Sigma-Aldrich) (and 4-OHT (Sigma-Aldrich) if the second cassette has UAS) for 36 hours, then washed and replaced

with media containing 100 nM AP1903 (MedChemExpress) and 10 μM TMP (Sigma-Aldrich) (each at the saturating concentration). After another 3 days, polyclonal cells with both A and B still activated were sorted by FACS as single cells into 384-well plates. The plates were checked under microscope after 4-5 days to eliminate wells without cells or with more than one colony. For wells that only have a single colony growing, cells were expanded, and subsequent screening was performed to obtain MultiFate-2 monoclones. Using a similar method, a monoclonal that could maintain the stability of the A+B+C state was selected as the MultiFate-3 line (Fig. S2.7B). All the MultiFate lines used in the current study are available from the corresponding author.

Flow cytometry

All samples were harvested from 24-well plates. Cells were first rinsed with 500 μl DPBS, and then trypsinized with 75 μl 0.25% trypsin for 3 min at 37ºC. Trypsin was neutralized by resuspending cells in 300 μl flow cytometry buffer containing Hank's Balanced Salt Solution (Life Technology) and 2.5 mg/ml Bovine Serum Albumin. Cell samples were then filtered by 40 μm cell strainers and analyzed by a flow cytometer (CytoFLEX, Beckman Coulter). We used a Matlab-based software package developed by Yaron Antebi to process flow cytometry data (https://antebilab.github.io/easyflow/).

Characterization of ZF transcription factors

To characterize transcriptional activation of different ZF transcription factor variants (Fig. 2A, Fig. 2B and S4), we transfected CHO-K1 cells in 24-well plates with 50ng ZF transcription factor plasmid (Table S2.2, construct MF08–MF62), 50ng reporter plasmid (Table S2, construct MF01–MF07) and 25ng EF1α-mTagBFP2. In the “ReporterOnly”

group, ZF transcription factor plasmid was replaced by an empty plasmid with only a constitutive promoter but no ZF transcription factor. In the “NoReporter” well, both ZF transcription factor plasmid and reporter plasmid were replaced by an empty plasmid. For ZF transcription factors with FKBP homodimerization domain (Fig. 2.2B), AP1903 was added to the transfection media. 36 hours after transfection, cells were harvested and analyzed by flow cytometry. To maximize the reporter dynamic range, we selected and compared highly transfected cells by gating cells with high levels of a BFP co-transfection marker. Median citrine fluorescence intensity of gated cells was used to calculate fold activation. To calculate fold activation, median fluorescence values of NoReporter samples, representing the cellular autofluorescence background, were first subtracted from ReporterOnly and Reporter+ZF samples. The ratio between background-subtracted Reporter+ZF value and background-subtracted ReporterOnly value is then the fold activation of that ZF transcription factor on that reporter.

Characterization of ZF transcription factor self-activation

Each self-activation construct (Table S2.2, construct MF63–MF69) was stably integrated into Tet3G-expressing CHO-K1 cells. After puromycin selection, polyclonal cells were transferred into media containing 500 ng/ml Dox to transiently express ZF transcription factors. After 24 hours of Dox treatment, cells were washed 3 times with regular media and transferred into media containing different concentrations of AP1903 and/or TMP to test how dimerization and/or protein stability affect self-activation. One sample of cells (Dox+ sample) continued to be cultured in 500 ng/ml Dox as the positive control. After another 72 hours, cells were harvested and analyzed by flow cytometry. Stable polyclonal cells showed a strong bimodal mCitrine distribution upon circuit activation (Fig. 2.2C,

middle). We used an empirical threshold at mCitrine = 10^4 fluorescence units to separate the population into mCitrine- (cells with no circuit integrated or integrated circuit cannot self-activate) and mCitrine+ (cells with integrated circuit activated) subpopulations. To only consider cells with at least one stably integrated activatable circuit, we normalized the mCitrine+ fraction of each sample to the mCitrine fraction of Dox+ sample, in which high concentrations of Dox should turn on all stably integrated activatable cassettes. This normalized mCitrine+ fraction was used to compare self-activation strength across different AP1903 and TMP combinations.

Assay showing inhibition of self-activation by competitive dimerization

Two monoclonal self-activation stable lines (with 42ZFR2AR39AR67A DNA-binding domain and either GCN4 or FKBP as the homodimerization domains, see Table S2.3) were selected since they showed spontaneous and homogeneous self-activation upon the addition of 100 nM AP1903 and 10 μM TMP (Fig. S2.5B). To whether competitive dimerization inhibits self-activation (Figs. 2.2D and S2.5B), we stably integrated plasmids (Table S2.2, construct MF72–MF80) expressing different transcription factor variants and a co-translational mCherry in these two monoclonal lines. After puromycin selection, cells were transferred into media containing 100 nM AP1903 + 10 μM TMP to permit self-activation, and measured by flow cytometry after another 72 hours. The mCherry+ cell population was gated for analysis. We introduced protein variants through stable integration, instead of transient transfection, to avoid nonspecific transcriptional interference by transient high expression of proteins during transfection, and to test inhibition of self-activation in a cellular environment better mimicking the MultiFate-2 and MultiFate-3 stable cell lines.

Fluorescence Activated Cell Sorting (FACS)

To separate MultiFate cells in distinct states for subsequent experiments (Figs. 2.3 to 2.5), cells were harvested and resuspended in sorting buffer (BD FACS Pre-Sort Buffer) supplemented with 1 U/ml DNAse I, AP1903 and TMP. Cells were then sorted into media containing different concentrations AP1903 and TMP, according to the experiments. Cell sorting was performed by Caltech Flow Cytometry Facility.

Flow cytometry measurement of long-term state stability of MultiFate cells

To characterize long-term state stability of MultiFate cells in different media conditions (Figs. 2.3 to 2.5), we trypsinized and transferred 4% of cells into fresh media with the same condition every three days. The remaining cells were resuspended in flow cytometry buffer and analyzed by a flow cytometer. The resulting two-dimensional (or three-dimensional for MultiFate-3) fluorescence intensity plots were then divided into four quadrants (or eight octants) by an empirical threshold in each of the two (or three) fluorescence channels. The exact values of empirical thresholds for different MultiFate cell lines are slightly different due to expression differences and are shown in the figure legends of Fig. S2.9 to S2.11 and S2.17. The percentage of cells in quadrants (or octants) were then calculated for each sample. The mean percentage of cells across three samples was plotted as a square (or a hexagon) with colored circles representing the percentages.

Time-lapse imaging

To visualize state stability of MultiFate cells (Figs. 2.3D and 2.5C), MultiFate cells from different states were mixed at equal ratio, and sparsely seeded in an imaging 24-well plate (μ-Plate 24 Well Black, ibidi) with media containing AP1903 and TMP. After 6 or 12

hours, we aspirate media to remove unattached cells and add fresh media containing AP1903 and TMP. Time-lapse images were acquired on an inverted Olympus IX81 fluorescence microscope with Zero Drift Control (ZDC), an ASI 2000XY automated stage, a Photometrics 95B camera (Teledyne Photometrics) and a 20x UPlanS/Apo objective (0.75 NA, Olympus). Fluorescent proteins were excited with an X-Cite XLED1 light source (Lumen Dynamics). Images were automatically acquired every one hour, controlled by MetaMorph software (Molecular Devices). Cells were kept in a custom-made environmental chamber enclosing the microscope, controlling a humidified, 37ºC and 5% CO_2 atmosphere. Media was changed every three days.

Measuring doubling time of MultiFate cells in different states

To measure the doubling time of different MultiFate lines in different states (Fig. S2.8), we first separated cells from different states using cell sorting. We then cultured cells from OFF state in regular media and cells from all other states in media containing AP1903 and TMP, so that cells do not change state during measurement. For each measurement, we plated the same number of cells in two wells in a 48-well plate. We counted one well after 24 (or 48) hours, and counted the other well after 72 hours. Note that the wells were still sub-confluent at 72 hours. The doubling time is then $\tau_D = \frac{\Delta t}{log_2(N_{t_2} - N_{t_1})}$, where Δt is the time difference between the two timepoints, N_{t_1} is the cell number at 24 (or 48) hours, and N_{t_2} is the cell number at 72 hours.

Robustness analysis of MultiFate circuit

All the stable fixed points attract surrounding cells in the deterministic model. However, random transcription factor concentration fluctuations arose from stochasticity of chemical reactions (intrinsic noise) (*43*) such as transcription and translation may switch cells from one stable fixed point to another. The robustness of a stable fixed point is measured by how infrequent cells from that stable fixed point spontaneously switch to other fixed points due to intrinsic noise. We used Gillespie simulations (*41*) to analyze the robustness of MultiFate circuit (Figs. 2.5D, S2.22 and S2.23). Molecular reactions and their propensities for Gillespie simulation are listed in Table S2.4. To quantify the robustness of a stable fixed point, we first simulate trajectories of cells starting from that stable fixed point. Then the robustness of that fixed point is quantified by a robustness score, defined by the fraction of cells not switching out of that stable fixed point at the end of simulation (1000 hours). The higher the robustness score, the more robust a stable fixed point is against intrinsic noise. In Fig. 2.5D, robust stable fixed points have robustness scores greater than 0.9, which means less than 10% of cells spontaneously escape at the end of stochastic simulations.

Attractor basin analysis of MultiFate circuit with N transcription factors

In theory, the attractor basin volume of a stable fixed point goes to infinity (except for the fixed point with all transcription factors OFF) if the concentration of each transcription factor has no limit. However, transcription factor concentrations are bounded by their maximum expression levels. Consequently, attractor basin volumes are finite and volumes of different basins can be compared. In the non-dimensionalized model,

transcription factor concentrations are confined to the interval [α, α + β], which corresponds to the equilibrium concentrations when a transcription factor is not self-activating and when it is fully self-activating, respectively. To calculate the approximate volumes of attractor basins (Fig. S2.22), we initialized an N-dimensional concentration grid with k points in each dimension, spaced at equal linear intervals, for a total of k^N points. Using these grid points as initial conditions, we numerically solved the differential equations to compute forward trajectories to their final stable fixed point, using the expanded MultiFate-N model (see below). We labeled each grid point based on which stable fixed point its trajectory ends at. The attractor basin volume of a stable fixed point was then approximated as $V_i = \frac{number\ of\ grid\ points\ that\ end\ at\ fixed\ point\ i}{total\ number\ of\ grid\ points} * \beta^N$, where β^N represents the total phase space volume.

Parameter screening of MultiFate-2 and MultiFate-3 circuits

Each parameter dependency plot in Figs. S2.1B and S2.2B represents a field of 100×100 points. The color of each point denotes the multistability type generated through MultiFate-2 (Fig. S2.1B) or MultiFate-3 (Fig. S2.2B) with the indicated parameter combination. To identify the multistability type of each parameter set, we initialized a 2- or 3-dimensional grid in the 2- or 3-dimensional space of transcription factor concentrations, for MultiFate-2 and MultiFate-3, respectively, with eight values for each transcription factor concentration. Using these grid points as starting points, we computed trajectories using the MultiFate differential equations. We then grouped the end points of these trajectories into clusters and used the centers (center of mass) of these clusters as estimated locations of stable fixed points. We double-checked the stability of the

estimated fixed points using standard linear stability analysis (*53*) type by the locations of these stable fixed points.

Asymmetry parameter fitting of MultiFate-2 and MultiFate-3 lines

While symmetric MultiFate-2 and MultiFate-3 models (Box 2.1) accurately predict the number of stable states for the experimental MultiFate-2 and MultiFate-3 circuits in different protein stability regimes, they cannot explain the bias of cells towards certain stable states when they transition away from an unstable state (for example, MultiFate-2.3 cells almost exclusively transition from the unstable OFF state to the B-only state in High TMP condition). This is because the dynamics of different transcription factors in the experimental MultiFate lines are asymmetric due to differences in the zinc fingers used, copy numbers of integrated cassettes and other factors. To explain the dynamics of different MultiFate lines, we added asymmetry parameters into the symmetric MultiFate models to construct asymmetric MultiFate models (see "Asymmetric MultiFate model" section in the Supplementary Information). To find the best fitted asymmetry parameter set for MultiFate-2 lines, we uniformly sampled 16 points for each of the four asymmetry parameters (r, m, κ, γ, each ranging from 0.5 to 2) and constructed a four-dimensional parameter space consisting of 16^4 = 65536 parameter sets. We first filtered out parameter sets that do not generate the expected numbers of stable states in High TMP (3 states) and Low TMP (2 states) conditions. For the remaining parameter sets, we used the Gillespie algorithm to simulate the exit of 200 cells from the OFF state in both High TMP and Low TMP conditions, and from the A+B state in the Low TMP condition (3 simulations total for each parameter set). We then compared cell fractions at the end of simulations with cell fractions at the

end of continuous culture (cf. Figs. S2.9 to S2.11, S2.17, S2.19, S2.20 and other two replicates) by calculating the mean squared error (MSE). The best fitted parameter set for each MultiFate-2 line is the one that has the lowest MSE. Finally, we validated the parameter fitting results by simulating 400 cells starting from OFF, A-only, B-only and A+B state in both High TMP and Low TMP conditions (8 simulations) and plotted the simulated cell fractions side-by-side with experimental cell fractions in Fig. S2.15A.

For the MultiFate-3 line, since it is constructed from MultiFate-2.2 line, we first chose around the best fitted parameter set for MultiFate-2.2 for each of the first four asymmetry parameters (r, m, κ, γ, each sampled 3 points including the best fitted parameter for MultiFate-2.2 and +/- 0.1). For the four new asymmetry parameters (r_2, m_2, κ_2, γ_2 for the new TF C), we again uniformly sampled 16 points for each of the four new asymmetry parameters ranging from 0.5 to 2. This results in an 8-dimensional parameter space consisting of 3^4 x 16^4 = 5308416 parameter sets. Using the same method for MultiFate-2 fitting, we found the best fitted parameter set for the MultiFate-3 line, validated the parameter fitting results by simulating 400 cells starting from 8 states in High TMP, Intermediate TMP and Low TMP conditions (24 simulations) and plotted the simulated cell fractions side-by-side with experimental cell fractions in Fig. S2.15B.

Data and code availability

All data generated and all the computational and data analysis and modeling code used in the current study are available at data.caltech.edu/records/1882.

2.10 Supplementary Information

Non-dimensionalization of MultiFate model

As shown in Box 2.1, each ordinary differential equation (ODE) for $[A_{tot}]$ and $[B_{tot}]$ consists of three terms: (i) a basal protein production rate α, (ii) a Hill function describing self-activation dynamics with maximal rate β, Hill coefficient n, and half-maximal activation at a homodimer concentration of K_M, and (iii) a protein degrade rate δ. We can then write:

$$\frac{d[A_{tot}]}{dt} = \alpha + \frac{\beta[A_2]^n}{K_M{}^n + [A_2]^n} - \delta[A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = \alpha + \frac{\beta[B_2]^n}{K_M{}^n + [B_2]^n} - \delta[B_{tot}]$$

Dimerization dynamics occur on a faster timescale than protein production and degradation (*52*). This separation of timescales permits us to assume that the distribution of monomer and dimer states remains close to equilibrium, generating the following relationships between the concentrations of monomers ($[A]$ and $[B]$), and dimers ($[A_2]$, $[B_2]$, and$[AB]$) based on the law of mass action:

$$[A]^2 = K_d[A_2]$$

$$[B]^2 = K_d[B_2]$$

$$2[A][B] = K_d[AB]$$

Here, because the two transcription factors share the same dimerization domain, homo- and hetero-dimerization are assumed to occur with equal dissociation constants, K_d. When deriving these three equations from the law of mass action, each monomer is counted twice in homodimerization reactions, and is counted once in the heterodimerization reaction, thus a factor of two is introduced in the third equation to account for this statistical difference. Additionally, conservation of mass implies that $[A_{tot}] = [A] + [AB] + 2[A_2]$, with a similar relationship for B.

Solving these equations produces expressions for the concentrations of the activating homodimers in terms of the total concentrations of A and B:

$$[A_2] = \frac{2[A_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d^2 + 8([A_{tot}] + [B_{tot}])\,K_d}}$$

$$[B_2] = \frac{2[B_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d^2 + 8([A_{tot}] + [B_{tot}])\,K_d}}$$

To non-dimensionalize the model, we rescale time in units of δ^{-1}, and concentrations in units of K_M. Then we have:

$$t \leftarrow t\delta \qquad [A_{tot}] \leftarrow [A_{tot}] / K_M \quad [B_{tot}] \leftarrow [B_{tot}] / K_M \qquad [A_2] \leftarrow [A_2] / K_M$$

$$[B_2] \leftarrow [B_2] / K_M \qquad \alpha \leftarrow \alpha/(K_M\delta) \qquad \beta \leftarrow \beta/(K_M\delta) \qquad K_d \leftarrow K_d/K_M$$

Here, the quantity to the left of the arrow is the parameter in the non-dimensionalized system. Thus, in the first assignment, the non-dimensionalized time, t, is equal to dimensionalized time multiplied by δ.

We can then write the system using the non-dimensionalized quantities:

$$\frac{\delta dK_M[A_{tot}]}{dt} = K_M\delta\alpha + \frac{K_M\delta\beta(K_M[A_2])^n}{K_M{}^n + (K_M[A_2])^n} - \delta K_M[A_{tot}]$$

$$\frac{\delta dK_M[B_{tot}]}{dt} = K_M\delta\alpha + \frac{K_M\delta\beta(K_M[B_2])^n}{K_M{}^n + (K_M[B_2])^n} - \delta K_M[B_{tot}]$$

and

$$K_M[A_2] = \frac{2K_M{}^2[A_{tot}]^2}{K_MK_d + 4K_M([A_{tot}] + [B_{tot}]) + \sqrt{K_M{}^2K_d{}^2 + 8K_M{}^2([A_{tot}] + [B_{tot}])K_d}}$$

$$K_M[B_2] = \frac{2K_M{}^2[B_{tot}]^2}{K_MK_d + 4K_M([A_{tot}] + [B_{tot}]) + \sqrt{K_M{}^2K_d{}^2 + 8K_M{}^2([A_{tot}] + [B_{tot}])K_d}}$$

After canceling δ and K_M from both side of equations, we obtain the non-dimensionalized MultiFate model:

$$\frac{d[A_{tot}]}{dt} = \alpha + \frac{\beta[A_2]^n}{1 + [A_2]^n} - [A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = \alpha + \frac{\beta[B_2]^n}{1 + [B_2]^n} - [B_{tot}]$$

$$[A_2] = \frac{2[A_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d{}^2 + 8([A_{tot}] + [B_{tot}])K_d}}$$

$$[B_2] = \frac{2[B_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d^2 + 8([A_{tot}] + [B_{tot}])K_d}}$$

This non-dimensionalization leaves us with four parameters: rescaled basal protein production rate, α, rescaled maximal protein production rate in the Hill function, β, rescaled dimerization dissociation constant, K_d, and Hill coefficient n.

Aparts from these four parameters, we used the mathematical model to predict how protein stability controls the number of stable fixed points in many parts of this study. In the non-dimensionalized model, the protein degradation rate, δ, does not appear explicitly but enters through the rescaling of α and β by $(\delta K_M)^{-1}$ as shown above. Thus, tuning protein stability is equivalent to multiplying bothα and β by a common factor, which we term the "protein stability factor" in this study.

MultiFate-2 model incorporating external inputs

To allow external control of MultiFate circuit by inducers, we designed binding sites for ERT2-Gal4 (induced by 4-OHT) and Tet3G (induced by Dox) at the promoters of TF A self-activation construct (Table S2.2, MF84) and TF B self-activation construct (Table S2.2, MF64), respectively. These two constructs were used to make the switchable MultiFate-2.3 cells (Table S2.3), in which we can independently control the expression of TF A and TF B by 4-OHT and Dox, respectively. Since promoters of these self-activation cassettes contain binding sites for both inducer-responsive activators (ERT2-Gal4 or Tet3G) and zinc finger transcription factor homodimers, transcriptional activation of these cassettes follows an OR logic, i.e. promoter is activated when inducer-responsive

activators or zinc finger transcription factor homodimers are bound. To model OR activation logic, we modified the Hill functions in the non-dimensionalized MultiFate model by adding an ind_1 (or ind_2) term, which represents the activation strength of the ERT2-Gal4 (or Tet3G) system, to both numerator and denominator:

$$\frac{d[A_{tot}]}{dt} = \alpha + \frac{\beta([A_2]^n + ind_1)}{1 + ([A_2]^n + ind_1)} - [A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = \alpha + \frac{\beta([B_2]^n + ind_2)}{1 + ([B_2]^n + ind_2)} - [B_{tot}]$$

In the experiments (Fig. 2.4B and S2.16), we only used saturating concentrations of 4-OHT or Dox to achieve a full activation of cassettes. To model this, we choose an arbitrary ind_1 (or ind_2) $= 100$ when inducer is added, so that the Hill function $\frac{[A_2]^n + ind_1}{1 + ([A_2]^n + ind_1)}$ (or $\frac{[B_2]^n + ind_2}{1 + ([B_2]^n + ind_2)}$) is close to its maximum 1, representing full promoter activation when a saturating concentration of inducer is added. When there is no inducer, ind_1 (or ind_2) $= 0$. Using this modified model, we simulated the state-switching dynamics shown in Fig. S2.3.

MultiFate model expanded to N transcription factors

The minimal MultiFate-2 model can be expanded in a straightforward way to include more transcription factors. To start, we consider the distribution of transcription factors X_1, X_2, X_3, … X_N among different dimerization states. For each transcription factor, the total concentration can be expressed as,

$$[X_{tot,\,i}] = [X_i] + \sum_{j \neq i} [X_i X_j] + 2[X_{2,\,i}]$$

$$for\ i = 1,\ 2,\ 3,\ \ldots\ N$$

Here $[X_i]$ denotes the concentration of transcription factor X_i monomers, $[X_{2,\,i}]$ denote the concentration of homodimers, and $[X_i X_j]$ denote the concentration of heterodimers formed by X_i and X_j ($i \neq j$). As mentioned above, we assume that homo- and hetero-dimerization occur with equal dissociation constants, K_d, reflecting the use of the same dimerization domain for both proteins. Because dimerization dynamics occur on a faster timescale than protein production and degradation, the protein dimerization states approximately follow their equilibrium values:

$$[X_i]^2 = K_d[X_{2,\,i}]$$

$$2[X_i][X_j] = K_d[X_i X_j]$$

Solving these equations produces a simple expression for the concentrations of the activating homodimers in terms of the total concentrations of all transcription factor species:

$$[X_{2,\,i}] = \frac{2[X_{tot,\,i}]^2}{K_d + 4\sum[X_{tot,\,i}] + \sqrt{K_d^{\,2} + 8K_d\sum[X_{tot,\,i}]}}$$

$$for\ i = 1,\ 2,\ 3,\ \ldots\ N$$

With these expressions, we can describe protein production and degradation dynamics using ODEs for $[X_{tot,i}]$ in a similar way as shown in Box 2.1. After non-dimensionalization and adding asymmetric parameters, we obtain ODEs for protein production and degradation in the expanded MultiFate model:

$$\frac{d[X_{tot,i}]}{dt} = r_i\alpha + \frac{m_i\beta[X_{2,i}]^n}{\kappa_i^n + [X_{2,i}]^n} - \gamma_i[X_{tot,i}]$$

$$for\ i = 1, 2, 3, \ldots N$$

where α represents the basal protein production, βrepresents the maximal protein production rate in the Hill function, n represents Hill coefficient and r_i, m_i, κ_i, and γ_i represents the asymmetric parameters for transcription factor X_i.

Asymmetric MultiFate model

While the symmetric MultiFate model precisely predicts many experimental results, we observed some asymmetric behaviors in MultiFate-2 and MultiFate-3 lines. For example, MultiFate-2.2 cells in A+B state preferentially migrated towards A-only state when transferred from the High TMP condition to the Low TMP condition (Fig. 2.3C). This kind of asymmetric behavior could result from several potential differences among integrated gene cassettes:

(i) A, B and C used three different ZF DNA-binding domains. As shown by the different fold changes in Fig. S4A, they could have different binding affinity to ZF binding sites, resulting in different $K_M$values, and different activated transcriptional rates, resulting in different β values.

(ii) The integration number of different genes and the genomic environment of different integrated cassettes could be different, which may affect basal and activated promoter expression, resulting in different values of α and β.

(iii) Due to sequence differences in ZF DNA-binding domains, the protein stability could be different for different genes, resulting in different values of δ.

To analyze such asymmetries, we now allow for distinct values of these parameters, indicated by subscripted A, B or C. While we allow asymmetry in these parameters, we still assume symmetry in others. Specifically, we maintain the same Hill coefficient, n, and dissociation constant for dimerization, K_d, for all factors, because they share the same transcriptional activation domain and the same homodimerization domain. With these assumptions, we can then write down an asymmetric dimensionalized model:

$$\frac{d[A_{tot}]}{dt} = \alpha_A + \frac{\beta_A [A_2]^n}{K_{MA}{}^n + [A_2]^n} - \delta_A [A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = \alpha_B + \frac{\beta_B [B_2]^n}{K_{MB}{}^n + [B_2]^n} - \delta_B [B_{tot}]$$

$$\frac{d[C_{tot}]}{dt} = \alpha_C + \frac{\beta_C [C_2]^n}{K_{MC}{}^n + [C_2]^n} - \delta_C [C_{tot}]$$

As above, we then non-dimensionalize the model by rescaling time in units of $\delta_A{}^{-1}$, and concentrations in units of K_{MA}. Then we have (after canceling K_{MA} and $\delta_A{}^{-1}$ from both sides of the equations):

$$\frac{d[A_{tot}]}{dt} = \alpha_A + \frac{\beta_A [A_2]^n}{1 + [A_2]^n} - [A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = (\alpha_B/\alpha_A) * \alpha_A + \frac{(\beta_B/\beta_A) * \beta_A [B_2]^n}{(K_{MB}/K_{MA})^n + [B_2]^n} - \delta_B/\delta_A [B_{tot}]$$

$$\frac{d[C_{tot}]}{dt} = (\alpha_C/\alpha_A) * \alpha_A + \frac{(\beta_C/\beta_A) * \beta_A [C_2]^n}{(K_{MC}/K_{MA})^n + [C_2]^n} - \delta_C/\delta_A [C_{tot}]$$

To further simplify these expressions we define additional parameter ratios, $r = \alpha_B/\alpha_A$, $r_2 = \alpha_C/\alpha_A$, $m = \beta_B/\beta_A$, $m_2 = \beta_C/\beta_A$, $\kappa = K_{MB}/K_{MA}$, $\kappa_2 = K_{MC}/K_{MA}$, and $\gamma = \delta_B/\delta_A$, $\gamma_2 = \delta_C/\delta_A$. We also let $\alpha = \alpha_A$, and $\beta = \beta_A$ for notational simplicity. With these definitions, we obtain the ODEs for protein production and degradation in the asymmetric MultiFate model:

$$\frac{d[A_{tot}]}{dt} = \alpha + \frac{\beta [A_2]^n}{1 + [A_2]^n} - [A_{tot}]$$

$$\frac{d[B_{tot}]}{dt} = r\alpha + \frac{m\beta [B_2]^n}{\kappa^n + [B_2]^n} - \gamma [B_{tot}]$$

$$\frac{d[C_{tot}]}{dt} = r_2\alpha + \frac{m_2\beta [C_2]^n}{\kappa_2^n + [C_2]^n} - \gamma_2 [C_{tot}]$$

with the same expressions of the $[A_2]$, $[B_2]$ and $[C_2]$ in terms of $[A_{tot}]$, $[B_{tot}]$ and $[C_{tot}]$ shown in the previous section.

We now have additional parameters that represent different types of asymmetries: r (and r_2) represents the ratio of basal protein production rates, m (and m_2) represents the ratio

of maximal protein production rates by self-activation, κ (and κ_2) represents the ratio of homodimer concentrations for half-maximal activation, and γ (and γ_2) represents the ratio of protein half-lifes. The symmetric MultiFate model is then a special case where all asymmetry parameters equal to 1.

MultiFate model with mRNA and protein dimerization dynamics

The treatment above lumps the processes of mRNA transcription and protein translation together into a single gene expression step and uses a steady-state approximation for dimerization dynamics. This model works accurately to predict the number and locations of stable fixed points. However, to capture the dynamics of cells during bifurcation and state-switching events, an expanded model that includes both mRNA and protein dimerization dynamics is required.

To incorporate mRNA dynamics into the model, we assume TF A, TF B and TF C mRNAs, denoted $[a]$, $[b]$ and $[c]$, are produced at a total rate equal to their basal transcription rate k_1 plus a homodimer-dependent transcriptional activation rate, which follows a Hill function of corresponding homodimer concentration, $[A_2]$, $[B_2]$ and $[C_2]$, with maximal rate k_2, Hill coefficient n, and half-maximal activation at a homodimer concentration of K_M. Each mRNA species is removed at a total rate δ_{mRNA}. For generality, we also allow asymmetry parameters, with r (or r_2) representing the ratio between TF A and TF B (or TF C) basal transcription rates ($r = k_{1,B}/k_{1,A}$ or $r_2 = k_{1,C}/k_{1,A}$), m representing the ratio between TF A and TF B (or TF C) maximal

rates ($m = k_{2,B}/k_{2,A}$ or $m = k_{2,C}/k_{2,A}$) , and κ (or κ_2) representing the ratio between TF A and TF B (or TF C) half-maximal homodimer concentrations ($\kappa = K_{MB}/K_{MA}$or K_{MC}/K_{MA}). We can then write:

$$\frac{d[a]}{dt} = k_1 + \frac{k_2[A_2]^n}{K_M{}^n + [A_2]^n} - \delta_{mRNA}[a]$$

$$\frac{d[b]}{dt} = rk_1 + \frac{mk_2[B_2]^n}{(\kappa K_M)^n + [B_2]^n} - \delta_{mRNA}[b]$$

$$\frac{d[c]}{dt} = r_2 k_1 + \frac{m_2 k_2 [C_2]^n}{(\kappa_2 K_M)^n + [C_2]^n} - \delta_{mRNA}[c]$$

where we let $k_1 = k_{1,A}$, $k_2 = k_{2,A}$ and $K_M = K_{MA}$ for notational simplicity.

Next, we describe the dynamics of TF A, TF B and TF C proteins in different dimerization forms with ODEs. For monomers of TF A, TF B and TF C, denoted $[A]$, $[B]$ and $[C]$, each equation consists of terms describing translation, protein removal, monomer association, dimer dissociation or conversion due to degradation of one of the constituent monomers. Here k_p is the translation rate, δ is the protein removal rate, k_{on} is the monomer association rate and k_{off} is the dimer dissociation rate. $[AB]$, $[AC]$ and $[BC]$ denotes the concentration of AB, AC and BC heterodimers.

The asymmetry parameter γ (or γ_2) represents the ratio of TF A and TF B (or TF C) removal rates ($\gamma = \delta_B/\delta_A$ or $\gamma_2 = \delta_C/\delta_A$), and we let $\delta = \delta_A$for notational simplicity:

$$\frac{d[A]}{dt} = k_p[a] - \delta[A] - 2k_{on}([A]^2 + [A][B] + [A][C])$$

$$+ k_{off}(2[A_2] + [AB] + [AC]) + \gamma\delta[AB] + \gamma_2\delta[AC]$$

$$\frac{d[B]}{dt} = k_p[b] - \gamma\delta[B] - 2k_{on}([B]^2 + [B][A] + [B][C])$$

$$+ k_{off}(2[B_2] + [AB] + [BC]) + \delta[AB] + \gamma_2\delta[BC]$$

$$\frac{d[C]}{dt} = k_p[c] - \gamma_2\delta[C] - 2k_{on}([C]^2 + [C][A] + [C][B])$$

$$+ k_{off}(2[C_2] + [AC] + [BC]) + \delta[AC] + \gamma\delta[BC]$$

For dimers, each equation consists of terms for association, dissociation, and removal:

$$\frac{d[A_2]}{dt} = k_{on}[A]^2 - k_{off}[A_2] - \delta[A_2]$$

$$\frac{d[B_2]}{dt} = k_{on}[B]^2 - k_{off}[B_2] - \gamma\delta[B_2]$$

$$\frac{d[C_2]}{dt} = k_{on}[C]^2 - k_{off}[C_2] - \gamma_2\delta[C_2]$$

$$\frac{d[AB]}{dt} = 2k_{on}[A][B] - k_{off}[AB] - \delta[AB] - \gamma\delta[AB]$$

$$\frac{d[AC]}{dt} = 2k_{on}[A][C] - k_{off}[AC] - \delta[AC] - \gamma_2\delta[AC]$$

$$\frac{d[BC]}{dt} = 2k_{on}[B][C] - k_{off}[BC] - \gamma\delta[BC] - \gamma_2\delta[BC]$$

Stochastic modeling of MultiFate circuits

To simulate the dynamics of the MultiFate-2 system during state-switching (Fig. S2.3) and bifurcation (Fig. 2.4A) events, to obtain best fitted asymmetry parameters for

different MultiFate lines (Fig. S2.15) and to test the robustness of MultiFate against intrinsic biological noise (Figs. 2.5D, S2.22 and S2.23), we constructed a stochastic model based on the same reactions represented by the ODEs in the above MultiFate-2 model with mRNA and protein dimerization dynamics. Molecular reactions and their propensities for Gillespie simulation (*41*) are listed in Table S2.3. All terms have a concentration unit of molecule number per cell (either mRNA or protein), and a time unit of hour. We performed Gillespie simulations using the biocircuits Python package (https://pypi.org/project/biocircuits/) with physiologically reasonable parameters (see below).

Physiologically reasonable parameter regimes

From existing literature and experimental measurements performed in this study, we estimated the physiologically reasonable regime for each dimensionalized parameter (K_M, δ, α, β, K_d, n, k_1, k_2, δ_{mRNA}, k_p, k_{on},k_{off}). Estimated values for these parameters are summarized in Table S2.4.

Since some measurement data are in the unit of concentration, while others are in the unit of molecules per cell, we first estimated the number of molecules equivalent to 1 nM in a CHO-K1 cell. The diameter of a CHO cell is ~14 μm (*56*), from which we can calculate the cell volume to be around 1.4×10^{-12} L (assuming it to be a sphere). With this rough approximation, **1 nM = 1 nM $\times$ 1.4×10^{-12} L $\times$ 6×10^{23} molecules/mol ≈ 800 molecules per CHO cell.** Below, we used this value to convert between molecules per cell and molarity.

The concentration for half-maximal activation, K_M, of our ZF activator homodimer, is used to rescale all concentrations in the non-dimensionalized MultiFate model (see above). We could not find a direct *in vivo* measurement of this value in the literature. However, the K_M of a monomeric ZF activator with a 3-finger Zif268 ZF domain was estimated to be ~600 nM (assuming the volume of yeast cells to be 40 μm^3) (*4*). Another *in vitro* study showed that by linking two Zif268 ZF domain with a linker, the resulting 6-finger ZF domain could bind to a 18bp DNA target site almost 70-fold stronger than a single Zif268 domain binding to a 9bp target site (*57*). Although *in vivo* and the *in vitro* K_M can differ by several orders of magnitudes due to the interactions of proteins with genomic background, the fold changes in K_M are in reasonable agreement between the *in vivo* and the *in vitro* results (*58*). Therefore, we estimated that an activator with a 6-finger ZF domain has a K_M of 600 nM / 70 ≈ 8 nM. Our ZF transcription factor homodimer should bind to 18bp target DNA site in a similar fashion with how 6-finger ZF domain does, thus we estimated our K_M to be comparable to, or (due to a more complicated genomic background in mammalian cells), slightly larger than the range of 8 – 20 nM. Based on this reasoning, **we used a K_M= 10 nM = 8000 molecules/cell in the model.**

Next, we estimated parameters related to protein production and removal dynamics. We used a protein removal rate $\delta = 0.1\ hr^{-1}$ for stable proteins in our system, based on an *in vivo* measurement of proteome half-life dynamics in living human cells (*59*). Our engineered ZF transcription factors have DHFR domains at their C-terminus, whose protein removal rate is controlled by TMP concentration. The dynamic range of this regulation is at least 20 fold (*39*). Therefore, we estimated the range of δto be between

0.1 hr^{-1} and 2 hr^{-1}, under saturating TMP condition and no TMP condition, respectively. **In the model, we used δ = 0.1 hr^{-1} for the "High TMP" condition, and δ = 0.2 hr^{-1} for the "Low TMP" condition.**

We next estimated k_2, δ_{mRNA} and k_p, which together are critical for establishing the levels and dynamics of mRNA and protein. In mammalian cells, average transcription rates were estimated to be ~2 mRNA/(gene•hr) (*60*). Based on previous stable cell line construction using PiggyBac in our lab (*61*), we estimated that a total of 50–100 gene cassettes were integrated during the construction of MultiFate-2, with 25–50 copies integrated each, for TF A and TF B. Maximal transcription rate k_2 was then estimated to be 50–100 mRNA/(cell•hr), and **we used an intermediate value k_2 = 80 mRNA/(cell•hr)**. For typical mRNA half-life, different studies have provided diverse values, ranging from 50 minutes to 9 hours (*60*, *62*). This corresponds to a mRNA removal rate δ_{mRNA} ranging from ln(2) / (9 hr) – ln(2) / (50 min) ≈ 0.077 – 0.83 hr^{-1}. Since these studies did not take mRNA dilution from cell division into consideration, **we used a value of δ_{mRNA} = 0.7 hr^{-1},** closer to the upper bound of the estimated range, in our stochastic model. For protein translation rate, **we used a value of k_p = 140 proteins/(mRNA•hr)** based on one of the above studies (*60*).

The value of βcan be estimated from the above parameters. Since the mRNA removal rate is much higher than the protein removal rate, we assumed that mRNA dynamics is approximately at steady state on the timescale of δ^{-1}. This assumption enabled us to estimate the **maximal protein production rate in the Hill function as**

$\beta = k_2 \times k_p/\delta_{mRNA}$ **= 16000 proteins/(cell•hr) = 20 nM/hr.** In the experiment, we observed that there was a fluorescence expression difference of around 25-50 fold between ON and OFF states (Fig. 2.3B). (The estimate of the OFF level is not limited by autofluorescence). Since expression in the OFF state comes from basal transcription, we estimated the basal transcription rate k_1 to be 25–50 fold smaller than k_2, giving a range of 1.152–4.608 mRNA/hr. From this, **we used an intermediate value of** k_1 **= 3.2 mRNA/(cell•hr)**. **Similarly, we can estimate the basal protein production rate** $\alpha = k_1 \times k_p/\delta_{mRNA}$ **= 640 proteins/(cell•hr) = 0.8 nM/hr.**

The final parameter related to protein production dynamics is the Hill coefficient n. Different studies obtained very different measurements of transcriptional Hill coefficients (*4*, *34*, *63*), ranging from 1–3.6. **Here we used a modest Hill coefficient of** n **= 1.5.**

Finally, we estimated parameters related to protein dimerization. The apparent dissociation constant K_d of FKBP homodimerization domain should depend on AP1903 concentrations. However, we could not find a direct measurement of this dependency. Therefore, we compared FKBP with another homodimerization domain GCN4 that we used in this study (Fig. 2.2A), which was shown to have a $K_d$of 10–20 nM (*64*). ZF transcription factors with FKBP (Fig. 2.2B) more strongly activated the reporter in 100 nM AP1903 media than ZF transcription factor with GCN4 did (Fig. S2.4B, BCRZF). Based on this observation, we reasoned that in media containing 100 nM AP1903, the apparent dissociation constant $K_d \leq 10$ nM. **We therefore used an estimate of** K_d **= 10 nM = 8000 molecules/cell in the model. For monomer association rate** k_{on}**, we used**

an intermediate value in the range of diffusion-limited association rates of k_{on} = 4x10^{5}/(M•s) = 1.8x10^{-3}/(protein•hr) (*65*)**.** These two values together produce a **dimer dissociation rate $k_{off} = K_d \times k_{on}$ = 14.4 hr^{-1}.**

Relationship between transcription factor concentrations and their co-expressed fluorescence proteins

The MultiFate models use the total concentrations of transcription factors as variables, whereas experimental MultiFate systems have the co-expressed fluorescent proteins as readouts. To understand the relationship between transcription factor concentrations ([A_{tot}] and [B_{tot}]), and their co-expressed fluorescent proteins, denoted FPA and FPB, we incorporated equations describing the dynamics of immature fluorescent proteins ([FPA_{im}] and [FPB_{im}]) and mature fluorescent proteins ([FPA_{m}] and [FPB_{m}]) into the MultiFate-2 model. Since fluorescent proteins are co-expressed with transcription factors, the production term of fluorescent protein have a similar form $\alpha + \frac{\beta[A_2]^n}{1+[A_2]^n}$, scaled by translational efficiency of IRES, denoted I_{eff}. Once produced, each immature fluorescent protein matures at a rate k_{mat}. Finally, either immature or mature fluorescent protein degrades and is diluted at a total rate δ_{FP}. We can then add a set of ODEs to the MultiFate-2 model:

$$\frac{d[FPA_{im}]}{dt} = I_{eff}(\alpha + \frac{\beta[A_2]^n}{1+[A_2]^n}) - k_{matA}[FPA_{im}] - \delta_{FPA}[FPA_{im}]$$

$$\frac{d[FPA_{m}]}{dt} = k_{matA}[FPA_{im}] - \delta_{FPA}[FPA_{m}]$$

with a similar set of equations for [FPB_{im}] and [FPB_{m}].

We used an I_{eff} = 0.5 in the model based on (*66*). Maturation rate for mCherry (k_{matA}) and mCitrine (k_{matB}) were calculated to be 1.12 hr^{-1} and 4.62 hr^{-1}, respectively, based on their estimated maturation time (*67*). In experimental MultiFate systems, all fluorescence proteins are fused with a PEST degron, which has a half-life of 2-6.5 hours (*31*, *68*), and we used a $\delta_{FP} = 0.35$ hr^{-1} based on a 2-hour half-life. All rates are then rescaled by the degradation rate of transcription factors (δ) to obtain a non-dimensionalized model. We also incorporated fluorescent protein translation, maturation and degradation into the MultiFate stochastic model and listed their propensities for Gillespie algorithm in Table S4. These models were used to simulate the dynamics of transcription factor concentrations and fluorescence readouts in the same cells.

We first used the stochastic model to simulate the relationship between transcription factors concentration and their mature fluorescent proteins for a single self-activation module (Fig. S2.6A). We chose 5 different transcription factor half-lives to obtain different distributions of transcription factor concentrations at steady states. While transcription factor concentrations vary in a wide range among these 5 conditions, fluorescent readouts show a strong bimodal distribution. This shows that positive autoregulation causes each transcription factor to express in a roughly binary (high or low) fashion: when transcription factor concentration is higher than the ‘self-activation threshold’ (defined by TF concentration where [Homodimers] = 1), activating transcription factor homodimers drive gene expression to ‘high’ state. Fluorescent proteins quickly saturate in the ‘high’ state, as shown by overlapping fluorescent protein

distributions in the ‘high’ expression state (Fig. S2.6A, middle right), whereas transcription factor concentrations are additionally affected by the protein half-life (Fig. S2.6A, middle left). This relationship between transcription factor concentrations and fluorescent readouts is further shown in the 2D scatter plot (Fig. S2.6A, right).

Since each transcription factor in a MultiFate circuit positively autoregulates itself, each transcription factor would express in a roughly binary fashion. Any stable state is thus a combination of these binary expression states, allowing us to distinguish them through fluorescence readouts.

To test this, we simulated the single-cell dynamics of the MultiFate-2 circuit in either type II tristable regime or bistable regime (Fig. S2.6B). In both regimes, fluorescence readouts are well separated into distinct clusters. Critically, each cluster can be unambiguously assigned to its underlying state. Consistent with results from a single self-activation module (Fig. S2.6A), although the TF A concentrations in A-only state (or TF B concentrations in B-only state) differ by more than 2 folds between the tristable regime and bistable regime, the mature mCherry (or mCitrine) only differ by about 10% and are almost indistinguishable on log scale. This matches with our experimental observations (Figs. S2.9 to S2.11). Together, these simulation results show that fluorescent reporters are sufficient to unambiguously identify the underlying states.

Finally, we asked how well the fluorescence readouts track the dynamics of cell state transition. To test this, we simulated MultiFate-2 cells switching from A-only state to B-only state in the bistable regime (similar to Fig. S2.3) with fluorescent proteins of different maturation times and half-lives. In particular, we measured the delay in time

between when transcription factor concentrations cross the state boundary (from $[A_{tot}] > [B_{tot}]$ to $[A_{tot}] < [B_{tot}]$) and when mature fluorescent proteins cross the state boundary (from $[FPA_m] > [FPB_m]$ to $[FPA_m] < [FPB_m]$) (Fig. S2.6C, left). Longer maturation time and longer fluorescent protein half-life both increase the time delay (Fig. S2.6C, right). Based on this, we chose three fluorescent proteins that have short maturation times (mCherry 37 minutes, mCitrine 9 minutes, mTurquoise2 34 minutes) (*67*), and fused a PEST degron to their C-terminus to shorten their half-life to 2-6.5 hours (*31*, *68*). With these modifications, we estimated the time delay between fluorescent readouts and transcription factor dynamics to be less than 6 hours. This time delay is small when compared to total switching time in the experiments (Fig. 2.4B), which spans several days.

Robustness of MultiFate circuit against intrinsic noise

In both flow cytometry plots (Figs. S2.9 to S2.11) and time-lapse images (Fig. S2.13), we found a small number of cells spontaneously escaped from their original states due to biological noise. While these cells were rare, they led us to ask about the robustness of MultiFate against biological noise, especially intrinsic noise resulting from stochasticity of chemical reactions such as transcription, translation and degradation (*43*).

We used the Gillespie algorithm (*41*) to simulate MultiFate circuits with intrinsic noise. We hypothesized that different stable steady states may have different robustness against these intrinsic noises. Cells in states with smaller attractor basins may be more likely to spontaneously switch to other states due to random concentration fluctuations introduced by intrinsic noise. To test this, we chose a MultiFate-3 type I quadrastable regime, in

which the OFF state has a much smaller attractor basin compared with other states (Fig. S2.22A, left). As expected, many cells from the OFF state spontaneously turn on one of the transcription factors to switch to one of the other three states (Fig. S2.22A, center), while all cells from the B-only state (and similarly for the A-only and the C-only state) remain in their original state (Fig. S2.22A, center). We quantify the robustness by the fraction of cells not changing states at the end of simulations, denoted as "robustness score". For this MultiFate-3 type I quadrastable regime, the OFF state has a smaller attractor basin and a lower robustness score than the other 3 states.

We next systematically test the relationship between attractor basin size and robustness score for a set of MultiFate-2 and MultiFate-3 regimes (Fig. S2.22B). We found that robustness score indeed is positively correlated with attractor basin size. However, there is no clear cutoff on the size of the attractor basin to separate robust stable states (robustness score = 1) and non-robust stable states, suggesting that robustness against intrinsic noise may be affected by other factors, such as promoter leakiness and ultrasensitivity. Therefore, we directly used the robustness score to determine whether a fixed point is robust against biological noise in Figs. 2.5D and S2.23.

Source of ultrasensitivity in MultiFate circuit

To generate multistability, a circuit should have both positive feedback and some levels of ultrasensitivity (i.e. effective Hill exponent greater than 1) (*69*, *70*). Although transcriptional activation can exhibit some ultrasensitivity in mammalian cells (represented by the Hill coefficient of $n = 1.5$ above) (*71*, *72*), parameter screening revealed that MultiFate generates multistability even when $n = 1$ (no ultrasensitivity

from transcriptional activation). This provokes the question of where the additional ultrasensitivity comes from in the MultiFate circuit. Two features of the MultiFate circuit could provide this additional ultrasensitivity. First, transcription factors homodimerize to self-activate, and such cooperativity has been shown to introduce ultrasensitivity (*73*, *74*). Indeed, homodimerization results in a $[A_{tot}]^2$ (or $[B_{tot}]^2$) term in the numerator of the expression for $[A_2]$ (or $[B_2]$) in Box 2.1, which we write again here for convenience:

$$[A_2] = \frac{2[A_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d^2 + 8([A_{tot}] + [B_{tot}])K_d}}$$

$$[B_2] = \frac{2[B_{tot}]^2}{K_d + 4([A_{tot}] + [B_{tot}]) + \sqrt{K_d^2 + 8([A_{tot}] + [B_{tot}])K_d}}$$

Its contribution to ultrasensitivity depends on the dimerization dissociation constant K_d. When homodimerization is strong (small K_d), the $[A_{tot}] + [B_{tot}]$ term dominates in the denominator, which cancels with the quadratic term $[A_{tot}]^2$ or $[B_{tot}]^2$ in the numerator. This makes the expression more linear, thus reducing the ultrasensitivity by homodimerization. Conversely, when dimerization is weak (large K_d), the K_d^2 term dominates in the denominator, which makes the expression more quadratic and increases the ultrasensitivity by homodimerization.

A second source of ultrasensitivity comes from mutual inhibition through heterodimerization, a prevalent feature in biology also known as molecular titration, which has been shown to introduce ultrasensitivity (*69*, *73*). Here, opposite to the case

with homodimerization, strong heterodimerization (small K_d) increases the ultrasensitivity introduced through molecular titration (*73*). Together, additional ultrasensitivity comes mainly from cooperativity through homodimerization when K_d is large, and mainly from molecular titration through heterodimerization when K_d is small (note that K_d values for homodimerization and heterodimerization are the same, since we are using the same dimerization domain for all transcription factors). This explains why the MultiFate circuit generates multistability in a wide K_d range (Figs. S2.1 and S2.2).

Modulating basal expression by modifying synthetic promoter sequences

To obtain the desired type II tristability in MultiFate-2 circuit, parameter screening (Fig. S2.1B) revealed that we should both avoid regimes with too high basal expression, where only A+B state is stable, and regimes with too low basal expression, where OFF state is stable. When building MultiFate self-activation modules, we found that original promoter basal expression is too low (Fig. S2.24, construct 1), as shown by low level of spontaneous self-activation upon the addition of AP1903+TMP. Therefore, we sought to increase the basal promoter expression by modifying promoter sequences. When characterizing different ZF transcription factors, we had an incidental finding that promoters containing the 9bp binding site GACGCTGCT for 42ZF (*32*) have higher basal expression. Basal promoter expression thus can be modulated by introducing different numbers of GACGCTGCT motifs at promoter regions (Fig. S2.24), and we introduced multiple repeats of this motif into final MultiFate constructs (Table S2.2).

Translating MultiFate circuits into other cell types

Many components used in MultiFate circuits generally work across different cell types, including transcriptional activation domains and protein dimerization domains. For example, the transcriptional activation domain VP16 has been shown to work in multiple cell lines (*33*). Some components in MultiFate may require modification when moving into a new context: First, some zinc finger DNA-binding domains were first developed in yeast (*32*). While the original zinc finger domains still work in this study, more recently optimized synthetic zinc finger domains (*75*) might be more desirable for translating MultiFate into human cells. Second, basal promoter expression (transcriptional leakiness) might differ among cell lines and genomic contexts. Improvements in the ability to modulate basal expression would help engineer MultiFate circuits in additional cell contexts.

The general strategy to engineer MultiFate in CHO cells should be applicable to other cell types. The basic module of MultiFate is the dimer-dependent self-activation circuit (Fig. 2.2C). In our experience, its performance directly affects the behaviors of the final MultiFate circuit. One should first test whether the self-activation circuit can robustly sustain its own expression, and whether the self-activation is dimer-dependent, when translating MultiFate in a new cell type. If self-activation cannot robustly sustain its own expression, it is possible that protein production rate (β in the model) is not high enough in the new cell type. In that case, one should consider using stronger transcriptional activation domains such as VP64 and p65 to boost the mRNA transcription. Alternatively, since β in the non-dimensionalized model is rescaled by K_M (homodimer

concentration for half-maximal activation), a smaller K_M results in larger rescaled β. To achieve a smaller K_M, one could use better zinc finger domains such as the new optimized / humanized zinc finger transcription factors mentioned above (*75*), which should increase binding affinity of the homodimers to the DNA and thus decrease the K_M. If self-activation is not dimer-dependent, one should consider modifying zinc fingers using the same mutation strategy of Figs. 2.2A and S2.4B.

Once the self-activation module works, one could follow the workflow in Fig. S2.7 to generate MultiFate cells. To make MultiFate-2 cells, one can first stably integrate two different self-activation modules into the desired cell types, then select for cells that can maintain a stable double-positive state using FACS or other cell sorting methods. This usually results in desired MultiFate-2 cells, since the MultiFate-2 model (Fig. S2.1B) shows that cells with a stable double-positive state can generate diverse multistability regimes. Similarly, to make MultiFate-3 cells, one can stably integrate a third self-activation module into the existing MultiFate-2 cells, then select for cells that can maintain a stable triple-positive state.

2.11 Supplementary Tables

Table S2.1 List of physiologically reasonable parameter regimes.

Parameters	Model	Estimated values	References
K_M	Deterministic and stochastic	10 nM or 8000 molecules/cell	(*4*, *57*, *58*)
δ	Deterministic and stochastic	0.1 hr^{-1} for "High TMP" condition; 0.2 hr^{-1} for "Low TMP" condition	(*39*, *59*)
n	Deterministic and stochastic	1.5	(*4*, *34*, *63*)
α	Deterministic	0.8 nM/hr	Derived
β	Deterministic	20 nM/hr	Derived
K_d	Deterministic	10 nM	(*64*) and current study
k_1	Stochastic	3.2 mRNA/hr	Current study
k_2	Stochastic	80 mRNA/hr	(*60*) and current study
δ_{mRNA}	Stochastic	0.7 hr^{-1}	(*60*, *62*)
k_p	Stochastic	140 proteins/(mRNA*hr)	(*60*)
k_{on}	Stochastic	$1.8x10^{-3}$/(protein*hr)	(*65*)
k_{off}	Stochastic	14.4 hr^{-1}	Derived
I_{eff}	Deterministic and stochastic	0.5	(*66*)
k_{mat}	Deterministic and stochastic	1.12 hr^{-1} for mCherry; 4.62 hr^{-1} for mCitrine	(*67*)
δ_{FP}	Deterministic and stochastic	0.35 hr^{-1}	(*68*)
Note: 1 nM in a CHO cell is equivalent to 800 molecules.			

Table S2.2 List of plasmids used in this study and their use in the figures.

Index	Construct name	Usage in this study	Figures or MultiFate lines
MF01	PB-2x(ErbB2bs_ErbB2bs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B,C
MF02	PB-2x(37bs_37bs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B,C
MF03	PB-2x(42bs_42bs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B,C
MF04	PB-2x(92bs_92bs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B
MF05	PB-2x(97bs_97bs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B
MF06	PB-2x(BCRbs_BCRbs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B,C
MF07	PB-2x(HIVbs_HIVbs)-TATA-3xNLS-Citrine-BGHpA	Reporter	2A, S4A,B
MF08	PB-CAG-ErbB2ZFWT-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF09	PB-CAG-ErbB2ZFWT-GCN4-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF10	PB-CAG-ErbB2ZFR39A-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF11	PB-CAG-ErbB2ZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF12	PB-CAG-ErbB2ZFR2AR39A-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF13	PB-CAG-ErbB2ZFR2AR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	2A, S4A, C
MF14	PB-CAG-ErbB2ZFR2AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF15	PB-CAG-ErbB2ZFR2AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	2A, S4A
MF16	PB-CAG-FKBP12F36V-BCRZFR39A-VP48-mCherry-BGHpA	Trancription factors	2B
MF17	PB-CAG-37ZFWT-VP48-mCherry-BGHpA	Trancription factors	S4B
MF18	PB-CAG-37ZFWT-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF19	PB-CAG-37ZFR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF20	PB-CAG-37ZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF21	PB-CAG-37ZFR2AR39A-VP48-mCherry-BGHpA	Trancription factors	S4B

MF22	PB-CAG-37ZFR2AR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF23	PB-CAG-37ZFR2AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF24	PB-CAG-37ZFR2AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF25	PB-CAG-37ZFR2AR11AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF26	PB-CAG-37ZFR2AR11AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B, C
MF27	PB-CAG-42ZFR2AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF28	PB-CAG-42ZFR2AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B, C
MF29	PB-CAG-92ZFWT-VP48-mCherry-BGHpA	Trancription factors	S4B
MF30	PB-CAG-92ZFWT-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF31	PB-CAG-92ZFR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF32	PB-CAG-92ZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF33	PB-CAG-92ZFR2AR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF34	PB-CAG-92ZFR2AR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF35	PB-CAG-92ZFR2AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF36	PB-CAG-92ZFR2AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF37	PB-CAG-97ZFWT-VP48-mCherry-BGHpA	Trancription factors	S4B
MF38	PB-CAG-97ZFWT-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF39	PB-CAG-97ZFR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF40	PB-CAG-97ZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF41	PB-CAG-97ZFR2AR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF42	PB-CAG-97ZFR2AR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B

MF43	PB-CAG-BCRZF-VP48-mCherry-BGHpA	Trancription factors	S4B
MF44	PB-CAG-BCRZF-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF45	PB-CAG-BCRZFR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF46	PB-CAG-BCRZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B, C
MF47	PB-CAG-HIV1ZFWT-VP48-mCherry-BGHpA	Trancription factors	S4B
MF48	PB-CAG-HIV1ZFWT-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF49	PB-CAG-HIV1ZFR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF50	PB-CAG-HIV1ZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF51	PB-CAG-HIV1ZFR2AR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF52	PB-CAG-HIV1ZFR2AR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF53	PB-CAG-HIV1ZFR2AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF54	PB-CAG-HIV1ZFR2AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF55	PB-CAG-HIV2ZFWT-VP48-mCherry-BGHpA	Trancription factors	S4B
MF56	PB-CAG-HIV2ZFWT-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF57	PB-CAG-HIV2ZFR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF58	PB-CAG-HIV2ZFR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF59	PB-CAG-HIV2ZFR2AR39A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF60	PB-CAG-HIV2ZFR2AR39A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF61	PB-CAG-HIV2ZFR2AR39AR67A-VP48-mCherry-BGHpA	Trancription factors	S4B
MF62	PB-CAG-HIV2ZFR2AR39AR67A-GCN4-VP48-mCherry-BGHpA	Trancription factors	S4B
MF63	PB-TRE3G-6x42bs-6x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-DHFR-IRES-m	Self-activation construct	2C

	Citrine-PEST-BGHpA		
MF64	PB-TRE3G-6x42bs-6x(37bs_37bs)-miniCMV-NLS-FKBP12F36V-37ZFR2AR11AR39AR67A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	S5A, MultiFate-2.1, MultiFate-2.3
MF65	PB-TRE3G-6x42bs-6x(92bs_92bs)-miniCMV-NLS-FKBP12F36V-92ZFR2AR39AR67A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	S5A
MF66	PB-TRE3G-6x42bs-6x(97bs_97bs)-miniCMV-NLS-FKBP12F36V-97ZFR39A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	S5A
MF67	PB-TRE3G-6x42bs-6x(ErbB2bs_ErbB2bs)-miniCMV-NLS-FKBP12F36V-ErbB2ZFR2AR39A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	S5A
MF68	PB-TRE3G-6x42bs-6x(HIVbs_HIVbs)-miniCMV-NLS-FKBP12F36V-HIV1ZFR2AR39A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	S5A
MF69	PB-TRE3G-6x42bs-6x(HIVbs_HIVbs)-miniCMV-NLS-FKBP12F36V-HIV2ZFR2AR39AR67A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	S5A
MF70	PB-TRE3G-6x(42bs_42bs)-miniPromo-42ZFR2AR39AR67A-GCN4-VP48-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	2D, S5B
MF71	PB-TRE3G-6x(42bs_42bs)-miniPromo-FKBP12F36V-42ZFR2AR39AR67A-VP48-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	2D, S5B
MF72	PB-CAG-IRES-mCherry-PEST-BGHpA (Control)	Protein perturbations	2D, S5B
MF73	PB-CAG-BCRZFR39A-GCN4-VP48-IRES-mCherry-PEST-BGHpA	Protein perturbations	2D, S5B
MF74	PB-CAG-FKBP12F36V-BCRZFR39A-VP48-IRES-mCherry-PEST-BGHpA	Protein perturbations	2D, S5B
MF75	PB-CAG-BCRZFR39A-GCN4-IRES-mCherry-PEST-BGHpA	Protein perturbations	S5B
MF76	PB-CAG-FKBP12F36V-IRES-mCherry-PEST-BGHpA	Protein perturbations	S5B
MF77	PB-CAG-BCRZFR39A-VP48-IRES-mCherry-PEST-BGHpA	Protein perturbations	2D, S5B
MF78	PB-CAG-BCRZFR39A-IRES-mCherry-PEST-BGHpA	Protein perturbations	S5B
MF79	PB-CAG-GCN4-IRES-mCherry-PEST-BGHpA	Protein perturbations	S5B
MF80	PB-CAG-VP48-IRES-mCherry-PEST-BGHpA	Protein	S5B

		perturbations	
MF81	PB-TRE3G-12x42bs-6x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-DHFR-IRES-mCherry-PEST-BGHpA	Self-activation construct	MultiFate-2.1
MF82	PB-TRE3G-12x42bs-10x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-DHFR-IRES-mCherry-PEST-BGHpA	Self-activation construct	MultiFate-2.2, MultiFate-3
MF83	PB-TRE3G-6x42bs-10x(37bs_37bs)-miniCMV-NLS-FKBP12F36V-37ZFR2AR11AR39AR67A-VP16-NLS-DHFR-IRES-mCitrine-PEST-BGHpA	Self-activation construct	MultiFate-2.2, MultiFate-3
MF84	PB-14xUAS-6x42bs-6x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-DHFR-IRES-mCherry-PEST-BGHpA	Self-activation construct	MultiFate-2.3
MF85	PB-14xUAS-12x42bs-10x(ErbB2bs_ErbB2bs)-miniCMV-NLS-FKBP12F36V-ErbB2ZFR2AR39A-VP16-NLS-DHFR-IRES-mTurquoise2-PEST-BGHpA	Self-activation construct	MultiFate-3
MF86	PB-EF1α-Tet3G-BGHpA	Inducible system	2C, S5A
MF87	PB-EF1α-Tet3G-P2A-ERT2-Gal4-BGHpA	Inducible system	All MultiFate cells
MF88	PB-TRE3G-6x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-FLAG-DHFR-IRES-mCherry-PEST-BGHpA	Self-activation construct	S24
MF89	PB-TRE3G-4x42bs-6x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-FLAG-DHFR-IRES-mCherry-PEST-BGHpA	Self-activation construct	S24
MF90	PB-TRE3G-6x42bs-6x(BCRbs_BCRbs)-miniCMV-NLS-FKBP12F36V-BCRZFR39A-VP16-NLS-FLAG-DHFR-IRES-mCherry-PEST-BGHpA	Self-activation construct	S24

Note:
PB = PiggyBac backbone; TRE3G = Tet3G binding site; UAS = ERT2-Gal4 binding site;
TATA, miniCMV, miniPromo are three different minimal promoters;
VP48, VP16 are two different transcriptional activation domains;
CAG = the constitutive CAG promoter (*30*);
EF1α = the constitutive EF1α promoter;
NLS = nuclear localization sequence;
IRES = internal ribosome entry site;
BGHpA = bovine growth hormone polyadenylation signal;
PEST = constitutive signal peptide for protein degradation (*31*);
42bs = both the 42ZF binding site and 9bp motif that increase promoter leakiness;
ZFbs_ZFbs = 18bp tandem ZF binding site pairs;
Construct maps in GenBank format are available at data.caltech.edu/records/1882.

Table S2.3 List of stable cell lines constructed for this study and their use in the figures.

Cell lines	Parental cells	Polyclonal or monoclonal	Integrated constructs	Figures	Additional procedures to screen monoclones
Tet3G-expressing CHO-K1	CHO-K1	Polyclonal	MF86	2C, 2D, S5	N/A
ERT2-Gal4-T2A-Tet3G expressing CHO-K1	CHO-K1	Polyclonal	MF87	3-5, S6-17	N/A
FKBP-BCRZFR39A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF63	2C	N/A
FKBP-37ZFR2AR11AR39AR67A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF64	S5A	N/A
FKBP-92ZFR2AR39AR67A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF65	S5A	N/A
FKBP-97ZFR39A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF66	S5A	N/A
FKBP-ErbB2ZFR2AR39A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF67	S5A	N/A
FKBP-HIV1ZFR2AR39A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF68	S5A	N/A
FKBP-HIV2ZFR2AR39AR67A-VP48-DHFR self-activation	Tet3G-expressing CHO-K1	Polyclonal	MF69	S5A	N/A
FKBP-BCRZFR39A-VP16-DHFR self-activation (with no 42bs in promoter)	CHO-K1	Polyclonal	MF88	S24	N/A
FKBP-BCRZFR39A-VP16-DHFR self-activation (with 4x 42bs in	CHO-K1	Polyclonal	MF89	S24	N/A

promoter)					
FKBP-BCRZFR 39A-VP16-DHF R self-activation (with 6x 42bs in promoter)	CHO-K1	Polyclonal	MF90	S24	N/A
42ZFR2AR39AR 67A-GCN4-VP4 8-DHFR self-activation	Tet3G-expressing CHO-K1	Monoclonal	MF70	2D, S5B	Obtained monoclone candidates by limiting dilution, induced candidates with 10 μM TMP and selected the monoclone that spontaneously and homogenously self-activate
FKBP-42ZFR2A R39AR67A-VP4 8-DHFR self-activation	Tet3G-expressing CHO-K1	Monoclonal	MF71	2D, S5B	Obtained monoclone candidates by limiting dilution, induced candidates with 100 nM AP1903 + 10 μM TMP and selected the monoclone that spontaneously and homogenously self-activate
MultiFate-2.1	ERT2-Gal4 -T2A-Tet3G expressing CHO-K1	Monoclonal	MF64, MF81	3C, 4B, S6	Induced the polyclonal population with 500 ng/ml Dox for 12 hours, then washed out Dox and changed to 100 nM AP1903 + 10 μM TMP for 3 days, FACS sorted monoclones that were mCherry+ and mCitrine+
MultiFate-2.2	ERT2-Gal4 -T2A-Tet3G expressing CHO-K1	Monoclonal	MF82, MF83	3C, S7	Induced the polyclonal population with 500 ng/ml Dox for 12 hours, then washed out Dox and changed to 100 nM AP1903 + 10 μM TMP for 3 days, FACS sorted monoclones that were mCherry+ and mCitrine+
MultiFate-2.3	ERT2-Gal4 -T2A-Tet3G expressing CHO-K1	Monoclonal	MF64, MF84	3C, 3D, 4A, S8, S9	Induced the polyclonal population with 500 ng/ml Dox and 75 nM 4-OHT for 12 hours, then washed out Dox and 4-OHT and changed to 100 nM AP1903 + 10 μM TMP for 3 days, FACS sorted

					monoclones that were mCherry+ and mCitrine+
MultiFate-3	MultiFate-2.2	Monoclonal	MF85	5B, 5C, S10-S14	Induced the polyclonal population with 500 ng/ml Dox and 75 nM 4-OHT for 12 hours, then washed out Dox and 4-OHT and changed to 100 nM AP1903 + 10 μM TMP for 3 days, FACS sorted monoclones that were mCherry+, mCitrine+ and mTurquoise2+
Promoter structures of different MultiFate lines: MultiFate-2.1 TF A promoter has Tet3G binding sites, 6x(BCRbs_BCRbs); TF B promoter has Tet3G binding sites, 6x(37bs_37bs); MultiFate-2.2 TF A promoter has Tet3G binding sites, 10x(BCRbs_BCRbs); TF B promoter has Tet3G binding sites, 10x(37bs_37bs); MultiFate-2.3 TF A promoter has ERT2-Gal4 binding sites, 6x(BCRbs_BCRbs); TF B promoter has Tet3G binding sites, 6x(37bs_37bs); MultiFate-3 TF A promoter has Tet3G binding sites, 10x(BCRbs_BCRbs); TF B promoter has Tet3G binding sites, 10x(37bs_37bs); TF C promoter has ERT2-Gal4 binding sites , 10x(ErbB2bs_ErbB2bs);					

Table S2.4 List of molecular reactions and their propensities for Gillespie simulation.

Reactions	Molecule update	Propensity
TF A mRNA transcription	a –> a + 1	$k_1 + k_2[A_2]^n/(K_M{}^n + [A_2]^n)$
TF A mRNA removal	a –> a - 1	$\delta_{mRNA}[a]$
TF B mRNA transcription	b –> b + 1	$rk_1 + mk_2[B_2]^n/((\kappa K_M)^n + [B_2]^n)$
TF B mRNA removal	b –> b - 1	$\delta_{mRNA}[b]$
TF C mRNA transcription	c –> c + 1	$r_2k_1 + m_2k_2[C_2]^n/((\kappa K_M)^n + [C_2]^n)$
TF C mRNA removal	c –> c - 1	$\delta_{mRNA}[c]$
TF A protein translation	A –> A + 1	$k_p[a]$
TF A protein removal	A –> A - 1	$\delta[A]$
TF B protein translation	B –> B + 1	$k_p[b]$
TF B protein removal	B –> B - 1	$\gamma\delta[B]$
TF C protein translation	C –> C + 1	$k_p[c]$
TF C protein removal	C –> C - 1	$\gamma_2\delta[C]$
TF A homodimerization	A –> A - 2, A_2 –> A_2 + 1	$k_{on}[A]^2 \times ([A] \geq 2)$
AA homodimer dissociation	A_2 –> A_2 - 1, A –> A + 2	$k_{off}[A_2]$
TF B homodimerization	B –> B - 2, B_2 –> B_2 + 1	$k_{on}[B]^2 \times ([B] \geq 2)$
BB homodimer dissociation	B_2 –> B_2 - 1, B –> B + 2	$k_{off}[B_2]$
TF C homodimerization	C –> C - 2, C_2 –> C_2 + 1	$k_{on}[C]^2 \times ([C] \geq 2)$
CC homodimer dissociation	C_2 –> C_2 - 1, C –> C + 2	$k_{off}[C_2]$

AB heterodimerization	A –> A - 1, B –> B - 1, AB –> AB + 1	$2k_{on}[A][B]$
AB heterodimer dissociation	AB –> AB - 1, A –> A + 1, B –> B + 1,	$k_{off}[AB]$
AC heterodimerization	A –> A - 1, C –> C - 1, AC –> AC + 1	$2k_{on}[A][C]$
AC heterodimer dissociation	AC –> AC - 1, A –> A + 1, C –> C + 1,	$k_{off}[AC]$
BC heterodimerization	B –> B - 1, C –> C - 1, BC –> BC + 1	$2k_{on}[B][C]$
BC heterodimer dissociation	BC –> BC - 1, B –> B + 1, C –> C + 1,	$k_{off}[BC]$
AA homodimer removal	A_2 –> A_2 - 1	$\delta[A_2]$
BB homodimer removal	B_2 –> B_2 - 1	$\gamma\delta[B_2]$
CC homodimer removal	C_2 –> C_2 - 1	$\gamma_2\delta[B_2]$
A removal in AB heterodimer	AB –> AB - 1, B –> B + 1	$\delta[AB]$
B removal in AB heterodimer	AB –> AB - 1, A –> A + 1	$\gamma\delta[AB]$
A removal in AC heterodimer	AC –> AC - 1, C –> C + 1	$\delta[AC]$
C removal in AC heterodimer	AC –> AC - 1, A –> A + 1	$\gamma_2\delta[AC]$
B removal in BC heterodimer	BC –> BC - 1, C –> C + 1	$\gamma\delta[BC]$
C removal in BC heterodimer	BC –> BC - 1, B –> B + 1	$\gamma_2\delta[BC]$
FPA_{im} protein translation	FPA_{im} –> FPA_{im} + 1	$I_{eff}k_p[a]$
FPB_{im} protein translation	FPB_{im} –> FPB_{im} + 1	$I_{eff}k_p[b]$
FPA_{im} protein maturation	FPA_{im} –> FPA_{im} - 1, FPA_m –> FPA_m + 1,	$k_{matA}[FPA_{im}]$

FPB_{im} protein maturation	FPB_{im} –> FPB_{im} - 1, FPB_m –> FPB_m + 1,	$k_{matB}[FPB_{im}]$
FPA_{im} protein removal	FPA_{im} –> FPA_{im} - 1	$\delta_{FPA}[FPA_{im}]$
FPB_{im} protein removal	FPB_{im} –> FPB_{im} - 1	$\delta_{FPB}[FPB_{im}]$
FPA_m protein removal	FPA_m –> FPA_m - 1	$\delta_{FPA}[FPA_m]$
FPB_m protein removal	FPB_m –> FPB_m - 1	$\delta_{FPB}[FPB_m]$

Note:
A, B, C, A_2, B_2, C_2, AB, AC, BC represent proteins of monomer A, monomer B, monomer C, homodimer AA, homodimer BB, homodimer CC, heterodimer AB, heterodimer AC, heterodimer BC, respectively.
a, b, c represents mRNAs of A, B and C, respectively.
FPA_{im}, FPB_{im}, FPA_m, FPB_m represents immature proteins of fluorescent protein A, immature proteins of fluorescent protein B, mature proteins of fluorescent protein A, mature proteins of fluorescent protein B, respectively.

2.12 References

1. S. R. y. Cajal, *Histology of the nervous system of man and vertebrates* (Oxford University Press, USA, 1995).

2. S. Huang, Multistability and Multicellularity: Cell Fates as High-Dimensional Attractors of Gene Regulatory Networks. *Computational Systems Biology* (2006), pp. 293–326.

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

4. C. M. Ajo-Franklin, D. A. Drubin, J. A. Eskin, E. P. S. Gee, D. Landgraf, I. Phillips, P. A. Silver, Rational design of memory in eukaryotic cells. *Genes Dev.* **21**, 2271–2276 (2007).

5. B. P. Kramer, A. U. Viretta, M. Daoud-El-Baba, D. Aubel, W. Weber, M. Fussenegger, An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22**, 867–870 (2004).

6. T. Lebar, U. Bezeljak, A. Golob, M. Jerala, L. Kadunc, B. Pirš, M. Stražar, D. Vučko, U. Zupančič, M. Benčina, V. Forstnerič, R. Gaber, J. Lonzarić, A. Majerle, A. Oblak, A. Smole, R. Jerala, A bistable genetic switch based on designable DNA-binding domains. *Nat. Commun.* **5**, 5007 (2014).

7. D. R. Burrill, M. C. Inniss, P. M. Boyle, P. A. Silver, Synthetic memory circuits for tracking human cell fate. *Genes Dev.* **26**, 1486–1497 (2012).

8. J. Santos-Moreno, E. Tasiudi, J. Stelling, Y. Schaerli, Multistable and dynamic CRISPRi-based synthetic circuits. *Nat. Commun.* **11**, 2746 (2020).

9. F. Wu, R.-Q. Su, Y.-C. Lai, X. Wang, Engineering of a synthetic quadrastable gene network to approach Waddington landscape and cell fate determination. *Elife*. **6** (2017), doi:10.7554/eLife.23702.

10. H. Y. Kueh, A. Champhekar, S. L. Nutt, M. B. Elowitz, E. V. Rothenberg, Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. *Science*. **341**, 670–673 (2013).

11. G. Yao, T. J. Lee, S. Mori, J. R. Nevins, L. You, A bistable Rb-E2F switch underlies the restriction point. *Nat. Cell Biol.* **10**, 476–482 (2008).

12. S. Huang, Y.-P. Guo, G. May, T. Enver, Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).

13. W. Xiong, J. E. Ferrell Jr, A positive-feedback-based bistable “memory module” that governs a cell fate decision. *Nature*. **426**, 460–465 (2003).

14. A. Gunne-Braden, A. Sullivan, B. Gharibi, R. S. M. Sheriff, A. Maity, Y.-F. Wang, A. Edwards, M. Jiang, M. Howell, R. Goldstone, R. Wollman, P. East, S. D. M. Santos, GATA3 Mediates a Fast, Irreversible Commitment to BMP4-Driven Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell*. **26**, 693–706.e9 (2020).

15. K. M. Loh, A. Chen, P. W. Koh, T. Z. Deng, R. Sinha, J. M. Tsai, A. A. Barkal, K. Y. Shen,

R. Jain, R. M. Morganti, N. Shyh-Chang, N. B. Fernhoff, B. M. George, G. Wernig, R. E. A. Salomon, Z. Chen, H. Vogel, J. A. Epstein, A. Kundaje, W. S. Talbot, P. A. Beachy, L. T. Ang, I. L. Weissman, Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell*. **166**, 451–467 (2016).

16. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. **126**, 663–676 (2006).

17. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

18. C. Sokolik, Y. Liu, D. Bauer, J. McPherson, M. Broeker, G. Heimberg, L. S. Qi, D. A. Sivak, M. Thomson, Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst*. **1**, 117–129 (2015).

19. A. Bhattacharya, N. E. Baker, A network of broadly expressed HLH genes regulates tissue-specific cell fates. *Cell*. **147**, 881–892 (2011).

20. H. Hosokawa, J. Ungerbäck, X. Wang, M. Matsumoto, K. I. Nakayama, S. M. Cohen, T. Tanaka, E. V. Rothenberg, Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity*. **48** (2018), pp. 1119–1134.e7.

21. S. Stefanovic, N. Abboud, S. Désilets, D. Nury, C. Cowan, M. Pucéat, Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *J. Cell Biol.* **186**, 665–673 (2009).

22. I. Aksoy, R. Jauch, J. Chen, M. Dyla, U. Divakar, G. K. Bogu, R. Teo, C. K. Leng Ng, W. Herath, S. Lili, A. P. Hutchins, P. Robson, P. R. Kolatkar, L. W. Stanton, Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J.* **32**, 938–953 (2013).

23. M. J. Thayer, S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, H. Weintraub, Positive autoregulation of the myogenic determination gene MyoD1. *Cell*. **58**, 241–248 (1989).

24. L. A. Neuhold, B. Wold, HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell*. **74**, 1033–1042 (1993).

25. Detailed materials and methods are available as supplementary materials.

26. S. Bessonnard, L. De Mot, D. Gonze, M. Barriol, C. Dennis, A. Goldbeter, G. Dupont, C. Chazaud, Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development*. **141**, 3637–3648 (2014).

27. P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, H. Singh, Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell*. **126**, 755–766 (2006).

28. T. Miyamoto, H. Iwasaki, B. Reizis, M. Ye, T. Graf, I. L. Weissman, K. Akashi, Myeloid or Lymphoid Promiscuity as a Critical Step in Hematopoietic Lineage Commitment.

Developmental Cell. **3** (2002), pp. 137–147.

29. J. J. Lohmueller, T. Z. Armel, P. A. Silver, A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Research*. **40** (2012), pp. 5180–5187.

30. J. Miyazaki, S. Takaki, K. Araki, F. Tashiro, A. Tominaga, K. Takatsu, K. Yamamura, Expression vector system based on the chicken beta-actin promoter directs efficient production of interleukin-5. *Gene*. **79**, 269–277 (1989).

31. H. Chassin, M. Müller, M. Tigges, L. Scheller, M. Lang, M. Fussenegger, A modular degron library for synthetic circuits in mammalian cells. *Nat. Commun.* **10**, 2013 (2019).

32. A. S. Khalil, T. K. Lu, C. J. Bashor, C. L. Ramirez, N. C. Pyenson, J. K. Joung, J. J. Collins, A synthetic biology framework for programming eukaryotic transcription functions. *Cell*. **150**, 647–658 (2012).

33. R. R. Beerli, D. J. Segal, B. Dreier, C. F. Barbas 3rd, Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14628–14633 (1998).

34. C. J. Bashor, N. Patel, S. Choubey, A. Beyzavi, J. Kondev, J. J. Collins, A. S. Khalil, Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science*. **364**, 593–597 (2019).

35. O. M. Subach, P. J. Cranfill, M. W. Davidson, V. V. Verkhusha, An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS One*. **6**, e28674 (2011).

36. P. S. Donahue, J. W. Draut, J. J. Muldoon, H. I. Edelstein, N. Bagheri, J. N. Leonard, The COMET toolkit for composing customizable genetic programs in mammalian cells. *Nat. Commun.* **11**, 779 (2020).

37. M. Elrod-Erickson, T. E. Benson, C. O. Pabo, High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure*. **6**, 451–464 (1998).

38. T. Clackson, W. Yang, L. W. Rozamus, M. Hatada, J. F. Amara, C. T. Rollins, L. F. Stevenson, S. R. Magari, S. A. Wood, N. L. Courage, X. Lu, F. Cerasoli Jr, M. Gilman, D. A. Holt, Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 10437–10442 (1998).

39. M. Iwamoto, T. Björklund, C. Lundberg, D. Kirik, T. J. Wandless, A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17**, 981–988 (2010).

40. S. S. Gerety, M. A. Breau, N. Sasai, Q. Xu, J. Briscoe, D. G. Wilkinson, An inducible transgene expression system for zebrafish and chick. *Development*. **140**, 2235–2243 (2013).

41. D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).

42. F. Notta, S. Zandi, N. Takayama, S. Dobson, O. I. Gan, G. Wilson, K. B. Kaufmann, J. McLeod, E. Laurenti, C. F. Dunant, J. D. McPherson, L. D. Stein, Y. Dror, J. E. Dick, Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*. **351**, aab2116 (2016).

43. P. S. Swain, M. B. Elowitz, E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12795–12800 (2002).

44. K. Pougach, A. Voet, F. A. Kondrashov, K. Voordeckers, J. F. Christiaens, B. Baying, V. Benes, R. Sakai, J. Aerts, B. Zhu, P. Van Dijck, K. J. Verstrepen, Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat. Commun.* **5**, 4868 (2014).

45. J. González, G. López, S. Argueta, X. Escalera-Fanjul, M. El Hafidi, C. Campero-Basaldua, J. Strauss, L. Riego-Ruiz, A. González, Diversification of Transcriptional Regulation Determines Subfunctionalization of Paralogous Branched Chain Aminotransferases in the Yeast. *Genetics*. **207**, 975–991 (2017).

46. L. Morsut, K. T. Roybal, X. Xiong, R. M. Gordley, S. M. Coyle, M. Thomson, W. A. Lim, Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell*. **164**, 780–791 (2016).

47. S. Toda, W. L. McKeithan, T. J. Hakkinen, P. Lopez, O. D. Klein, W. A. Lim, Engineering synthetic morphogen systems that can program multicellular patterning. *Science*. **370**, 327–331 (2020).

48. K. A. Schwarz, N. M. Daringer, T. B. Dolberg, J. N. Leonard, Rewiring human cellular input–output using modular extracellular sensors. *Nature Chemical Biology*. **13** (2017), pp. 202–209.

49. I. Moraga, J. B. Spangler, J. L. Mendoza, M. Gakovic, T. S. Wehrman, P. Krutzik, K. C. Garcia, Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers. *Elife*. **6** (2017), doi:10.7554/eLife.22882.

50. K. S. Stapornwongkul, M. de Gennes, L. Cocconi, G. Salbreux, J.-P. Vincent, Patterning and growth control in vivo by an engineered GFP gradient. *Science*. **370**, 321–327 (2020).

51. Y. Ma, M. W. Budde, M. N. Mayalu, J. Zhu, R. M. Murray, M. B. Elowitz, Synthetic mammalian signaling circuits for robust cell population control. *bioRxiv*, doi:10.1101/2020.09.02.278564.

52. M. R. Bennett, D. Volfson, L. Tsimring, J. Hasty, Transient Dynamics of Genetic Regulatory Networks. *Biophysical Journal*. **92** (2007), pp. 3501–3512.

53. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Hachette UK, 2014).

54. K. J. Polach, J. Widom, A Model for the Cooperative Binding of Eukaryotic Regulatory Proteins to Nucleosomal Target Sites. *Journal of Molecular Biology*. **258** (1996), pp. 800–812.

55. T. Buder, A. Deutsch, M. Seifert, A. Voss-Böhme, CellTrans: An R Package to Quantify

Stochastic Cell State Transitions. *Bioinformatics and Biology Insights*. **11** (2017), p. 117793221771224.

56. X. Pan, C. Dalm, R. H. Wijffels, D. E. Martens, Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Appl. Microbiol. Biotechnol.* **101**, 8101–8113 (2017).

57. Q. Liu, D. J. Segal, J. B. Ghiara, C. F. Barbas 3rd, Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 5525–5530 (1997).

58. L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, R. Phillips, Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* **15**, 125–135 (2005).

59. E. Eden, N. Geva-Zatorsky, I. Issaeva, A. Cohen, E. Dekel, T. Danon, L. Cohen, A. Mayo, U. Alon, Proteome half-life dynamics in living human cells. *Science*. **331**, 764–768 (2011).

60. B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control. *Nature*. **473**, 337–342 (2011).

61. K. L. Frieda, J. M. Linton, S. Hormoz, J. Choi, K.-H. K. Chow, Z. S. Singer, M. W. Budde, M. B. Elowitz, L. Cai, Synthetic recording and in situ readout of lineage information in single cells. *Nature*. **541**, 107–111 (2017).

62. B. Schwalb, M. Michel, B. Zacher, K. Frühauf, C. Demel, A. Tresch, J. Gagneur, P. Cramer, TT-seq maps the human transient transcriptome. *Science*. **352**, 1225–1228 (2016).

63. J. J. Muldoon, V. Kandula, M. Hong, P. S. Donahue, J. D. Boucher, N. Bagheri, J. N. Leonard, Model-guided design of mammalian genetic programs. *Cold Spring Harbor Laboratory* (2020), p. 2020.09.30.320853.

64. J. A. Zitzewitz, O. Bilsel, J. Luo, B. E. Jones, C. R. Matthews, Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*. **34**, 12812–12819 (1995).

65. M. Schlosshauer, D. Baker, Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. *Protein Sci.* **13**, 1660–1669 (2004).

66. S. Paulous, C. E. Malnou, Y. M. Michel, K. M. Kean, A. M. Borman, Comparison of the capacity of different viral internal ribosome entry segments to direct translation initiation in poly(A)-dependent reticulocyte lysates. *Nucleic Acids Res.* **31**, 722–733 (2003).

67. E. Balleza, J. M. Kim, P. Cluzel, Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods*. **15**, 47–51 (2018).

68. L. He, R. Binari, J. Huang, J. Falo-Sanjuan, N. Perrimon, In vivo study of gene expression with an enhanced dual-color fluorescent transcriptional timer. *Elife*. **8** (2019), doi:10.7554/eLife.46181.

69. J. E. Ferrell Jr, S. H. Ha, Ultrasensitivity part II: multisite phosphorylation, stoichiometric

inhibitors, and positive feedback. *Trends Biochem. Sci.* **39**, 556–569 (2014).

70. J. E. Ferrell Jr, S. H. Ha, Ultrasensitivity part III: cascades, bistable switches, and oscillators. *Trends Biochem. Sci.* **39**, 612–618 (2014).

71. J. A. Miller, J. Widom, Collaborative competition mechanism for gene activation in vivo. *Mol. Cell. Biol.* **23**, 1623–1632 (2003).

72. L. A. Mirny, Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22534–22539 (2010).

73. N. E. Buchler, M. Louis, Molecular titration and ultrasensitivity in regulatory networks. *J. Mol. Biol.* **384**, 1106–1119 (2008).

74. C. Hsu, V. Jaquet, M. Gencoglu, A. Becskei, Protein Dimerization Generates Bistability in Positive Feedback Loops. *Cell Rep.* **16**, 1204–1210 (2016).

75. D. V. Israni, H.-S. Li, K. A. Gagnon, J. D. Sander, K. T. Roybal, J. Keith Joung, W. W. Wong, A. S. Khalil, Clinically-driven design of synthetic gene regulatory programs in human cells. *bioRxiv*, doi:10.1101/2021.02.22.432371.

Chapter 3. Reconstitution of BMP gradient reveals the sufficient circuit for ligand shuttling

Abstract

Morphogens pattern developing tissues by forming concentration gradients. Understanding how the shapes of these gradients are modulated can provide key insights into how natural tissues are patterned, and allow engineering of synthetic patterns. Many extracellular modulators have been identified to regulate BMP gradient formations during development, but how they work together to shape BMP gradients remains unclear. Here, combining in vitro bottom-up gradient reconstitution and mathematical modeling, we systematically analyzed a gradient modulation circuit consisting of mouse BMP4 and its modulators, Chordin, Twsg, and BMP-1. We found that this circuit can give rise to diverse gradient modulation capabilities such as gradient lengthening and gradient suppression. The full circuit is sufficient for the active process of ligand shuttling and can generate a non-monotonic displaced gradient. Bottom-up gradient reconstitution thus provides a scalable platform for understanding how morphogen gradients are shaped by extracellular modulators, and using these modulators to engineer novel synthetic multicellular patterns.

3.1 Introduction

In multicellular organisms, a class of long-range signaling molecules called morphogens pattern developmenting tissues (*1–4*). A major challenge in developmental biology is to understand how morphogens are used to pattern diverse tissue types. In the simplest

models of morphogenetic patterning, morphogen synthesis, degradation and diffusion establish spatial concentration gradients, which allow signal receiving cells to infer their spatial positions and control fate decisions (*5*). However, natural morphogen pathways use additional extracellular modulators that interact with one another and with morphogens to actively shape gradients in more complex ways (*6–8*). The roles of many of these components have been elucidated in specific developmental contexts, but the design principles that govern the regulatory architecture of patterning systems remain unclear. Being able to understand these principles would enable us to control tissue patterning more precisely and even rationally engineer more complex, synthetic tissue patterns.

Most approaches have relied on analysis of pattern formation in the context of wild-type and mutant embryos and tissues. These ‘top-down’ approaches have provided fundamental insights into the identities of morphogens and their regulators, their connectivity, and their general role in the control of gradient shape and scale (*9*). However, natural developmental patterning systems use multiple interacting components concurrently, re-use them pleiotropically across distinct embryo contexts, and often offer limited experimental accessibility. These features make it difficult to isolate a specific set of components, test their sufficiency for pattern formation, and more generally understand the full repertoire of potential pattern formation behaviors they can generate. In contrast, recent work (*10*, *11*) showed that reconstituting pattern formation circuits from the bottom up could address these issues and provide a powerful and complementary method into natural patterning systems. This bottom-up approach enables one to study these pattern formation circuits in an isolated, quantitative manner.

The Bone Morphogenetic Protein (BMP) pathway is one of the most important and best-studied metazoan morphogen pathways. BMP ligands form gradients that control dorsal-ventral patterning of early embryos (*12*, *13*), the neural tube (*14*), and limb bud formation (*15*), among many other processes. BMP is particularly notable for its extensive use of extracellular modulators. One interaction that is particularly interesting is interactions among BMP, Chordin, Twisted gastrulation (Twsg), and BMP-1 (Fig. 3.1A). Chordin can bind to BMP and prevent it from binding to receptors for signaling (*16*), and Twsg can form ternary complexes with BMP-Chordin to further strengthen this interaction (*17–20*). Thus, Chordin is mostly known as a BMP antagonist. However, this interaction may serve another purpose: it prevents BMP from receptor mediated internalization and degradation (*21*). Thus it is postulated that Chordin can "shuttle" BMP ligands on the cell surface and increase their diffusibility. Furthermore, BMP-1 (a homolog of Tolloid in Drosophila) protease can cleave Chordin, which releases BMP from the ternary complex for signaling (*22*). Ligand shuttling enabled by a circuit with BMP, Chordin, Tsg, and BMP-1 has been suggested to produce non-intuitive gradient shapes during the embryonic development of Drosophila (*23*, *24*), Xenopus (*25*), and mouse (*26*). However, two papers refute the role of BMP ligand shuttling by Chordin in zebrafish embryos, despite the presence of all circuit components (*27*, *28*). This controversy may come from the limitation of top-down approaches: with the same circuit, ligand shuttling and certain gradient shapes may happen in one parameter regime, but not the other. One developmental context may only allow us to explore a limited set of parameter regimes, within which some circuit behaviors may not appear.

Here, we use a bottom-up gradient reconstitution system to isolate the BMP gradient formation process, analyze it in space and time, and perturb it through systematic, quantitative control of key modulators, including Chordin, Twsg, and BMP-1. We show that gradient reconstitution can provide a quantitative understanding of how BMP and a handful of modulators interact with one another in space and time to enable active patterning behaviors such as shuttling, along with other modulatory capabilities (Fig. 3.1B). These results provide a foundation for a systematic bottom-up reconstitution of diverse patterns and pattern-forming mechanisms.

Fig. 3.1

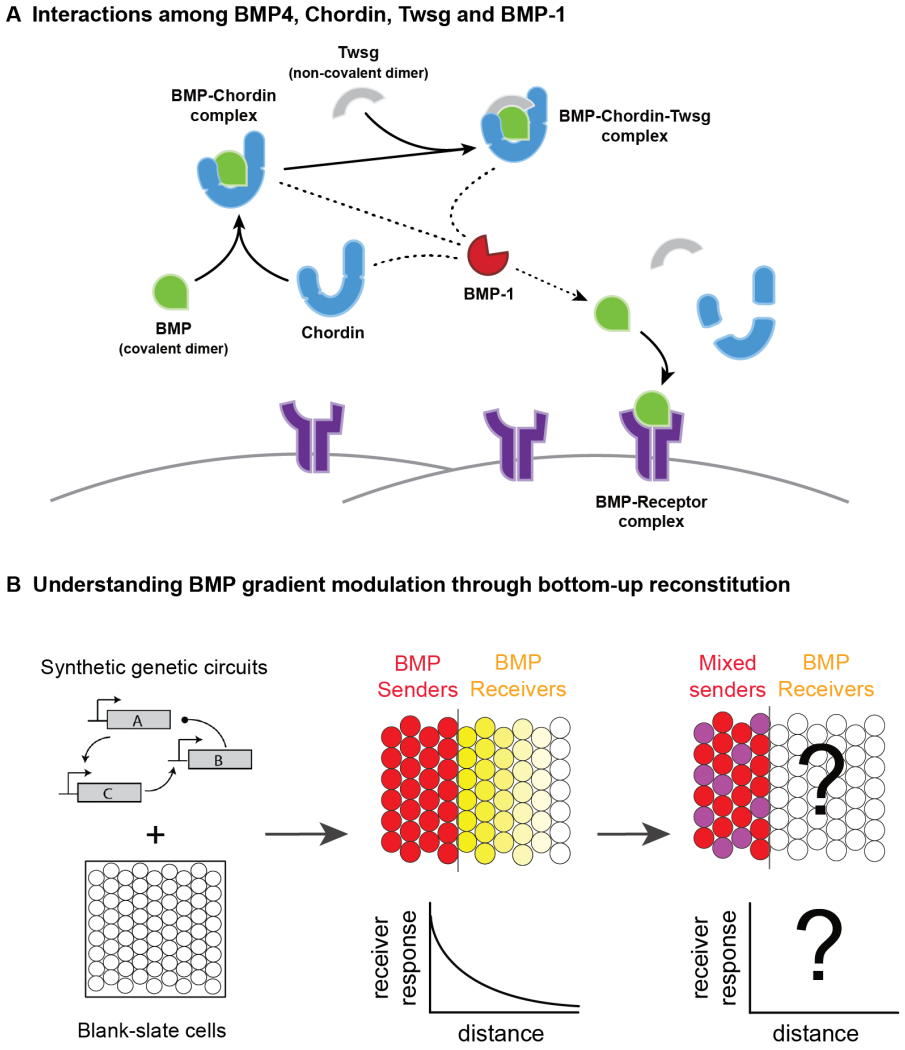

Fig. 3.1 BMP4 gradient modulation by Chordin, Twsg, and BMP-1 can be understood through gradient reconstitution.

(A) Chordin can bind to BMP ligands, and Twsg strengthens this interaction. BMP-1 can cleave Chordin in its free form or complex forms and release BMP for binding to receptor and signaling. (B) To reconstitute BMP gradients, we choose blank-slate cells that (1) express minimal levels of morphogen pathway components except for receptors, and (2) have minimal cellular state changes in response to morphogen signaling. Synthetic genetic components are engineered into blank-slate cells to make morphogen (or modulator) senders and morphogen receivers (left). Engineered senders and receivers are co-cultured in different configurations (middle and right) to reconstitute BMP gradients.

3.2 BMP gradients can be reconstituted in confluent NMuMG monolayers

Reconstitution of BMP gradients requires sender cells that produce BMP morphogen and receiver cells that can respond to it. To focus on gradient dynamics, and decouple them from proliferation, these cells should ideally form a contact-inhibited confluent monolayer. They should also transmit BMP signals with minimal biological responses such as differentiation. Based on these requirements, we used the NAMRU mouse mammary gland (NMuMG) epithelial cell line, which was previously shown to form confluent monolayers and to provide a high dynamic range of BMP signaling (*29*, *30*). We engineered NMuMG sender cells that express BMP4 ligands together with a co-expressed mCherry reporter under the control of the inducer 4-hydroxytamoxiofen (4-OHT) (Fig. 3.2A). For receiver cells, we utilized a previously described NMuMG cell line incorporating a H2B-Citrine reporter under the control of a BMP-responsive regulatory element, based on BMP response elements in the natural BMP target gene Id1 (*30*). In this cell line, H2B-Citrine fluorescence is regulated by BMP in a dose-dependent fashion that correlates with the levels of phosphorylated Smad1/5 and with expression of endogenous BMP target genes (*30*). We verified that sender cells were able to activate the receiver cells in a dose-dependent fashion across their full dynamic response range (Fig. 3.2B).

To reconstitute gradient formation, we used a similar sequential plating protocol previously developed in our lab (*10*). Briefly, we first plated senders in a confined region using a PDMS insert (Ibidi). After senders were plated, we removed the insert and plated receivers in an unconfined region. We then induced the secretion of BMP4 from the sender region with 4-OHT, and imaged the BMP4 signaling gradient using fluorescent

microscopes (Fig. 3.2C). With this system, a quasi-1D gradient formed along the direction perpendicular to the interface between the senders and receivers (Fig. 3.2D), similar to the geometry of various developmental tissues like a Drosophila wing disc (*31*). Time-lapse movies showed that, at saturating 4-OHT induction, BMP4 signaling gradients grew over time in amplitude and lengthscale, and eventually reached a steady state in around 48 hours (Fig. 3.2D). This behavior is consistent with a simple model of BMP4 gradient formation based on ligand production, diffusion and removal. Together, these results show that BMP gradients can be reconstituted on a monolayer of engineered sender and receiver NMuMG cells.

Fig. 3.2

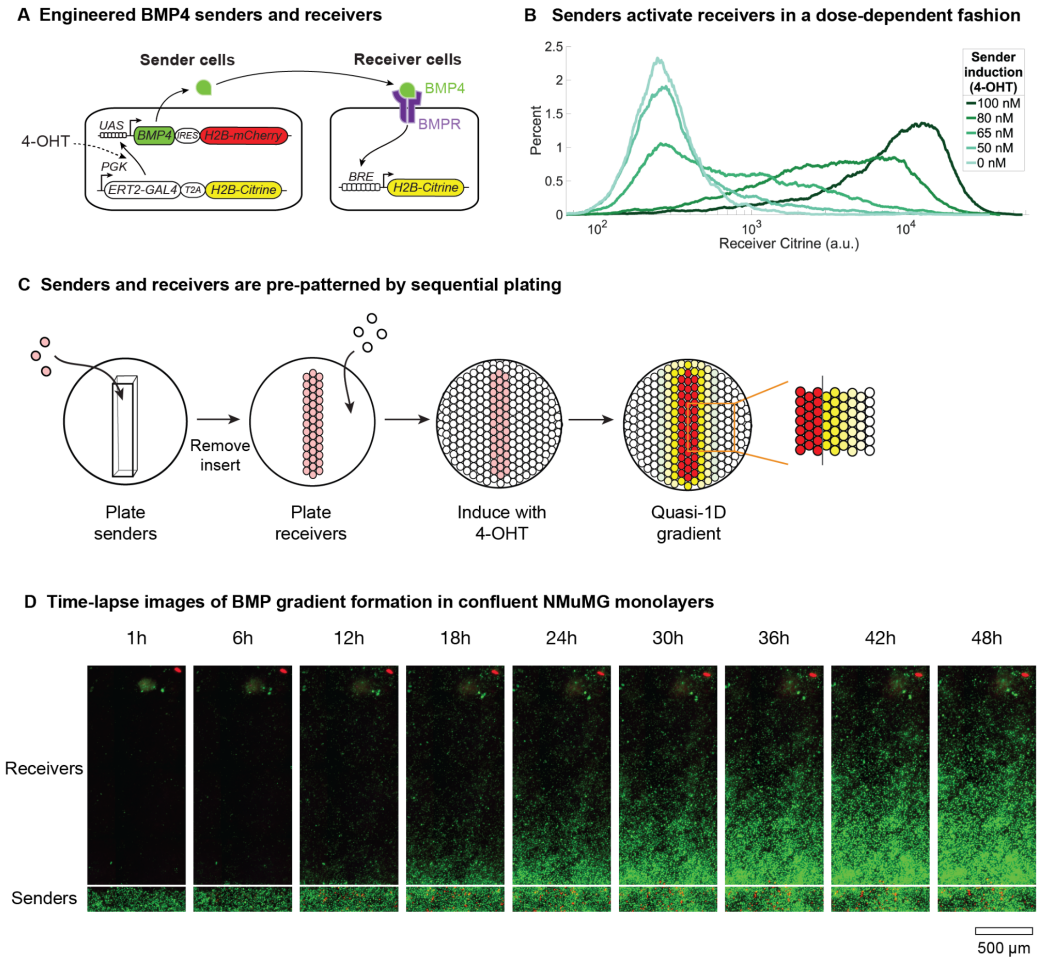

Fig. 3.2 BMP gradients can be reconstituted in vitro.
(A) An 4-OHT inducible BMP4 sending module is integrated into NMuMG cells to construct sender cells. Receiver cells are from (*30*), in which a H2B-Citrine reporter under the control of a BMP-responsive regulatory element (BRE) is integrated into the cells. UAS denotes upstream activating sequence that can be bound by Gal4; PGK denotes a mammalian constitutive promoter from the gene 3-phosphoglycerate kinase; BMPR denotes BMP receptors, which are endogenously expressed by NMuMG cells. (B) Senders activate Citrine responses in receiver cells in a 4-OHT dose-dependent manner. Sender cells and receiver cells were mixed and plated together in a 96-well plate. After 24hrs, different concentrations of 4-OHT were added to the culture media and cell fluorescence was measured by flow cytometry. Sender cells were separated from receiver cells in the flow data by their elevated mCherry expression, then receiver Citrine responses were plotted. (C) The sequential plating protocol is adopted from (*30*). One modification in this study is that during the induction step, instead of adding liquid media, we added 50% Matrigel and allowed it to solidify to prevent protein loss into liquid media during gradient formation. (D) Time-lapse imaging reveals gradient formation dynamics. In the sender region (bottom), 5,000 BMP4 sender cells are mixed with 35,000 filler cells that constitutively express Citrine.

3.3 Interactions among BMP4 and its modulators are recapitulated in the reconstitution system

During development, a set of extracellular modulators including Chordin, Twsg, and BMP-1 interact with each other and with BMP4 ligands to regulate and spatially redistribute BMP activity (Fig. 3.1A) (*6*, *7*). We next sought to determine whether these components could function in the reconstituted system. To determine how these factors together control the level of BMP signaling, we constructed a dual sender cell line that expresses BMP4 and Chordin (Sender-BC) under the control of 4-OHT and doxycycline (Dox), respectively (Fig. 3.3A). We also constructed a modified receiver cell line, Receiver-B1, containing the same reporter system but with an abscisic acid (ABA) inducible expression of BMP-1 protease. In some experiments we also incorporated different concentrations of recombinant mouse Twsg protein (R&D Systems).

We first focused on Chordin, an inhibitor that binds to BMP4 and prevents it from signaling. We used 4-OHT to induce BMP4 expression in the co-culture of Sender-BC and Receiver-B1 cells, and analyzed activation of Receiver-B1 cells using flow cytometry (Fig. 3.3B, left) and fluorescence imaging (Fig. 3.3B, right). Sender-BC cells efficiently activated Receiver-B1 cells (Fig. 3.3B, lane 1). As expected, this activation was reduced by co-induction of Chordin expression by doxycycline (Fig. 3.3B, lane 5). Addition of recombinant Twsg (rTwsg), which facilitates Chordin-BMP4 interactions (*17*), strengthened inhibition by Chordin, allowing it to fully suppress BMP4 signaling (Fig. 3.3B, lane 6), while rTwsg addition by itself had minimal effects (Fig. 3.3B, lane 2). By contrast, expression of the BMP-1 protease, which cleaves Chordin to release BMP4 ligands, partially relieved BMP4 suppression by Chordin (Fig. 3.3B, lane 7 and 8).

Together, these results show that interactions among BMP4, Chordin, Twsg, and BMP-1 can be recapitulated in our cell culture system.

Fig. 3.3

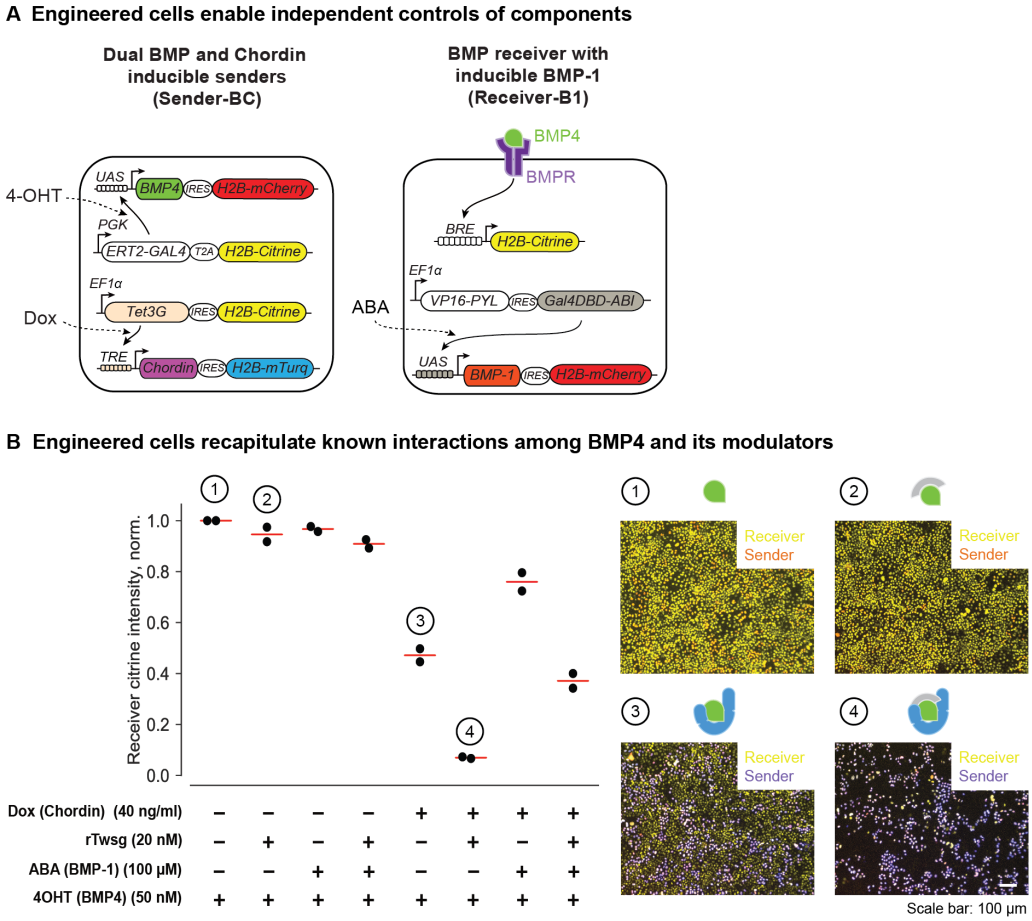

Fig. 3.3 Engineered cells recapitulate known interactions among BMP4, Chordin, Twsg, and BMP-1.

(A) On top of sender cells in Fig. 3.2A, we added a Dox-inducible Chordin expression module to make a dual BMP and Chordin inducible sender line (Sender-BC). This allows independent control of BMP4 and Chordin expression. For Receiver-B1, we integrated an ABA-inducible BMP-1 expression module in the receiver cells in Fig. 3.2A. EF1α denotes a mammalian constitutive promoter from the gene human elongation factor-1α; TRE denotes Tet3G response elements; (B) In the co-culture experiment, 5,000 Sender-BC were mixed with 45,000 Receiver-B1 and plated into a 96-well plate. After 24hrs, inducers were added to the media and cell fluorescence was measured by flow cytometry after another 36hrs. Images of cells (right) were taken right before flow cytometry measurements. On the left plot, each dot represents one biological replicate, and each red line indicates the mean of two replicates.

3.4 Mathematical modeling predicts gradient features

To understand what types of patterns could be generated by these interactions, we developed a simple mathematical model of BMP and its modulators based on previous models (*23–25*, *32*) but specifically incorporating the components analyzed here (Fig. 3.4A, Supplementary Information). In the model, BMP4 is secreted from sender cells, can be internalized through binding to receptors, undergoes diffusion, and forms complexes with Chordin. We further assume that BMP4 and Chordin form a more stable complex in the presence of Twsg, and BMP-1 cleaves Chordin with a rate dependent on BMP-1 expression. We used existing biochemical measurements (Supplementary Information) to estimate key parameters, including the diffusion coefficient, D, and the binding and unbinding rates between BMP4 and receptors or Chordin, and the rate of receptor-mediated internalization of BMP4 ligands. We also assume that the fluorescent reporter for pathway activity is synthesized at a rate that is a Hill function of the concentration of ligand-receptor signaling complexes. We simulated the model in one spatial dimension, parallel to the gradient, corresponding to the quasi-1D gradients in the experiment.

Using this model, we asked how each modulator should affect the signaling gradient.. We first analyzed the simplest case, in which BMP4 and Chordin are produced from the sender cell region. Simulations showed that Chordin both lengthened the gradient and reduced its amplitude, consistent with the dual role of BMP4-Chordin interactions (*21*) (Fig. 3.4B). In this case, BMP4 ligands can be released from the BMP4-Chordin complex due to a weak affinity between the two. In the presence of Twsg globally, Chordin was predicted to strongly bind to BMP4, and thus could completely abrogate the signaling

gradient (Fig. 3.4C). Strikingly, further introducing BMP-1 expression in receiver cells generated a non-monotonic displaced gradient, because BMP-1 can cleave Chordin and release BMP4 for signaling again. The gradient peak amplitude and peak displacement (peak distance to sender-receiver interface) were concomitantly controlled by the level of BMP-1, with higher BMP-1 expression both elevating the peak amplitude and reducing the peak displacement (Fig. 3.4D). Together, mathematical modeling demonstrated that interactions among BMP4, Chordin, Twsg, and BMP-1 enable diverse gradient modulation capabilities.

Fig. 3.4

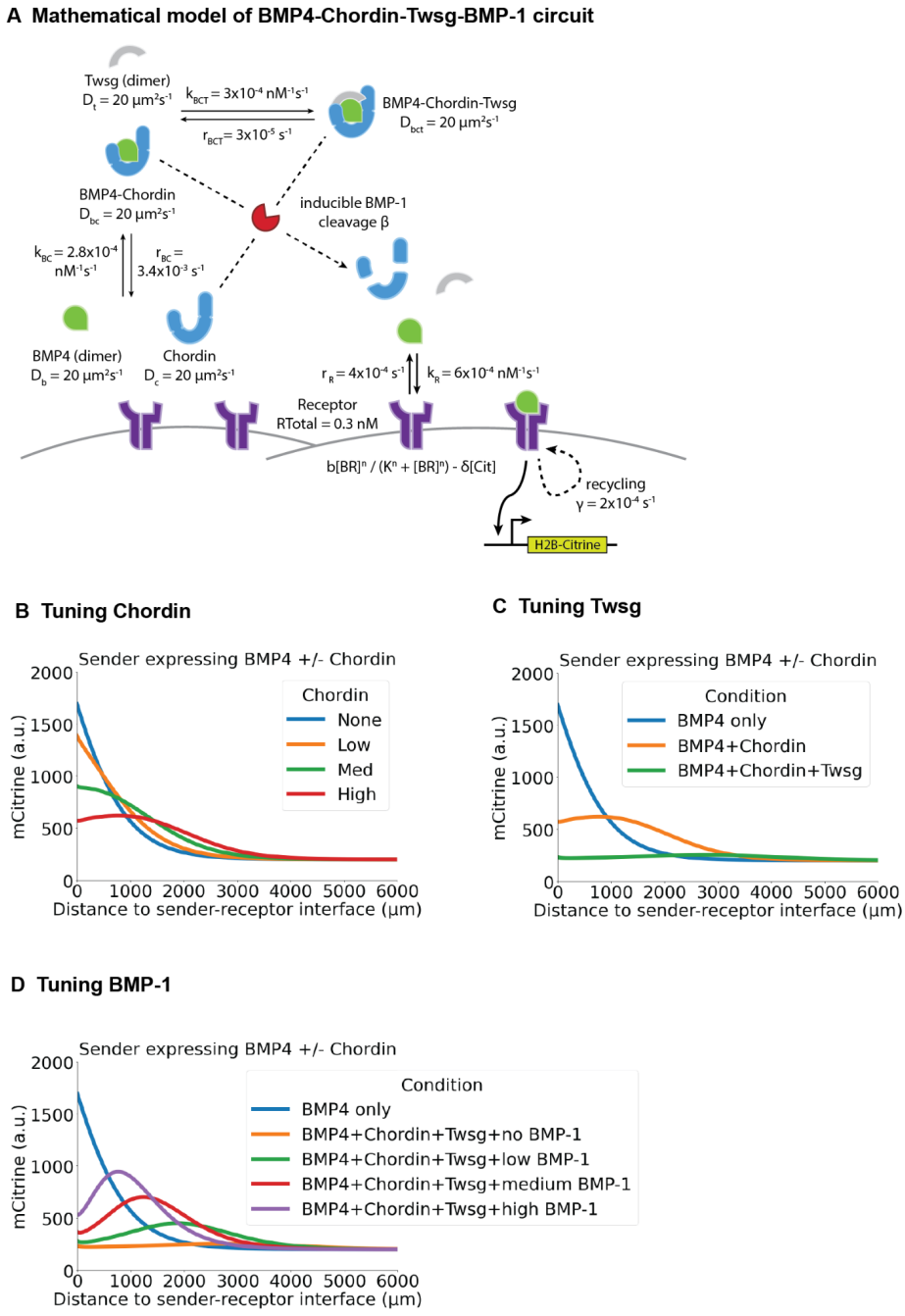

Fig. 3.4 Mathematical modeling shows diverse gradient modulation capabilities by Chordin, Twsg, and BMP-1.

(A) Mathematical model of BMP4, Chordin, Twsg and BMP-1 circuit. Model parameters were estimated based on existing literature (Supplementary Information). (B-D) Tuning different circuit components enables diverse gradient modulation capabilities, including (B) gradient lengthening by tuning Chordin, (C) gradient suppression by tuning Twsg, and (D) gradient displacement by tuning BMP-1.

3.5 Ligand shuttling can be reconstituted by a minimal circuit consisting of BMP4, Chordin, Twsg, and BMP-1

To test these predictions, we followed the same sequential plating protocol mentioned above to set up experimental gradients using engineered sender cells and receiver cells. In the sender region, we separated BMP4 and Chordin sending capabilities into two different sender lines (Sender-B and Sender-C, respectively) (Fig. 3.5A), because through both changing cell numbers and inducer concentrations, a two-sender system gave us more flexibility than using the dual sender line. We used the inducible BMP-1 receivers mentioned above (Receiver-B1) in the receiver region.

Similar to the simple sender-receiver gradient (Fig. 3.2D), BMP4 production from the sender region generated an exponentially decaying gradient at 48hrs. Induction of Chordin production from the sender region lengthened the gradient while reduced its amplitude (Fig. 3.5B). As predicted by the model (Fig. 3.4B), this lengthening effect was dose-dependent, with increased Chordin production resulting in longer gradient lengthscales and lower amplitudes (Fig. 3.5B). Consistent with co-culture experiments (Fig. 3.3B) and model predictions (Fig. 3.4C), global rTwsg addition greatly enhanced the inhibitory effects of Chordin, leading to a complete suppression of BMP4 signaling gradients (Fig. 3.5C). Finally, we tested the full circuit with BMP4, Chordin, Twsg and BMP-1 expressed from cells or added as recombinant proteins. Strikingly, when all components were present, the BMP4 signaling gradient was displaced from the sender region (Fig. 3.5D). Crucially, the displaced gradient extended to a distance beyond that reached when only BMP4 was expressed (Fig. 3.5D), suggesting substantial ligand shuttling. The model predicts that the peak displacement and amplitude can be controlled

by the level of BMP-1 (Fig. 3.4D). To test this prediction, we held the level of BMP4, Chordin, and rTwsg constant, and tuned the level of BMP-1 by ABA. Increasing expression of BMP-1, as shown by increasing the level of mCherry fluorescence in the receiver region, decreased the peak displacement and increased the peak amplitude (Fig. 3.5D), consistent with model predictions (Fig. 3.4D). Together, these experimental results confirmed model predictions that a minimal circuit with BMP4, Chordin, Twsg, and BMP-1 enables ligand shuttling and allows for diverse gradient modulation.

Fig. 3.5

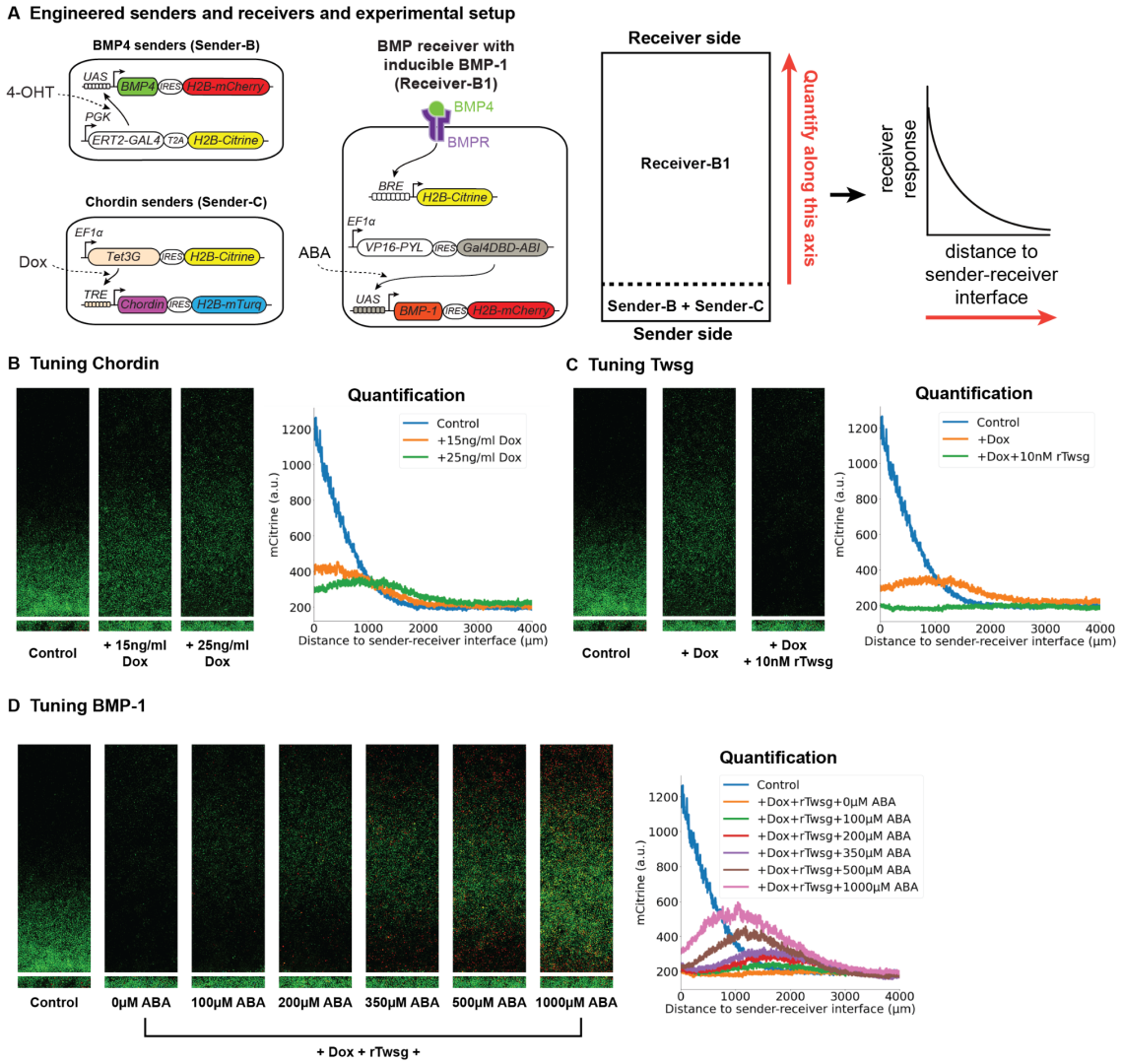

Fig. 3.5 The BMP4-Chordin-Twsg-BMP-1 circuit is sufficient for ligand shuttling. (A) (Left) BMP4 and Chordin were secreted by two different senders (Sender-B and Sender-C, respectively), and Receiver-B1 (also used in Fig. 3.3) was used in the receiver region. (Right) To set up gradient reconstitution experiments, 5,000 Sender-B and 35,000 Sender-C were mixed and first plated in a confined sender region (bottom side). Then Receiver-B1 cells were plated at 100% confluency throughout the well. Gradient was quantified along an axis perpendicular to the sender-receiver interface. (B-D) Consistent with modeling, tuning different circuit components enables diverse gradient modulation capabilities, including (B) gradient lengthening, (C) gradient suppression, and (D) gradient displacement. 2500 nM 4-OHT was added to all samples. Control sample only had 4-OHT added. In (C), the concentration of Dox is 25 ng/ml. In (D), the concentration of Dox is 25 ng/ml, and the concentration of rTwsg is 10 nM.

3.6 Discussion

A fundamental question in development is how a handful of mobile extracellular regulators work together to generate complex, and even counter-intuitive, developmental patterns. The BMP system provides a classic, well-studied example in which the BMP family of morphogens and their modulators work together to generate diverse patterns across different developmental contexts. Although specific examples have been studied in details, it has remained unclear what kinds of spatial patterns BMP and its modulators can produce, what roles different components play in pattern formation, and whether these components are sufficient to synthetically generate pattern formation de novo. In particular, a circuit with BMP, Chordin, Twsg, and BMP-1 has been suggested to enable ligand shuttling and generation of certain gradient shape such as a displaced gradient (*25*), but two studies in zebrafish embryos showed that BMP ligands were not shuttled despite the presence of all circuit components (*27*, *28*). This controversy may result from a limitation of top-down approaches such as genetic perturbation experiments in embryos, since the same circuit may behave very differently in distinct developmental contexts. We sought to comprehensively explore the behaviors of this circuit using a bottom-up reconstitution approach. Through adding one component at a time and gradually increasing the circuit complexity, we found that this circuit had diverse gradient modulation capabilities, both in the mathematical model and in experiments. Chordin by itself suppresses the gradient amplitude and increases the gradient length in a dose-dependent manner (Fig. 3.4B and Fig. 3.5B). In the presence of Twsg, Chordin can completely suppress the gradient formation (Fig. 3.4C and Fig. 3.5C). Finally, in the presence of Chordin, Twsg, and BMP-1, the full circuit is sufficient to displace BMP4

signaling gradients away from BMP sending source, to a place that cannot be reached by BMP4 diffusion itself. This demonstrates that a circuit with BMP, Chordin, Twsg, and BMP-1 is sufficient for ligand shuttling and gradient displacement. Tuning BMP-1 level further changed the peak positions and amplitudes of the displace gradients, revealing additional gradient modulation capabilities.

Biochemical and genetics studies showed that interactions among BMP, Chordin, Twsg, and BMP-1 are largely conserved across vertebrates (*17*, *20*, *33–35*), and the mathematical models from this and other studies showed that a circuit with these interactions is sufficient to enable ligand shuttling. Bottom-up gradient reconstitution experiments from this study confirmed these model predictions, suggesting that this circuit behavior is largely conserved across evolution. What are the functional implications of shuttling? Ligand shuttling by Chordin, Twsg, and BMP-1, combined with BMP signaling dependent feedback, has been suggested to confer robustness to gradient formation (*23*, *24*) and allow for gradient scaling (*25*). However, alternative mechanisms are also proposed for these behaviors (*36*), provoking the question of whether ligand shuttling combined with feedback is sufficient for robust gradient formation and gradient scaling. This kind of question can be answered by the bottom-up reconstitution approach used in this study. It would be interesting to reconstitute more complex gradient modulation and find minimal circuits sufficient for certain modulation using this approach.

Apart from Chordin, Twsg, and BMP-1, BMP gradients are modulated by many other extracellular modulators (*6*, *7*). Our system provides a scalable platform to test these modulators individually and in combinations. For example, the extracellular modulator

CV2 has been shown to assist ligand shuttling in mouse vertebral field (*37*). Though it is not expressed in our reconstituted system thus not required for ligand shuttling, it would be interesting to see how the addition of CV2 changes gradient modulation capabilities by Chordin, Twsg, and BMP-1, and what other modulation modes it enables.

Using synthetic and natural cell signaling components, biologists have begun to design more and more pattern formation programs in multicellular systems (*11*, *38–41*). The ability to reconstitute ligand shuttling behaviors using a small set of natural proteins provide additional programmable components to design complex multicellular patterns. Combined with synthetic cell fate control systems such as MultiFate (*42*), one can design a larger feedback circuit, where the morphogen gradient instructs synthetic cell fate decisions, and cells in different states express distinct combinations of ligands and modulators. This kind of circuit could provide a foundation for engineering complex synthetic tissue patterning systems.

3.7 Supplementary Information

The mathematical model of BMP4-Chordin-Twsg-BMP-1 circuit

Here we introduce the mathematical model of BMP4-Chordin-Twsg-BMP-1 circuit, which is based on previous models (*23–25*, *32*) but specifically incorporating the components analyzed here (Fig. 3.4A), including mobile extracellular components BMP4 ($[BMP4]$), Chordin ($[Chordin]$), Twsg ($[Twsg]$), BMP4-Chordin complex ($[BMP4\text{–}Chordin]$), and BMP4-Chordin-Twsg ($[BMP4\text{–}Chordin\text{–}Twsg]$) complex, as well as immobile components receptors ($[R]$), BMP4-receptor complex ($[BR]$), and fluorescent reporters ($[Cit]$). Since BMP-1 is uniformly expressed by receiver cells, we

used the parameter β to represent its Chordin cleavage efficiency, which depends on the expression level of BMP-1.

We used a set of reaction-diffusion partial differential equations for mobile components. Thus, their equations have two parts. The first part is a diffusion term $D\nabla^2 c$, where c can be $[BMP4]$, $[Chordin]$, $[Twsg]$, $[BMP4\text{–}Chordin]$, $[BMP4\text{–}Chordin\text{–}Twsg]$, and D can be D_b, D_c, D_t, D_{bc}, and D_{bct}, correspondingly. We used the effective diffusion coefficients for each mobile component, taking into account the effects of their interactions with extracellular matrix components such as heparan sulfate polyglycans, but not BMP4-receptor interactions, which will be considered separately. The estimated diffusion coefficients for these mobile components vary among studies (*28*, *43*), and we chose the same intermediate value of 20 $\mu m^2/s$ among these estimates.

The second part contains reaction terms. First, BMP4 can bind to Chordin weakly, with estimated association rate $k_{BC} = 2.8 \times 10^{-4}$ $nM^{-1}s^{-1}$ and dissociation rate $r_{BC} = 3.4 \times 10^{-3}$ s^{-1} from (*44*). Then BMP4-Chordin complex can further interact with Twsg to form BMP4-Chordin-Twsg complex, with association rate k_{BCT} and dissociation rate r_{BCT}. There is no measurement for k_{BCT} and r_{BCT}. We used a similar $k_{BCT} = 3 \times 10^{-4}$ $nM^{-1}s^{-1}$ with k_{BC}, and a $r_{BCT} = 3 \times 10^{-5}$ s^{-1} such that the dissociation constant is 0.1 nM. We assumed this low dissociation constant because the BMP4-Chordin-Twsg complex has been suggested to be very stable once formed (*17–20*). Furthermore, BMP4 can bind to receptors with estimated association rate $k_R = 6 \times 10^{-4}$ $nM^{-1}s^{-1}$ and dissociation rate $r_R = 4 \times 10^{-4}$ s^{-1} from (*45*). Finally,

Chordin in its free form or in complex forms can be cleaved by BMP-1 with a rate β dependent on BMP-1 expression level, and BMP4 and Twsg can be released from BMP4-Chordin and BMP4-Chordin-Twsg complex once Chordin is cleaved. Thus we can write,

$$\partial[BMP4]/\partial t = D_b\nabla^2[BMP4] - k_{BC}[BMP4][Chordin] + r_{BC}[BMP4\text{–}Chordin]$$

$$- k_R[BMP4][R] + r_R[BR] + \beta[BMP4\text{–}Chordin]$$

$$+ \beta[BMP4\text{–}Chordin\text{–}Twsg]$$

$$\partial[Chordin]/\partial t = D_c\nabla^2[Chordin] - k_{BC}[BMP4][Chordin]$$

$$+ r_{BC}[BMP4\text{–}Chordin] - \beta[Chordin]$$

$$\partial[Twsg]/\partial t = D_t\nabla^2[Twsg] - k_{BCT}[Twsg][BMP4\text{–}Chordin]$$

$$+ r_{BCT}[BMP4\text{–}Chordin\text{–}Twsg] + \beta[BMP4\text{–}Chordin\text{–}Twsg]$$

$$\partial[BMP4\text{–}Chordin]/\partial t = D_{bc}\nabla^2[BMP4\text{–}Chordin] + k_{BC}[BMP4][Chordin]$$

$$- r_{BC}[BMP4\text{–}Chordin] - k_{BCT}[Twsg][BMP4\text{–}Chordin]$$

$$+ r_{BCT}[BMP4\text{–}Chordin\text{–}Twsg] - \beta[BMP4\text{–}Chordin]$$

$$\partial[BMP4\text{–}Chordin\text{–}Twsg]/\partial t = D_{bct}\nabla^2[BMP4\text{–}Chordin\text{–}Twsg]$$

$$+ k_{BCT}[Twsg][BMP4\text{–}Chordin]$$

$$- r_{BCT}[BMP4\text{–}Chordin\text{–}Twsg]$$

$$- \beta[BMP4\text{–}Chordin\text{–}Twsg]$$

Note that Twsg can also interact with BMP4 and Chordin individually (*46*), so there exist multiple interaction routes of forming the final BMP4-Chordin-Twsg complex. However, since interactions between Twsg and BMP4 or Chordin are much weaker than interactions between BMP4 and Chordin, we only consider one interaction route (BMP4 first interacts with Chordin, then BMP4-Chordin interacts with Twsg) in our model.

For the immobile BMP4 receptors ($[R]$), other than reversible interactions with BMP4 ligands, we also considered receptor-mediated internalization and degradation of BMP4 ligands [ref]. Once the BMP4-receptor complex ($[BR]$) is internalized, we assumed that the BMP4 ligand is degraded and the receptor is recycled with a rate $\gamma = 2 \times 10^{-4}$ s^{-1} estimated from (*47*). Thus we can write,

$$\partial[R]/\partial t = - k_R[BMP4][R] + r_R[BR] + \gamma[BR]$$

$$\partial[BR]/dt = k_R[BMP4][R] - r_R[BR] - \gamma[BR]$$

From these equations we can see that the total receptor concentration ($[R] + [BR]$) is held constant. There is no direct estimate of typical BMP receptor density on the cell surface, and the estimate for the densities of other receptors vary from 10^2 to 10^3 molecules/μm^2 (*48–50*). We used a relatively low density of 150 molecules/μm^2, which corresponds to 0.3 nM, given a 800 μm thick Matrigel layer on top of the cell.

Finally, we assume fluorescent reporter production rate follows a Hill function with BMP4-receptor complex concentrations as variable. Thus we can write,

$$\partial[Cit]/\partial t = b[BR]^n/(K^n + [BR]^n) - \delta[Cit]$$

We used a degradation rate $\delta = 1 \times 10^{-5}$ s^{-1}, corresponding to a ~20hr Citrine fluorescent protein half-life. Parameters in Hill function b, n, and K were acquired by data fitting.

3.8 References

1. A. Turning, The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. **237**, 37–72 (1952).

2. G. Struhl, K. Struhl, P. M. Macdonald, The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*. **57**, 1259–1273 (1989).

3. W. Driever, C. Nüsslein-Volhard, A gradient of bicoid protein in Drosophila embryos. *Cell*. **54**, 83–93 (1988).

4. W. Driever, C. Nüsslein-Volhard, The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*. **54**, 95–104 (1988).

5. J. B. Gurdon, P.-Y. Bourillot, Morphogen gradient interpretation. *Nature*. **413**, 797–803 (2001).

6. L. Zakin, E. M. De Robertis, Extracellular regulation of BMP signaling. *Current Biology*. **20**, R89–R92 (2010).

7. D. Umulis, M. B. O'Connor, S. S. Blair, The extracellular regulation of bone morphogenetic protein signaling. *Development*. **136**, 3715–3728 (2009).

8. J. Capdevila, J. C. I. Belmonte, Extracellular modulation of the Hedgehog, Wnt and TGF-β signalling pathways during embryonic development. *Current Opinion in Genetics & Development*. **9**, 427–433 (1999).

9. T. Tabata, Genetics of morphogen gradients. *Nature Reviews Genetics*. **2**, 620–630 (2001).

10. P. Li, J. S. Markson, S. Wang, S. Chen, V. Vachharajani, M. B. Elowitz, Morphogen gradient reconstitution reveals Hedgehog pathway design principles. *Science*. **360**, 543–548 (2018).

11. R. Sekine, T. Shibata, M. Ebisuya, Synthetic mammalian pattern formation driven by differential diffusivity of Nodal and Lefty. *Nat. Commun.* **9**, 5456 (2018).

12. K. A. Wharton, R. P. Ray, W. M. Gelbart, An activity gradient of decapentaplegic is necessary for the specification of dorsal pattern elements in the Drosophila embryo. *Development*. **117**, 807–822 (1993).

13. L. Dale, F. C. Wardle, A gradient of BMP activity specifies dorsal–ventral fates in early Xenopus embryos. *Seminars in Cell & Developmental Biology*. **10**, 319–326 (1999).

14. L. Marchant, C. Linker, P. Ruiz, N. Guerrero, R. Mayor, The inductive properties of mesoderm suggest that the neural crest cells are specified by a BMP gradient. *Developmental Biology*. **198**, 319–329 (1998).

15. J. Raspopovic, L. Marcon, L. Russo, J. Sharpe, Modeling digits. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science*. **345**, 566–570 (2014).

16. S. Piccolo, Y. Sasai, B. Lu, E. M. De Robertis, Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell*. **86**, 589–598 (1996).

17. J. J. Ross, O. Shimmi, P. Vilmos, A. Petryk, H. Kim, K. Gaudenz, S. Hermanson, S. C. Ekker, M. B. O'Connor, J. L. Marsh, Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature*. **410**, 479–483 (2001).

18. C. Chang, D. A. Holtzman, S. Chau, T. Chickering, E. A. Woolf, L. M. Holmgren, J. Bodorova, D. P. Gearing, W. E. Holmes, A. H. Brivanlou, Twisted gastrulation can function as a BMP antagonist. *Nature*. **410**, 483–487 (2001).

19. I. C. Scott, I. L. Blitz, W. N. Pappano, S. A. Maas, K. W. Y. Cho, D. S. Greenspan, Homologues of Twisted gastrulation are extracellular cofactors in antagonism of BMP signalling. *Nature*. **410**, 475–478 (2001).

20. O. Shimmi, M. B. O'Connor, Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the Drosophila embryo. *Development*. **130**, 4673–4682 (2003).

21. M. B. O'Connor, D. Umulis, H. G. Othmer, S. S. Blair, Shaping BMP morphogen gradients in the Drosophila embryo and pupal wing. *Development*. **133**, 183–193 (2006).

22. I. C. Scott, I. L. Blitz, W. N. Pappano, Y. Imamura, T. G. Clark, B. M. Steiglitz, C. L. Thomas, S. A. Maas, K. Takahara, K. W. Cho, D. S. Greenspan, Mammalian BMP-1/Tolloid-related metalloproteinases, including novel family member mammalian Tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. *Dev. Biol.* **213**, 283–300 (1999).

23. A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, N. Barkai, Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature*. **419**, 304–308 (2002).

24. C. M. Mizutani, Q. Nie, F. Y. M. Wan, Y.-T. Zhang, P. Vilmos, R. Sousa-Neves, E. Bier, J. L. Marsh, A. D. Lander, Formation of the BMP activity gradient in the Drosophila embryo. *Dev. Cell*. **8**, 915–924 (2005).

25. D. Ben-Zvi, B.-Z. Shilo, A. Fainsod, N. Barkai, Scaling of the BMP activation gradient in

Xenopus embryos. *Nature*. **453**, 1205–1211 (2008).

26. L. Zakin, C. A. Metzinger, E. Y. Chang, C. Coffinier, E. M. De Robertis, Development of the vertebral morphogenetic field in the mouse: interactions between Crossveinless-2 and Twisted Gastrulation. *Dev. Biol.* **323**, 6–18 (2008).

27. J. Zinski, Y. Bu, X. Wang, W. Dou, D. Umulis, M. C. Mullins, Systems biology derived source-sink mechanism of BMP gradient formation. *Elife*. **6** (2017), doi:10.7554/eLife.22199.

28. A. P. Pomreinke, G. H. Soh, K. W. Rogers, J. K. Bergmann, A. J. Bläßle, P. Müller, Dynamics of BMP signaling and distribution during zebrafish dorsal-ventral patterning. *Elife*. **6** (2017), doi:10.7554/eLife.25861.

29. H. Yano, H. Uchida, T. Iwasaki, M. Mukai, H. Akedo, K. Nakamura, S. Hashimoto, H. Sabe, Paxillin α and Crk-associated substrate exert opposing effects on cell migration and contact inhibition of growth through tyrosine phosphorylation. *Proceedings of the National Academy of Sciences*. **97**, 9076–9081 (2000).

30. Y. E. Antebi, J. M. Linton, H. Klumpe, B. Bintu, M. Gong, C. Su, R. McCardell, M. B. Elowitz, Combinatorial Signal Perception in the BMP Pathway. *Cell*. **170**, 1184–1196.e24 (2017).

31. A. Kicheva, P. Pantazis, T. Bollenbach, Y. Kalaidzidis, T. Bittig, F. Jülicher, M. González-Gaitán, Kinetics of Morphogen Gradient Formation. *Science*. **315**, 521–525 (2007).

32. A. Eldar, D. Rosin, B.-Z. Shilo, N. Barkai, Self-enhanced ligand degradation underlies robustness of morphogen gradients. *Dev. Cell*. **5**, 635–646 (2003).

33. H. Troilo, A. V. Zuk, R. B. Tunnicliffe, A. P. Wohl, R. Berry, R. F. Collins, T. A. Jowitt, G. Sengle, C. Baldock, Nanoscale structure of the BMP antagonist chordin supports cooperative BMP binding. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13063–13068 (2014).

34. H. Troilo, A. L. Barrett, A. V. Zuk, M. P. Lockhart-Cairns, A. P. Wohl, C. P. Bayley, R. Dajani, R. B. Tunnicliffe, L. Green, T. A. Jowitt, G. Sengle, C. Baldock, Structural characterization of twisted gastrulation provides insights into opposing functions on the BMP signalling pathway. *Matrix Biol.* **55**, 49–62 (2016).

35. P. Blader, S. Rastegar, N. Fischer, U. Strähle, Cleavage of the BMP-4 Antagonist Chordin by Zebrafish Tolloid. *Science*. **278**, 1937–1940 (1997).

36. H. Inomata, T. Shibata, T. Haraguchi, Y. Sasai, Scaling of Dorsal-Ventral Patterning by Embryo Size-Dependent Degradation of Spemann's Organizer Signals. *Cell*. **153**, 1296–1311 (2013).

37. L. Zakin, E. Y. Chang, J.-L. Plouhinec, E. M. De Robertis, Crossveinless-2 is required for the relocalization of Chordin protein within the vertebral field in mouse embryos. *Dev. Biol.* **347**, 204–215 (2010).

38. M. Matsuda, M. Koga, K. Woltjen, E. Nishida, M. Ebisuya, Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat. Commun.* **6**, 6195 (2015).

39. S. Toda, L. R. Blauch, S. K. Y. Tang, L. Morsut, W. A. Lim, Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science*. **361**, 156–162 (2018).

40. S. Toda, W. L. McKeithan, T. J. Hakkinen, P. Lopez, O. D. Klein, W. A. Lim, Engineering synthetic morphogen systems that can program multicellular patterning. *Science*. **370**, 327–331 (2020).

41. K. S. Stapornwongkul, M. de Gennes, L. Cocconi, G. Salbreux, J.-P. Vincent, Patterning and growth control in vivo by an engineered GFP gradient. *Science*. **370**, 321–327 (2020).

42. R. Zhu, J. M. Del Rio-Salgado, J. Garcia-Ojalvo, M. B. Elowitz, Synthetic multistability in mammalian cells. *Science*. **375**, eabg9765 (2022).

43. S. Zhou, W.-C. Lo, J. L. Suhalim, M. A. Digman, E. Gratton, Q. Nie, A. D. Lander, Free extracellular diffusion creates the Dpp morphogen gradient of the Drosophila wing disc. *Curr. Biol.* **22**, 668–675 (2012).

44. F. Rentzsch, J. Zhang, C. Kramer, W. Sebald, M. Hammerschmidt, Crossveinless 2 is an essential positive feedback regulator of Bmp signaling during zebrafish gastrulation. *Development*. **133**, 801–811 (2006).

45. T. Kirsch, J. Nickel, W. Sebald, BMP-2 antagonists emerge from alterations in the low-affinity binding epitope for receptor BMPR-II. *EMBO J.* **19**, 3314–3324 (2000).

46. J.-L. Zhang, Y. Huang, L.-Y. Qiu, J. Nickel, W. Sebald, von Willebrand Factor Type C Domain-containing Proteins Regulate Bone Morphogenetic Protein Signaling through Different Recognition Mechanisms. *Journal of Biological Chemistry*. **282**, 20002–20014 (2007).

47. S. N. Roed, P. Wismann, C. R. Underwood, N. Kulahin, H. Iversen, K. A. Cappelen, L. Schäffer, J. Lehtonen, J. Hecksher-Soerensen, A. Secher, J. M. Mathiesen, H. Bräuner-Osborne, J. L. Whistler, S. M. Knudsen, M. Waldhoer, Real-time trafficking and signaling of the glucagon-like peptide-1 receptor. *Mol. Cell. Endocrinol.* **382**, 938–949 (2014).

48. M. Wang, M. Misakian, H.-J. He, P. Bajcsy, F. Abbasi, J. M. Davis, K. D. Cole, I. V. Turko, L. Wang, Quantifying CD4 receptor protein in two human CD4+ lymphocyte preparations for quantitative flow cytometry. *Clin. Proteomics*. **11**, 43 (2014).

49. B. H. Meyer, J.-M. Segura, K. L. Martinez, R. Hovius, N. George, K. Johnsson, H. Vogel, FRET imaging reveals that functional neurokinin-1 receptors are monomeric and reside in membrane microdomains of live cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2138–2143 (2006).

50. Y. Chen, A. C. Munteanu, Y.-F. Huang, J. Phillips, Z. Zhu, M. Mavros, W. Tan, Mapping receptor density on live cells by using fluorescence correlation spectroscopy. *Chemistry*. **15**, 5327–5336 (2009).

Chapter 4. Concluding remarks

4.1 Bottom-up engineering and reconstitution reveal design principles of multicellular circuits in mammalian cells

Multicellular circuits control the development of multicellular organisms, programming a variety of multicellular processes such as cell proliferation, cell differentiation, cell movement, and cell signaling (*1*). Top-down approaches have identified key components for many multicellular circuits, but the design principles of these circuits remained generally unclear, due to the limitations from top-down approaches to study these circuits in an isolated context and in a quantitative and systematic manner. In this thesis, we have taken an alternative, complementary approach: engineering and reconstituting multicellular circuits from bottom up to gain quantitative understanding of their design principles.

In the first project, we summarized two prevalent features of natural cell fate control circuits: positive autoregulation and promiscuous protein dimerization, and engineered a synthetic multistable circuit architecture named MultiFate based on these two features. MultiFate circuits recapitulate many properties of the natural cell fate control system. They robustly generate multiple cellular states, each mitotically heritable for weeks. Similar to the idea of cell fate reprogramming (*2*), they allow controlled switching of cells among these states through transient transcription factor overexpression. Through externally controlling key circuit parameters such as protein stability, we were able to modulate the state stability and recapitulate irreversible cell state transitions observed during stem cell differentiation (*3*). Finally, because cross-inhibition among transcription

factors is implemented through promiscuous protein-protein dimerization, MultiFate architecture is easily expandable by 'plugging in' additional transcription factors, without re-engineering the existing circuit. Since promiscuous protein-protein dimerization among fate regulators is very prevalent in natural multistable circuits (*4–10*), the same design principle may have a related role, allowing the evolution of new cell states through transcription factor duplication and subfunctionalization (*7*, *8*, *11*, *12*).

In the second project, we used a gradient reconstitution system to dissect a gradient modulation circuit consisting of a morphogen BMP4, and its modulators, Chordin, Twsg, and BMP-1. The reconstitution system allows us to add circuit components one at a time to slowly increase the complexity, and individually tune the expression level of each circuit component. Through this systematic exploration, we revealed a diverse gradient modulation capabilities by this circuit, including gradient lengthening, gradient suppression, and gradient displacement. In particular, our reconstitution system confirms the sufficiency of this circuit for active ligand shuttling, a phenomenon observed in some model organisms (*13–16*) but not others (*17*, *18*).

One common theme of these two projects is the close agreement of mathematical modeling and experimental results. Both mathematical models do not have precise quantitative parameter values for many molecular interactions, but the qualitative behaviors possible with these two circuits can be enumerated, explained and predicted from simple properties of the components and their interactions. These mathematical models are best synergistic with our bottom-up approaches, where we could comprehensively explore these qualitative behaviors to validate model predictions because a multicellular circuit can be studied in a relatively isolated context, and circuit

complexity and the abundance and activities of circuit components can be well-controlled.

4.2 Future perspectives

The tools and systems developed here can be applied to understand the design of other multicellular circuits. In the first project, we could couple MultiFate to synthetic cell-cell communication systems such as synNotch (*19*, *20*), MESA (*21*), synthekines (*22*), engineered GFP (*23*), and auxin (*24*), and navigate cells through a series of fate choices, recapitulating cell behaviors associated with normal development. In the second project, the reconstitution system can be repurposed to study how other BMP extracellular modulators, individually or in combination, change the shape of BMP gradients. It can also be used to reconstitute more complex circuits (*13–15*) through adding feedback to the BMP4-Chordin-Twsg-BMP-1 circuit. These circuits have been suggested to confer robustness to gradient formation (*25*, *26*) and allow for gradient scaling (*15*), but the minimal circuits sufficient for these more complex behaviors have not been shown experimentally.

Finally, the MultiFate circuit and BMP4-Chordin-Twsg-BMP-1 circuit could be combined to create novel synthetic tissue patterns. On one hand, MultiFate state-switching can be controlled by BMP4 signaling, such that cells switch to different cellular states based on the strength of BMP4 signaling. On the other hand, cells in different MultiFate states can express distinct combinations of BMP4 ligands and modulators that reshape the BMP4 signaling gradients. Similar reaction-diffusion systems have been shown to create diverse synthetic patterns in vitro (*27*, *28*). An advantage of

combining MultiFate and BMP4-Chordin-Twsg-BMP-1 circuits is the availability of tractable mathematical models for both circuits. Combining two mathematical models should allow us to first computationally explore the possible synthetic patterns enabled by the combined circuit, and then use the model to guide the circuit design for certain desired patterns. This kind of circuit could provide a foundation for engineering complex synthetic tissue patterning systems in the future.

4.3 References

1. P. Weiss, THE SO-CALLED ORGANIZER AND THE PROBLEM OF ORGANIZATION IN AMPHIBIAN DEVELOPMENT. *Physiological Reviews*. **15**, 639–674 (1935).

2. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. **126**, 663–676 (2006).

3. S. Huang, Y.-P. Guo, G. May, T. Enver, Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713 (2007).

4. C. Sokolik, Y. Liu, D. Bauer, J. McPherson, M. Broeker, G. Heimberg, L. S. Qi, D. A. Sivak, M. Thomson, Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Syst*. **1**, 117–129 (2015).

5. A. Bhattacharya, N. E. Baker, A network of broadly expressed HLH genes regulates tissue-specific cell fates. *Cell*. **147**, 881–892 (2011).

6. H. Hosokawa, J. Ungerbäck, X. Wang, M. Matsumoto, K. I. Nakayama, S. M. Cohen, T. Tanaka, E. V. Rothenberg, Transcription Factor PU.1 Represses and Activates Gene Expression in Early T Cells by Redirecting Partner Transcription Factor Binding. *Immunity*. **48** (2018), pp. 1119–1134.e7.

7. S. Stefanovic, N. Abboud, S. Désilets, D. Nury, C. Cowan, M. Pucéat, Interplay of Oct4 with Sox2 and Sox17: a molecular switch from stem cell pluripotency to specifying a cardiac fate. *J. Cell Biol.* **186**, 665–673 (2009).

8. I. Aksoy, R. Jauch, J. Chen, M. Dyla, U. Divakar, G. K. Bogu, R. Teo, C. K. Leng Ng, W. Herath, S. Lili, A. P. Hutchins, P. Robson, P. R. Kolatkar, L. W. Stanton, Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J.* **32**, 938–953 (2013).

9. M. J. Thayer, S. J. Tapscott, R. L. Davis, W. E. Wright, A. B. Lassar, H. Weintraub, Positive autoregulation of the myogenic determination gene MyoD1. *Cell*. **58**, 241–248 (1989).

10. L. A. Neuhold, B. Wold, HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. *Cell*. **74**, 1033–1042

(1993).

11. K. Pougach, A. Voet, F. A. Kondrashov, K. Voordeckers, J. F. Christiaens, B. Baying, V. Benes, R. Sakai, J. Aerts, B. Zhu, P. Van Dijck, K. J. Verstrepen, Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat. Commun.* **5**, 4868 (2014).

12. J. González, G. López, S. Argueta, X. Escalera-Fanjul, M. El Hafidi, C. Campero-Basaldua, J. Strauss, L. Riego-Ruiz, A. González, Diversification of Transcriptional Regulation Determines Subfunctionalization of Paralogous Branched Chain Aminotransferases in the Yeast. *Genetics*. **207**, 975–991 (2017).

13. A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, N. Barkai, Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature*. **419**, 304–308 (2002).

14. C. M. Mizutani, Q. Nie, F. Y. M. Wan, Y.-T. Zhang, P. Vilmos, R. Sousa-Neves, E. Bier, J. L. Marsh, A. D. Lander, Formation of the BMP activity gradient in the Drosophila embryo. *Dev. Cell*. **8**, 915–924 (2005).

15. D. Ben-Zvi, B.-Z. Shilo, A. Fainsod, N. Barkai, Scaling of the BMP activation gradient in Xenopus embryos. *Nature*. **453**, 1205–1211 (2008).

16. L. Zakin, C. A. Metzinger, E. Y. Chang, C. Coffinier, E. M. De Robertis, Development of the vertebral morphogenetic field in the mouse: interactions between Crossveinless-2 and Twisted Gastrulation. *Dev. Biol.* **323**, 6–18 (2008).

17. J. Zinski, Y. Bu, X. Wang, W. Dou, D. Umulis, M. C. Mullins, Systems biology derived source-sink mechanism of BMP gradient formation. *Elife*. **6** (2017), doi:10.7554/eLife.22199.

18. A. P. Pomreinke, G. H. Soh, K. W. Rogers, J. K. Bergmann, A. J. Bläßle, P. Müller, Dynamics of BMP signaling and distribution during zebrafish dorsal-ventral patterning. *Elife*. **6** (2017), doi:10.7554/eLife.25861.

19. L. Morsut, K. T. Roybal, X. Xiong, R. M. Gordley, S. M. Coyle, M. Thomson, W. A. Lim, Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell*. **164**, 780–791 (2016).

20. S. Toda, W. L. McKeithan, T. J. Hakkinen, P. Lopez, O. D. Klein, W. A. Lim, Engineering synthetic morphogen systems that can program multicellular patterning. *Science*. **370**, 327–331 (2020).

21. K. A. Schwarz, N. M. Daringer, T. B. Dolberg, J. N. Leonard, Rewiring human cellular input–output using modular extracellular sensors. *Nature Chemical Biology*. **13** (2017), pp. 202–209.

22. I. Moraga, J. B. Spangler, J. L. Mendoza, M. Gakovic, T. S. Wehrman, P. Krutzik, K. C. Garcia, Synthekines are surrogate cytokine and growth factor agonists that compel signaling through non-natural receptor dimers. *Elife*. **6** (2017), doi:10.7554/eLife.22882.

23. K. S. Stapornwongkul, M. de Gennes, L. Cocconi, G. Salbreux, J.-P. Vincent, Patterning and growth control in vivo by an engineered GFP gradient. *Science*. **370**, 321–327 (2020).

24. Y. Ma, M. W. Budde, M. N. Mayalu, J. Zhu, R. M. Murray, M. B. Elowitz, Synthetic mammalian signaling circuits for robust cell population control. *bioRxiv*, doi:10.1101/2020.09.02.278564.

25. A. Eldar, R. Dorfman, D. Weiss, H. Ashe, B.-Z. Shilo, N. Barkai, Robustness of the BMP morphogen gradient in Drosophila embryonic patterning. *Nature*. **419**, 304–308 (2002).

26. C. M. Mizutani, Q. Nie, F. Y. M. Wan, Y.-T. Zhang, P. Vilmos, R. Sousa-Neves, E. Bier, J. L. Marsh, A. D. Lander, Formation of the BMP activity gradient in the Drosophila embryo. *Dev. Cell*. **8**, 915–924 (2005).

27. M. Matsuda, M. Koga, K. Woltjen, E. Nishida, M. Ebisuya, Synthetic lateral inhibition governs cell-type bifurcation with robust ratios. *Nat. Commun.* **6**, 6195 (2015).

28. R. Sekine, T. Shibata, M. Ebisuya, Synthetic mammalian pattern formation driven by differential diffusivity of Nodal and Lefty. *Nat. Commun.* **9**, 5456 (2018).